JON BONSO, KAYNE RODRIGO,
& SAMANTHA SERVO

AWS CERTIFIED
AI PRACTITIONER
AIF-C01
EXAM

Tutorials Dojo Study Guide

TABLE OF CONTENTS

INTRODUCTION

We live in an era where artificial intelligence (AI) and machine learning (ML) are revolutionizing industries and enhancing human capabilities. AWS, as a pioneer in cloud computing, offers an extensive array of AI/ML services that empower businesses to innovate and thrive. The AWS Certified AI Practitioner (AIF-C01) certification is designed to validate your understanding of AI, ML, and generative AI concepts and their applications on AWS, as well as your ability to identify appropriate AWS services to implement AI solutions. This certification is ideal for professionals who are familiar with AI/ML technologies and use them in their roles but do not necessarily build AI/ML solutions themselves.

AWS has consistently been recognized as the leading cloud provider in the market[1]. They are committed to continuously enhancing their services to ensure customer success and satisfaction. By earning the AWS Certified AI Practitioner certification, you demonstrate your expertise in identifying appropriate AI solutions based on specific business needs, which can significantly elevate your career prospects. This certification signals to employers your proficiency in leveraging AWS AI services to drive innovation and efficiency. It also positions you for career growth and higher earnings, as employers are willing to pay a premium for AI-skilled workers.

The journey to mastering AI and ML on AWS is both challenging and rewarding. The AWS Certified AI Practitioner credential serves as a robust foundation for further specializations in AI and ML. Whether you aim to become a data scientist, ML engineer, or other AI/ML-focused roles, this certification will equip you with the knowledge and skills necessary to excel in your chosen field. The credential also opens doors to new career opportunities and professional growth in a rapidly evolving technological landscape. AWS provides comprehensive resources and training to help you prepare for the exam and achieve your certification goals.

This Study Guide eBook is designed to provide you with the knowledge and practical skills needed to pass the AWS Certified AI Practitioner (AIF-C01) exam. It includes essential concepts, exam domains, tips for success, sample questions, cheat sheets, and other relevant information about the AIF-C01 exam. The guide begins with an overview of the exam structure, offering insights into the types of questions, the exam domains, the scoring scheme, and the benefits of passing the exam.

The contents are organized according to the official AIF-C01 exam guide, covering all pertinent AWS topics for each exam domain. The guide discusses various AWS concepts, related AI/ML services, and technical implementations to provide a clear understanding of what to expect on the actual exam.

Note: We've created these study guides and cheat sheets as supplementary resources. To enhance your exam readiness, we recommend using them alongside our high-quality practice exams, which will help you assess your knowledge and identify areas for improvement.

1 https://aws.amazon.com/blogs/aws/aws-named-as-a-leader-in-gartners-infrastructure-as-a-service-iaas-magic-quadrant-for-the-9th-consecutiveyear/

AWS CERTIFIED AI PRACTITIONER AIF-C01 EXAM OVERVIEW

Amazon Web Services began its Global Certification Program in 2013 with the primary purpose of validating the technical skills and knowledge of IT Professionals in building secure and reliable cloud-based applications using the AWS Cloud. In April 2013, AWS launched its first-ever AWS Certification test called the AWS Certified Solutions Architect Associate exam. This was followed by the AWS Certified SysOps Administrator and AWS Certified Developer Associate exams.

Amazon has been continuously expanding and updating its certification program year after year. They launched a series of Professional and Specialty-level certifications that cover various topics like DevOps, machine learning, data analytics, advanced networking, and many others. As the number of AWS services increases, a new and updated version of the AWS certification exam is released regularly to reflect the recent service changes and include the new knowledge areas.

In August 2024, AWS launched the AWS Certified AI Practitioner (AIF-C01) exam, a foundational certification for individuals who are familiar with AI/ML technologies on AWS. The AI Practitioner exam evaluates your understanding of concepts and use cases related to artificial intelligence (AI), machine learning (ML), and generative AI. This certification is ideal for roles such as business analysts, IT support, marketing professionals, product or project managers, and sales professionals. The exam has a duration of 90 minutes and consists of 65 questions. There are no prerequisites, allowing you to take the exam without the need for any prior certification, degree, or training.

The exam contains a mixture of scenario-based and multiple-choice questions, including multiple-response formats. The scenario-based questions have one correct answer and three incorrect responses, while the multiple-response format requires you to select two or more correct answers out of five or more options. Additionally, AWS has introduced three new question types: Ordering, where you get a list of 3 to 5 responses to complete a specific task, and you need to select the right ones and place them in the correct order; Matching, where you get a list of responses to match to 3 to 7 prompts; and Case Study, which involves detailed real-world scenarios requiring in-depth analysis and application of AWS AI principles. The exam costs **100 USD** and can be taken either at a local testing center or online from the comfort of your home.

The AI Practitioner certification exam has a total of 65 questions that you should complete within 90 minutes or one hour and a half. The score range for this test is from 100 to 1,000, with a minimum passing score of 700. AWS is using a scaled scoring model to equate scores across multiple exam types that may have different difficulty levels. An email of your result will be sent to you after a few days, and the complete score report will be available to your AWS Certification account afterward.

Individuals who unfortunately did not pass the AWS exam must wait for 14 days before they are allowed to retake the exam. There is no hard limit on the number of exam attempts, so you can try again and again until you pass the exam. Take note that on each attempt, the full registration price of the exam must be paid.

