AWS Certified AI Practitioner (AIF-C01) Exam Guide

Introduction

The AWS Certified AI Practitioner (AIF-C01) exam is intended for individuals who can effectively demonstrate overall knowledge of AI/ML, generative AI technologies, and associated AWS services and tools, independent of a specific job role.

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

- Understand AI, ML, and generative AI concepts, methods, and strategies in general and on AWS.
- Understand the appropriate use of AI/ML and generative AI technologies to ask relevant questions within the candidate's organization.
- Determine the correct types of AI/ML technologies to apply to specific use cases.
- Use AI, ML, and generative AI 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, and Amazon SageMaker) 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 the AWS global infrastructure, including the concepts of AWS Regions, Availability Zones, and edge locations
- 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

Refer to the Appendix for a list of in-scope AWS services and features and a list of out-of-scope AWS services and features.

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.
- **Case study:** Has one scenario with two or more questions about the scenario. The scenario is the same for each question in the case study. Each question in the case study will be evaluated separately. You will receive credit for each question that you answer correctly in the case study.

Unanswered questions are scored as incorrect; there is no penalty for guessing. The exam includes 50 questions that affect your score.[1]

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 these questions for future use as scored questions. These 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.

---

[1] Does not apply to the beta version of the exam. You can find more information about beta exams in general on the [AWS Certification website](https://).

Content outline

This exam guide includes weightings, content domains, and task statements for the exam. This guide does not provide a comprehensive list of the content on the exam. However, additional context for each task statement is available to help you prepare for the exam.

The exam has the following content domains and weightings:

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

Domain 1: Fundamentals of AI and ML

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

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 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).
- Explain the capabilities of AWS managed AI/ML services (for example, SageMaker, Amazon Transcribe, Amazon Translate, Amazon Comprehend, Amazon Lex, Amazon Polly).

Task Statement 1.3: Describe the ML development lifecycle.

Objectives:

- Describe components of an ML pipeline (for example, data collection, exploratory data analysis [EDA], data pre-processing, feature engineering, model training, hyperparameter tuning, evaluation, deployment, monitoring).
- Understand sources of ML 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 ML pipeline (for example, SageMaker, Amazon SageMaker Data Wrangler, Amazon SageMaker Feature Store, Amazon SageMaker Model Monitor).

- Understand fundamental concepts of ML operations (MLOps) (for example, experimentation, repeatable processes, scalable systems, managing technical debt, achieving production readiness, model monitoring, model re-training).
- Understand model performance metrics (for example, accuracy, Area Under the ROC Curve [AUC], F1 score) and business metrics (for example, cost per user, development costs, customer feedback, return on investment [ROI]) to evaluate ML models.

Domain 2: Fundamentals of Generative AI

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

Objectives:
- Understand foundational generative AI concepts (for example, tokens, chunking, embeddings, vectors, prompt engineering, transformer-based LLMs, foundation models, multi-modal models, diffusion models).
- Identify potential use cases for generative AI models (for example, image, video, and audio generation; summarization; chatbots; translation; code generation; customer service agents; search; recommendation engines).
- Describe the foundation model lifecycle (for example, data selection, model selection, pre-training, fine-tuning, evaluation, deployment, feedback).

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

Objectives:
- Describe the advantages of generative AI (for example, adaptability, responsiveness, simplicity).
- Identify disadvantages of generative AI solutions (for example, hallucinations, interpretability, inaccuracy, nondeterminism).
- Understand various factors to select appropriate generative AI models (for example, model types, performance requirements, capabilities, constraints, compliance).
- Determine business value and metrics for generative AI applications (for example, cross-domain performance, efficiency, conversion rate, average revenue per user, accuracy, customer lifetime value).