Your AWS Certification Account will have a record of your complete exam results within 5 business days of completing your exam. The score report contains a table of your performance for each exam domain, which indicates whether you met the competency level required for these domains or not. AWS uses a compensatory scoring model, which means that you do not necessarily need to pass each and every individual section.

You will pass this exam as long as you get an overall score of 700 across 4 domains. Each section has a specific score weighting that translates to the number of questions; hence, some sections have more questions than others. Your Score Performance table highlights your strengths and weaknesses that you need to improve on.

Exam Details

The AWS Certified AI Practitioner (AIF-C01) exam is designed for individuals who can effectively demonstrate a comprehensive understanding of AI/ML, generative AI technologies, and related AWS services and tools, regardless of their specific job role.

The exam includes various question types such as multiple-choice questions, where you select one correct response out of four options, and multiple-response questions, where you choose two or more correct responses out of five or more options. Additionally, it features ordering questions that require arranging responses in the correct order to complete a specified task, matching questions that involve pairing responses to a set of prompts correctly, and case study questions based on a scenario with two or more related questions. Each question is evaluated separately, and you receive credit for each correct response. You have the flexibility to take the exam either via online proctoring or at a testing center near you.

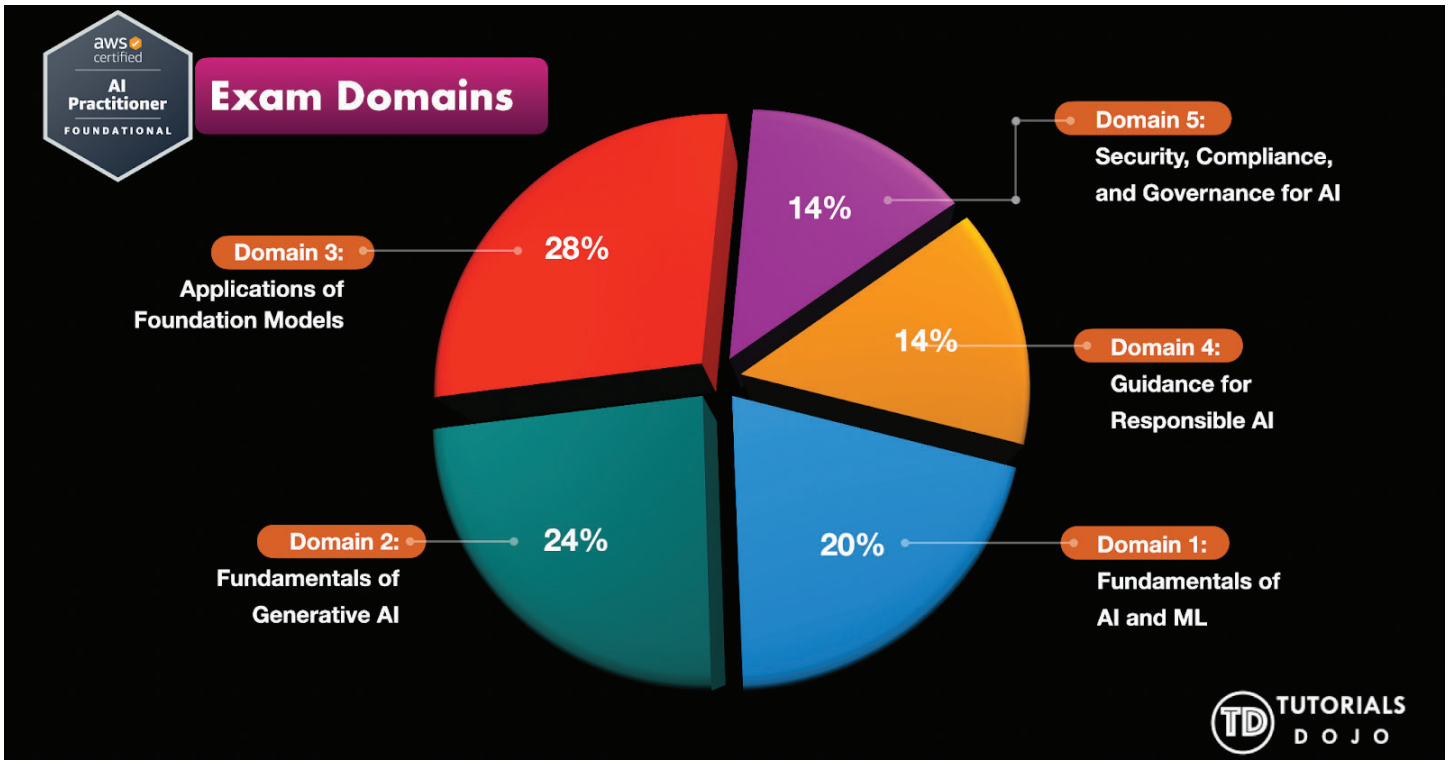
Exam Code:	AIF-C01
Prerequisites:	None
No. of Questions:	65
Score Range:	100-1000
Cost:	100 USD
Passing Score:	700
Time Limit:	90 minutes

Exam Domains

The **AWS Certified AI Practitioner (AIF-C01)** exam has five different domains, each with a corresponding weight and topic coverage. The domains are as follows:

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

One exam domain is comprised of several task statements. A task statement is a sub-category of the exam domain that contains the required cloud concepts, knowledge, and skills for you to accomplish a particular task or activity in AWS. In the AWS Certified AI Practitioner (AIF-C01) test, the **Domain 3: Applications of Foundation Models** has the biggest weighting in the exam at 28%, so expect to see a lot of foundation models-related scenarios in the exam. Conversely, the exam domain with the least exam weighting is **Domain 4: Guidelines for Responsible AI** and **Domain 5: Security, Compliance, and Governance for AI Solutions** so you have to limit the time you spend studying under this knowledge area.



Domain 1: Fundamentals of AI and ML

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

Domain 2: Fundamentals of Generative AI

2.1. Explain the basic concepts of generative AI.
2.2. Understand the capabilities and limitations of generative AI for solving business problems.
2.3. Describe AWS infrastructure and technologies for building generative AI applications.

Domain 3: Applications of Foundation Models

3.1. Describe design considerations for applications that use foundation models.
3.2. Choose effective prompt engineering techniques.
3.3. Describe the training and fine-tuning process for foundation models.
3.4. Describe methods to evaluate foundation model performance.

Domain 4: Guidelines for Responsible AI

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

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

5.1. Explain methods to secure AI systems.
5.2. Recognize governance and compliance regulations for AI systems.

Exam-Related AWS Topics and Services

The official exam guide contains a list of key tools, technologies, and concepts that may show up on the AI Practitioner AIF-C01 test. Keep in mind that this is just a non-exhaustive list of the tools and technologies that may or may not appear on the exam. This list can change at any time and is primarily given to test-takers to help them understand the general scope of services, features, or technologies for this certification. In addition, the general tools and technologies in this list appear in no particular order.

Here are the topics, AWS services, and concepts that you should focus on for your upcoming exam. You have to review your knowledge on the following:

• AI/ML • AI/ML Use Cases and Applications • Generative AI technologies • Types of AI/ML technologies • AWS shared responsibility model • Management and governance	• Amazon Translate • Amazon Comprehend • Amazon Lex • Amazon Polly • Amazon Bedrock • Amazon SageMaker AI

• AWS global infrastructure • Prompt Engineering • Foundation model • Foundation model lifecycle • Advantages of generative AI (for example, adaptability, responsiveness, simplicity). • Disadvantages of generative AI solutions (for example, hallucinations, interpretability, inaccuracy, nondeterminism). • Retrieval Augmented Generation (RAG) • Recall-Oriented Understudy for Gisting Evaluation [ROUGE] • Bilingual Evaluation Understudy [BLEU] • BERTScore • Features of responsible AI • Amazon Transcribe	• Amazon SageMaker JumpStart • Amazon SageMaker Clarify • Amazon SageMaker Data Wrangler • Amazon SageMaker Model Monitor • Amazon SageMaker Feature Store • Amazon Augmented AI [Amazon A2I] • Amazon SageMaker Model Cards • Amazon Macie • Data cataloging • International Organization for Standardization [ISO] • System and Organization Controls [SOC] • Data lifecycles • AWS Support plans • AWS Well-Architected Framework

Remember that out of the 5 exam domains, the **Applications of Foundation Models** domain has the biggest coverage in the exam, at 28 percent. This means that more than a quarter of the questions in the entire AWS Certified AI Practitioner exam cover the applications of foundation models.

The Appendix section of the exam guide also includes a list of relevant AWS services that you should focus on, so in your exam, make sure that you review the following AWS services.

When it comes to Analytics, ensure you study AWS Data Exchange, Amazon EMR, AWS Glue, AWS Glue DataBrew, AWS Lake Formation, Amazon OpenSearch Service, Amazon QuickSight, and Amazon Redshift.

For Cloud Financial Management, Compute, and Containers, you should learn about AWS Budgets, AWS Cost Explorer, Amazon EC2, Amazon Elastic Container Service (Amazon ECS), and Amazon Elastic Kubernetes Service (Amazon EKS).

In the Databases category, focus on Amazon DocumentDB (with MongoDB compatibility), Amazon DynamoDB, Amazon ElastiCache, Amazon MemoryDB, Amazon Neptune, and Amazon RDS.

The Machine Learning section includes essential services such as 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 AI, Amazon Textract, Amazon Transcribe, and Amazon Translate.

The AI Practitioner exam covers a handful of services related to Management and Governance, including AWS CloudTrail, Amazon CloudWatch, AWS Config, AWS Trusted Advisor, and AWS Well-Architected Tool.

For the Security, Identity, and Compliance category, prepare to see a range of AWS services that you can use to secure your enterprise applications and AWS resources. These include AWS Artifact, AWS Audit Manager, AWS Identity and Access Management (IAM), Amazon Inspector, AWS Key Management Service (AWS KMS), Amazon Macie, and AWS Secrets Manager. Pay attention to how these services work together and know the appropriate AWS service to use for a particular use cases or situation.

Lastly, for Networking and Content Delivery, make sure to understand Amazon CloudFront and Amazon VPC. For Storage, review Amazon S3 and Amazon S3 Glacier.

Exam Scoring System

You can get a score from 100 to 1,000 with a minimum passing score of **700** when you take the AWS Certified AI Practitioner exam. AWS uses a scaled scoring model to associate scores across multiple exam types that may have different levels of difficulty. Your complete score report will be sent to you by email 1 - 5 business days after your exam.

For individuals who unfortunately do not pass their exams, you must wait 14 days before you are allowed to retake the exam. There is no hard limit on the number of attempts you can retake an exam. Once you pass, you'll receive various benefits such as a discount coupon which you can use for your next AWS exam.