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

Objectives:
- Identify AWS services and features to develop generative AI applications (for example, Amazon SageMaker JumpStart; Amazon Bedrock; PartyRock, an Amazon Bedrock Playground; Amazon Q).
- Describe the advantages of using AWS generative AI services to build applications (for example, accessibility, lower barrier to entry, efficiency, cost-effectiveness, speed to market, ability to meet business objectives).
- Understand the benefits of AWS infrastructure for generative AI applications (for example, security, compliance, responsibility, safety).
- Understand cost tradeoffs of AWS generative AI services (for example, responsiveness, availability, redundancy, performance, regional coverage, token-based pricing, provision throughput, custom models).

Domain 3: Applications of Foundation Models

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

Objectives:
- Identify selection criteria to choose pre-trained models (for example, cost, modality, latency, multi-lingual, model size, model complexity, customization, input/output length).
- Understand 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 base).
- Identify AWS services that help store embeddings within vector databases (for example, Amazon OpenSearch Service, Amazon Aurora, Amazon Neptune, Amazon DocumentDB [with MongoDB compatibility], Amazon RDS for PostgreSQL).

- Explain the cost tradeoffs of various approaches to foundation model customization (for example, pre-training, fine-tuning, in-context learning, RAG).
- Understand the role of agents in multi-step tasks (for example, Agents for Amazon Bedrock).

Task Statement 3.2: Choose effective prompt engineering techniques.

Objectives:
- Describe the concepts and constructs of prompt engineering (for example, context, instruction, negative prompts, model latent space).
- Understand techniques for prompt engineering (for example, chain-of-thought, zero-shot, single-shot, few-shot, prompt templates).
- Understand 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).

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

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

Task Statement 3.4: Describe methods to evaluate foundation model performance.

Objectives:
- Understand approaches to evaluate foundation model performance (for example, human evaluation, benchmark datasets).
- Identify relevant metrics to assess foundation model performance (for example, Recall-Oriented Understudy for Gisting Evaluation [ROUGE], Bilingual Evaluation Understudy [BLEU], BERTScore).
- Determine whether a foundation model effectively meets business objectives (for example, productivity, user engagement, task engineering).

Domain 4: Guidelines for Responsible AI

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).
- Understand how to use tools to identify features of responsible AI (for example, Guardrails for Amazon Bedrock).
- Understand responsible practices to select a model (for example, environmental considerations, sustainability).
- Identify legal risks of working with generative AI (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).
- Understand 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:
- Understand the differences between models that are transparent and explainable and models that are not transparent and explainable.
- Understand the tools to identify transparent and explainable models (for example, Amazon SageMaker Model Cards, open source models, data, licensing).
- Identify tradeoffs between model safety and transparency (for example, measure interpretability and performance).
- Understand principles of human-centered design for explainable AI.

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

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).
- Understand the concept of source citation and documenting data origins (for example, data lineage, data cataloging, SageMaker Model Cards).
- Describe best practices for secure data engineering (for example, assessing data quality, implementing privacy-enhancing technologies, data access control, data integrity).
- Understand 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).

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

Objectives:

- Identify regulatory compliance standards for AI systems (for example, International Organization for Standardization [ISO], System and Organization Controls [SOC], algorithm accountability laws).
- 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).

Appendix

In-scope AWS services and features

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

Analytics:

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

Cloud Financial Management:

- AWS Budgets
- AWS Cost Explorer

Compute:

- Amazon EC2

Containers:

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

Database:

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

Machine Learning:

- Amazon Augmented AI (Amazon A2I)
- Amazon Bedrock
- Amazon Comprehend
- Amazon Fraud Detector
- Amazon Kendra
- Amazon Lex
- Amazon Personalize
- Amazon Polly
- Amazon Q
- Amazon Rekognition
- Amazon SageMaker
- Amazon Textract
- Amazon Transcribe
- Amazon Translate

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 and features

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:

Analytics:

- AWS Clean Rooms
- Amazon CloudSearch
- Amazon FinSpace
- 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 Honeycode
- Amazon Pinpoint
- Amazon Simple Email Service (Amazon SES)
- AWS Supply Chain
- AWS Wickr
- Amazon WorkDocs
- 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

Survey

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