Once you receive your score report via email, the result should also be saved in your AWS Certification account already. The score report contains a table of your performance on each domain and it will indicate whether you have met the level of competency required for these domains. Take note that you do not need to achieve competency in all domains for you to pass the exam. At the end of the report, there will be a score performance table that highlights your strengths and weaknesses which will help you determine the areas you need to improve on.

		Score Performance	
Section	**% of Scored Items**	**Needs Improvement**	**Meets Competencies**
Domain 1: Fundamentals of AI and ML	20%		
Domain 2: Fundamentals of Generative AI	24%		
Domain 3: Applications of Foundation Models	28%		
Domain 4: Guidelines for Responsible AI	14%		
Domain 5: Security, Compliance, and Governance for AI Solutions	14%		

Exam Benefits

If you successfully passed any AWS exam, you will be eligible for the following benefits:

- **Exam Discount** - You'll get a 50% discount voucher that you can apply for your recertification or any other exam you plan to pursue. To access your discount voucher code, go to the "Benefits" section of your AWS Certification Account, and apply the voucher when you register for your next exam.

- **Certification Digital Badges** - You can showcase your achievements to your colleagues and employers with digital badges on your email signatures, Linkedin profile, or on your social media accounts. You can also show your Digital Badge to gain exclusive access to Certification Lounges at AWS re:Invent, regional Appreciation Receptions, and select AWS Summit events. To view your badges, simply go to the "Digital Badges" section of your AWS Certification Account.

You can visit the official AWS Certification FAQ page to view the frequently asked questions about getting AWS Certified and other information about the AWS Certification: https://aws.amazon.com/certification/faqs/.

AWS CERTIFIED AI PRACTITIONER (AIF-C01) EXAM STUDY GUIDE

The AWS Certified AI Practitioner or AWS AIF-C01 exam is designed to assess the candidate's understanding of AI and machine learning concepts within the AWS environment. This certification covers most, if not all, fundamental knowledge that one should know when venturing into the Cloud, together with AI/ML applications. The AWS AIF-C01 course intends to provide practitioners with a fundamental understanding of the AWS Cloud without having to dive deep into the technicalities. This includes the AWS Global Infrastructure, best practices in using AWS Cloud, pricing models, technical support options, and many more. You can view the complete details and guidelines for the certification exam here.

In addition to core cloud concepts, the exam emphasizes AI/ML services and solutions available through AWS, such as Amazon SageMaker AI, Rekognition, and Polly. It also touches on the basics of machine learning workflows, from data collection to model deployment and monitoring. A candidate will be tested on their ability to apply AI/ML models in real-world scenarios, leveraging AWS tools to solve business challenges. While no deep programming expertise is required, a solid understanding of machine learning principles, data handling, and AWS services is essential to passing the exam. This certification is ideal for those aiming to advance their careers in AI, machine learning, or cloud computing while learning about AWS's vast array of AI solutions.

What to review

1. AWS AI and ML Services

Familiarize yourself with the wide range of AWS services designed for artificial intelligence (AI) and machine learning (ML). Understand their key features and use cases, and learn how these services can help you in building, deploying, and managing AI and ML solutions efficiently. The available services include tools for data preparation, model building, training, deployment, and inference. Additionally, understanding the AWS Cloud and its various services is crucial. This includes:

- Core AWS services like Amazon EC2, Amazon S3, AWS Lambda, and Amazon SageMaker AI, and their operational applications.
- The AWS shared responsibility model for security and compliance.
- AWS Identity and Access Management (AWS IAM) for managing security and access to AWS resources.
- The AWS global infrastructure, including the concepts of AWS Regions, Availability Zones, and edge locations.
- AWS services pricing models.

To quickly view over the different categories, you may visit this link. For a comprehensive introduction, this AWS whitepaper contains an overview of the different AWS services along with their definitions and use cases.

2. Fundamentals of Artificial Intelligence (AI) and Machine Learning (ML)

Grasp the foundational concepts of artificial intelligence and machine learning. Understand the differences between supervised, unsupervised, and reinforcement learning, as well as common algorithms. Learn how these technologies are applied to solve various problems across different domains.

3. Prompt Engineering

Delve into the art of prompt engineering to effectively customize AI models. Learn about different types of prompts and how to use prompt templates to enhance model responses without changing the underlying model architecture.

4. Data Wrangling and Preprocessing

Understand the importance of data preparation in the machine learning pipeline. Discover how AWS services can simplify and automate data wrangling tasks to improve the quality and reliability of your machine learning models.

5. Model Training and Evaluation

Gain insights into training machine learning models, selecting appropriate algorithms, tuning hyperparameters, and evaluating model performance. Learn key metrics like accuracy, precision, recall, and F1 score to evaluate your models' performance.

6. Deployment and Operationalization

Study the process of deploying machine learning models into production environments. Learn about best practices for monitoring, maintaining, and scaling deployed models to ensure they perform well over time.

7. Ethical and Responsible AI

Understand the ethical considerations in AI, such as fairness, transparency, and bias detection. Learn best practices for implementing AI solutions responsibly to mitigate ethical risks.

8. Real-World Applications

Examine real-world case studies and examples of how various industries are leveraging AWS AI and ML technologies. This will help you understand practical applications and the impact of these technologies in different sectors.

How to review

As with any exam, the very first step is always the same - **KNOWING WHAT TO STUDY**. Although we have already enumerated them in the previous section, I highly suggest you go over the AWS Certified AI Practioner AIF-C01 Exam Guide again and see the exam contents.

AWS already has a vast number of free resources available for you to prepare for the exam. I suggest you first read Overview of Amazon Web Services whitepaper, and gain a good understanding of the different AWS concepts and services. Check out the amazing Tutorials Dojo cheat sheets to supplement your review for this section. Also check out this article: **Top 5 FREE AWS Review Materials**.

Beyond understanding the core AWS services, it's essential to dive into AWS's AI/ML offerings. Familiarize yourself with services like Amazon SageMaker AI, Amazon Bedrock, Amazon Rekognition, Amazon Lex, Amazon Polly, and many more, as these are crucial for the exam.

AWS also offers a comprehensive digital course on Machine Learning, consisting of a collection of free courses designed to enhance your understanding and skills. This collection includes:

- AWS Machine Learning Services Overview
- AWS DeepRacer
- Data Science
- Math for Machine Learning
- MLS-C01 Exam Readiness

These courses not only cover the MLS-C01 exam topics but also provide valuable insights for the AWS Certified Machine Learning Engineer – Associate MLA-C01 and AWS Certified AI Practitioner AIF-C01 certifications.

Common Exam Scenarios

Scenario	Solution
Domain 1: Fundamentals of AI and ML	
A branch of computer science focused on addressing cognitive challenges typically linked to human intelligence.	Artificial Intelligence
It is a type of AI focused on developing methods that enable machines to learn and understand.	Machine Learning
It is an AI technique that enables computers to process data in ways that mimic the human brain.	Deep Learning

It is a technique of artificial intelligence that teaches computers to interpret data in a way inspired by the human brain.	Neural Networks
It allows machines to identify people, places, and objects in photos with accuracy similar to humans, all while operating much faster and more efficiently.	Computer Vision
.Analyzing data sets to summarize their main characteristics, frequently using visualizations to reveal patterns, trends, and relationships.	Exploratory Data Analysis (EDA)
It is a branch of artificial intelligence that focuses on how computers interact with human languages.	Natural Language Processing (NLP)
A performance metric for classification models that indicates the model's effectiveness in distinguishing between classes at different thresholds.	Area Under the ROC Curve (AUC)
Domain 2: Fundamentals of Generative AI	
The process of guiding generative artificial intelligence (generative AI) to produce specific outputs.	Prompt Engineering
These models are created to handle inputs from multiple sources, such as text, images, audio, and video.	Multi-modal Models
It involves training an already established model on a new dataset instead of starting from the beginning. This technique, known as transfer learning, can produce reliable models from smaller datasets and requires less training time.	Fine-tuning
A numerical representation of real-world objects used by machine learning (ML) and artificial intelligence (AI) systems to comprehend complex knowledge domains like humans do.	Embedding
Generative AI is capable of adapting to various activities and domains by learning from data and creating content tailored to specific situations or needs. Its flexibility allows it to be applied across a wide range of sectors.	Adaptability

It generates content in real-time, leading to quicker responses and more dynamic interactions. This is particularly advantageous for chatbots, virtual assistants, and other interactive applications that require immediate feedback.	Responsiveness
AI language models can simplify challenging tasks by automating content generation processes. For instance, they can produce text that resembles human writing, thus reducing the time and effort required for content development.	Simplicity
With JumpStart, you can quickly evaluate, compare, and choose FMs for tasks like summarizing articles and creating images based on established quality and accountability criteria.	Amazon SageMaker JumpStart
It is a fully managed solution that provides access to high-performing foundation models (FMs) from leading AI startups and Amazon through a common API.	Amazon Bedrock
It is an Amazon Bedrock Playground that enables users to easily and intuitively build generative AI applications. The platform offers a fun, hands-on environment where users can create a variety of AI-driven applications in just a few steps.	PartyRock
Domain 3: Application of Foundation Models	
It is a technique used to help a model generalize from a few examples. The model leverages these examples to make more accurate predictions without the need for re-training or fine-tuning.	Few-shot Prompt Engineering
It is a method for customizing a pre-trained FM by fine-tuning the model on a specific task or domain-specific information.	Domain adaptation fine-tuning
A technique that utilizes labeled examples to enhance the performance of a model for a specific task.	Instruction-based fine-tuning
One continuously improves their model by analyzing feedback from earlier versions. In reinforcement	Reinforcement learning

learning, an agent learns through trial and error while interacting with its environment.	
It is a method to optimize the output of a large language model (LLM) by referencing a knowledge base containing company-specific or industry-specific data.	Retrieval Augmented Generation
A managed service for search, monitoring, and data analysis that provides real-time search and analytics for various applications.	Amazon OpenSearch Service
A relational database compatible with MySQL and PostgreSQL, designed for high performance and availability while supporting complex applications.	Amazon Aurora
A fully managed graph database service designed for efficient storage and querying of highly connected datasets, making it perfect for graph-based applications.	Amazon Neptune
In generative models, these prompts indicate which content to exclude from the generated output.	Negative prompts
The conceptual space where machine learning models transform input data into feature representations that are used to generate outputs.	Model Latent Space
It is a technique that uses human feedback to help machine learning models make predictions more efficiently and accurately while maximizing rewards.	Reinforcement Learning from Human Feedback (RLHF)
A metric for assessing text summarization quality by comparing the overlap between produced and reference summaries.	Recall-Oriented Understudy for Gisting Evaluation (ROUGE)
A metric used to evaluate the quality of machine-translated text by measuring the similarity between the machine output and human reference translations.	Bilingual Evaluation Understudy (BLEU)
A metric for assessing text generation models by comparing token-level similarities using BERT embeddings.	BERTScore

Domain 4: Guidelines for Responsible AI	
It outlines the procedures and principles that ensure AI systems are transparent and trustworthy, while also minimizing potential risks and negative effects.	Responsible AI
A tool for machine learning that offers insights to enhance model fairness and transparency by analyzing potential biases in datasets and model predictions.	Amazon SageMaker Clarify
A service that simplifies incorporating human review into machine learning predictions, ensuring high-quality outcomes.	Amazon Augmented AI (Amazon A2I)
Documents that describe important details about machine learning models, such as performance metrics, intended uses, and compliance information.	Amazon SageMaker Model Cards
Domain 5: Security, Compliance, and Governance for AI Solutions	
A fully managed data security and privacy solution utilizing machine learning to identify, categorize, and safeguard sensitive data in AWS.	Amazon Macie
A service that offers secure, private connectivity between VPCs and AWS services, ensuring scalable access to critical resources.	AWS PrivateLink
A security vulnerability occurs when malicious input is used in prompts to alter the output of language models.	Prompt Injection
A global organization that sets standards and provides guidelines for various industries to ensure quality, safety, and efficiency.	International Organization for Standardization (ISO)
A collection of reports that provide details about the controls at a service organization, related to security, availability, processing integrity, confidentiality, or privacy.	System and Organization Controls (SOC)
A framework for identifying and managing security risks of generative AI models.	Generative AI Security Scoping Matrix

Validate Your Knowledge

When you are feeling confident with your review, it is best to validate your knowledge through sample exams. **Tutorials Dojo** offers a very useful and well-reviewed set of practice tests for the AI Practitioner exam takers **here** to help you prepare well. Each test contains many unique questions which will surely help you verify if you have missed out on anything important that might appear on your exam. You can pair these practice exams with this study guide eBook for comprehensive preparation.

If you have scored well on the **Tutorials Dojo AWS Certified AI Practitioner Practice Tests** and you think you are ready, then go earn your certification with your head held high. However, if you find certain areas challenging, take the time to review them again and pay attention to any hints in the questions that can guide you to the correct answers. If you're not entirely confident about passing, consider rescheduling your exam to allow yourself more preparation time. In the end, the efforts you invest will undoubtedly pay off.

Sample Practice Test Questions:

Question 1

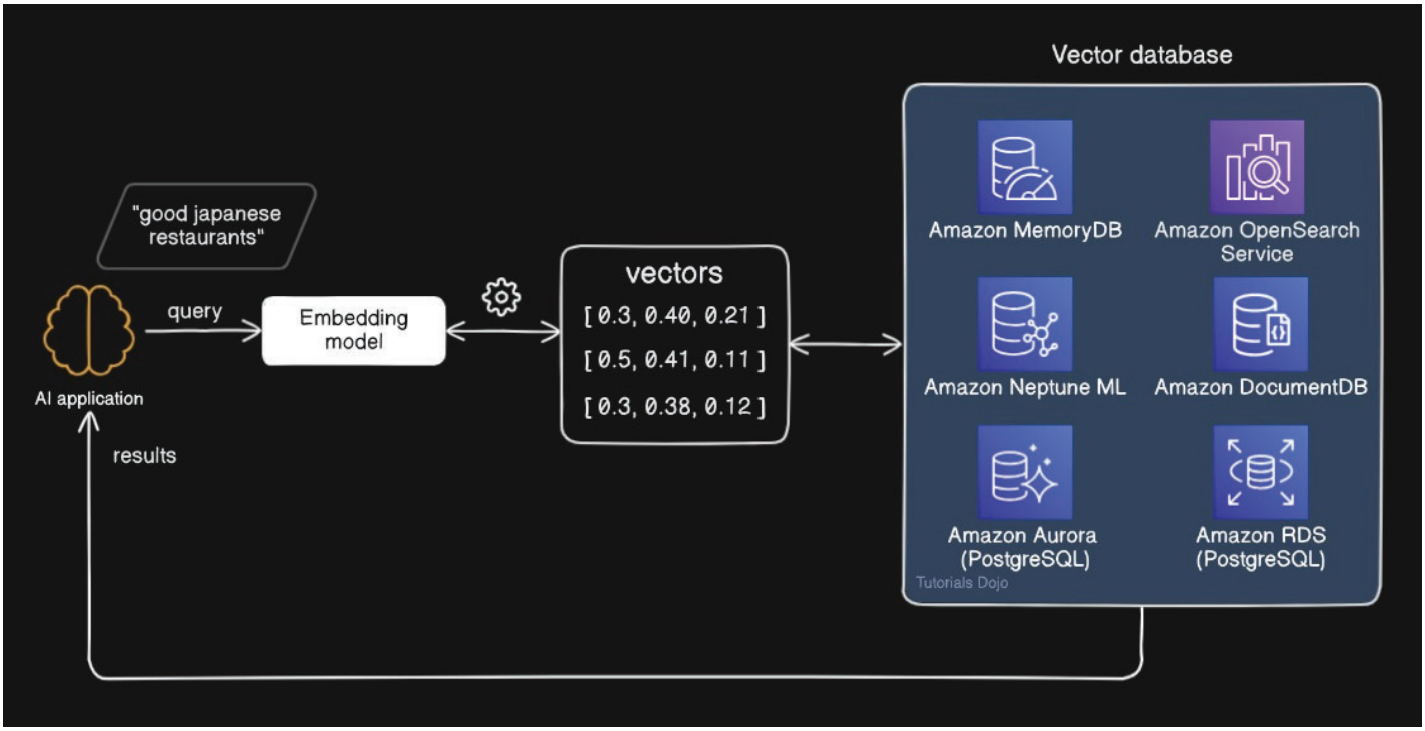
A company has a recommender system that generates embeddings from customer interaction data. They want to improve the speed and efficiency of retrieving similar product recommendations.

Which AWS services are best suited for implementing vector search to optimize the recommendation system? (Select THREE).

1. Amazon OpenSearch Service
2. Amazon Neptune ML
3. Amazon S3
4. Amazon DocumentDB (with MongoDB compatibility)
5. Amazon Redshift
6. Amazon Quicksight

Correct Answer: 1,2,4

Vector search is a method used in machine learning to find similar data points to a given data point by comparing their vector representations using distance or similarity metrics. The closer the two vectors are in the vector space, the more similar the underlying items are considered to be. This technique helps capture the semantic meaning of the data. This approach is useful in various applications, such as recommendation systems, natural language processing, and image recognition.



A vector database, specifically designed to handle vector representations of data efficiently, excels in storing, indexing, and retrieving vectors compared to traditional databases. Traditional databases typically handle scalar values and are optimized for transactional operations, which can be limiting for AI-driven queries that require rapid computation of vector similarities. In contrast, vector databases use specialized indexing algorithms, such as approximate nearest neighbor (ANN) search, which significantly speeds up query times and maintains high accuracy, even in large datasets. This makes them particularly advantageous for AI applications, where quick and precise retrieval of similar items based on complex, high-dimensional data is crucial. These capabilities allow for more dynamic and responsive AI systems, such as real-time personalized recommendation engines and instant image or voice recognition services.

Here are some services in AWS that you can use for your vector database requirements:

- **Amazon OpenSearch Service** makes it easy for you to perform interactive log analytics, real-time application monitoring, website search, and more. For vector databases, you can read about k-Nearest Neighbor (k-NN) search in OpenSearch Service.
- **Amazon Aurora PostgreSQL-Compatible Edition** and **Amazon Relational Database Service (Amazon RDS)** for PostgreSQL support the pgvector extension to store embeddings from machine learning (ML) models in your database and to perform efficient similarity searches.
- **Amazon Neptune ML** is a capability of Neptune that uses Graph Neural Networks (GNNs), an ML technique purpose-built for graphs, to make easy, fast, and more accurate predictions using graph data.
- **Amazon MemoryDB** supports storing millions of vectors, with single-digit millisecond query and update response times, and tens of thousands of queries per second (QPS) at greater than 99% recall.

- **Amazon DocumentDB (with MongoDB compatibility)** supports vector search, a new capability that enables you to store, index, and search millions of vectors with millisecond response times. With vector search for Amazon DocumentDB, you can simply set up, operate, and scale databases for your ML applications.

Hence, the correct answers are:

- Amazon OpenSearch Service
- Amazon Neptune ML
- Amazon DocumentDB (with MongoDB compatibility)

The option that says: **Amazon S3** is incorrect. This service is primarily an object storage service designed for storing and retrieving large amounts of data. While it can store vectors as files, it lacks built-in capabilities for indexing and searching vectors efficiently. For vector search, you need a service that supports real-time querying and similarity search, which S3 does not provide.

The option that says: **Amazon Redshift** is incorrect because this is mainly a data warehouse service designed for running complex queries on large datasets. Although it may be possible to implement vector search using custom User Defined Functions (UDFs), this approach would involve additional complexity and introduce potentially higher query latency.

The option that says: **Amazon Quicksight** is incorrect. This service is simply a business intelligence and visualization service for creating dashboards and reports. It is not designed for vector search or indexing.

References:
https://aws.amazon.com/what-is/vector-databases/
https://docs.aws.amazon.com/neptune-analytics/latest/userguide/vector-similarity.html
https://docs.aws.amazon.com/documentdb/latest/developerguide/vector-search.html
https://docs.aws.amazon.com/opensearch-service/latest/developerguide/serverless-vector-search.html

Check out these cheat sheets for Amazon OpenSearch Service, Amazon Neptune, and Amazon DocumenDB:
https://tutorialsdojo.com/amazon-neptune/
https://tutorialsdojo.com/amazon-documentdb/
https://tutorialsdojo.com/amazon-opensearch-service/

Question 2

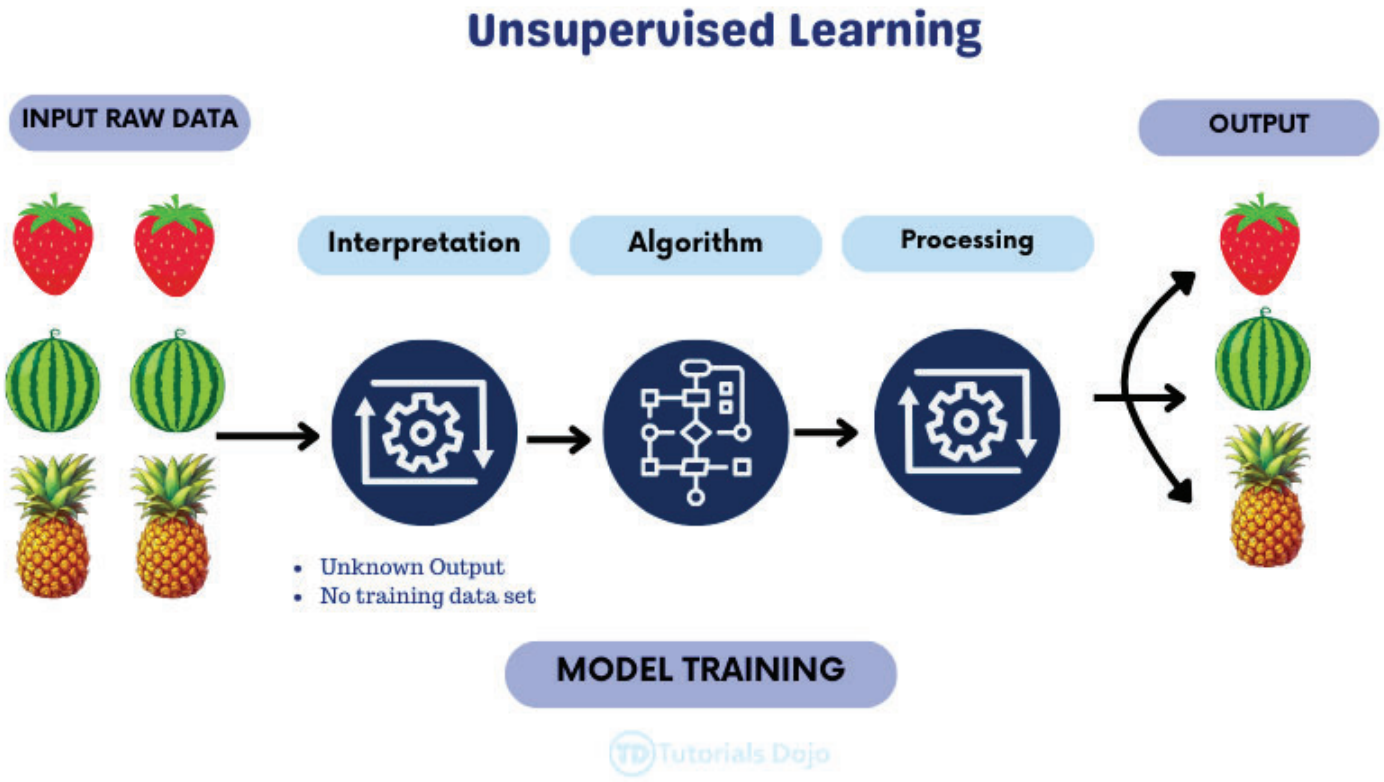
Which machine learning approach is used to classify and organize unlabeled data by identifying hidden patterns without requiring predefined categories or labels?

1. Transfer Learning
2. Reinforcement Learning

3. Few-shot Learning
4. Unsupervised Learning

Correct Answers: 4

Unsupervised learning is a machine learning technique used to analyze and cluster unlabeled data. Unlike supervised learning, where models are trained with labeled datasets, unsupervised learning algorithms operate without any predefined categories or labels. These algorithms discover hidden patterns, structures, and relationships in the data. One of the most common applications of unsupervised learning is clustering, where data points with similar characteristics are grouped. Another use case includes dimensionality reduction, which simplifies data by reducing the number of variables while maintaining its important features. This is highly effective when handling large datasets.



Amazon SageMaker AI supports unsupervised learning for various tasks, such as clustering, anomaly detection, and association rule learning. Amazon SageMaker AI provides pre-built algorithms such as K-Means and Principal Component Analysis (PCA) that can be used to analyze unlabeled datasets. By utilizing SageMaker's built-in unsupervised learning capabilities, users can quickly build and deploy models for tasks like customer segmentation, recommendation engines, or detecting anomalies in system logs. SageMaker streamlines the entire machine learning workflow, from data preparation to deployment, making it easier to leverage unsupervised learning for real-world business needs.

Hence, the correct answer is: **Unsupervised Learning.**

The option that says: **Transfer Learning** is incorrect because this type of learning uses a pre-trained model from one task or domain and applies it to a different but related task. Transfer learning is typically used when there is a shortage of labeled data in the target domain but an abundance of labeled data in a related domain. This approach is not designed for working with unlabeled data or discovering hidden patterns, making it unsuitable for the scenario described.

The option that says: **Reinforcement Learning** is incorrect because this primarily focuses on learning through an agent's interactions with an environment. In reinforcement learning, the agent receives feedback in the form of rewards or penalties based on its actions and learns to maximize cumulative rewards over time. This approach is not meant for classifying or grouping unlabeled data, but rather for optimizing decision-making processes in dynamic environments.

The option that says: **Few-shot Learning** is incorrect because it involves training models with a very limited amount of labeled data, often just a few examples. Few-shot learning is particularly useful when labeled data is scarce, but it still requires some level of labeled data to function. It does not apply to situations with no labeled data, as is the case with unsupervised learning.

References:
https://docs.aws.amazon.com/sagemaker/latest/dg/algorithms-unsupervised.html
https://aws.amazon.com/compare/the-difference-between-machine-learning-supervised-and-unsupervised/

Check out this Amazon SageMaker AI Cheat Sheet:
https://tutorialsdojo.com/amazon-sagemaker/