

1. Record Nr.	UNINA9911034867603321
Autore	Elango Vikram
Titolo	AWS Certified AI Practitioner Study Guide : Foundational (AIF-C01) Exam
Pubbl/distr/stampa	Newark : , : John Wiley & Sons, Incorporated, , 2025 ©2026
ISBN	1-394-32821-4 1-394-40669-X 1-394-32820-6
Edizione	[1st ed.]
Descrizione fisica	1 online resource (322 pages)
Collana	Sybex Study Guide Series
Altri autori (Persone)	GangasaniVivek SubramanianShreyas
Disciplina	006.3
Soggetti	Artificial intelligence - Examinations
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Cover -- Half Title Page -- Title Page -- Copyright -- Contents at a Glance -- Contents -- Introduction -- Assessment Test -- Answers to Assessment Test -- Part I: Introduction to AI and ML -- Chapter 1: Basic AI Concepts and Terminology -- A Brief History of AI -- Diving Deeper into Terms You Should Know -- The Learning Paradigms: Supervised, Unsupervised, and Reinforcement Learning -- When to Use the Different Types of Learning -- The Deep Learning Revolution -- Training: Learning Model Parameters Values -- Feature Engineering and Data Preprocessing -- Evolution of Specialized Architectures for Complex Data -- The Transformer Revolution -- Generative AI -- The Relationship Among AI, ML, and Deep Learning -- Hierarchical Relationship -- Artificial Intelligence -- Machine Learning -- Deep Learning -- Similarities Among AI, ML, and DL -- Differences Among AI, ML, and DL -- Understanding Data Types in AI Models -- Labeled vs. Unlabeled Data -- Structured Data -- Tabular Data -- Time-Series Data -- Log Data -- Unstructured Data -- Text Data -- Image Data -- Video Data -- Audio Data -- Making Predictions Using Trained Models -- Batch Inference -- Real-Time Inference -- Asynchronous Inference -- Summary -- Exam Essentials -- Review Questions -- Chapter 2:

Basic Concepts of Generative AI -- A New Way to Interact with AI -- From Text to Numbers: Tokens, Chunking, and Embeddings -- The Transformer Architecture and Foundation Models -- From Embeddings to Attention -- Beyond Attention: Why Transformers Are So Powerful -- Beyond Text: Multi-modal Models -- Diffusion Models -- Key Use Cases for Multi-modal AI -- Prompt Engineering -- The Upsides and Downsides of Gen AI -- Summary -- Exam Essentials -- Review Questions -- Part II: Building AI Applications with AWS -- Chapter 3: Applications of AI and ML in Real-World Use Cases. Key Trends in AI and ML Applications -- Automation of Repetitive Tasks -- Predictive Analytics and Forecasting -- Personalization -- Enhanced Decision-Making -- Cost Optimization -- Use Cases Unsuitable for AI and ML Applications -- Choosing the Right ML Techniques for Different Use Cases -- Regression -- Data Labeling Requirements for Regression -- Regression Metrics to Evaluate -- Mapping Domain Use Cases to Regression -- Classification -- Data Labeling Requirements for Classification -- Classification Metrics to Evaluate -- Mapping Domain Use Cases to Classification -- Clustering -- Mapping Domain Use Cases to Clustering -- Clustering Metrics to Evaluate -- Use Cases and Applications for Deep Learning Algorithms -- High-Level Approach for Deep Learning Workflows -- Computer Vision (CV) Use Cases -- Natural Language Processing (NLP) Use Cases -- Generative AI Use Cases -- Consumer-Focused Applications -- Enterprise Applications -- Summary -- Exam Essentials -- Review Questions -- Chapter 4: AWS AI and ML Services -- An Overview of AWS Managed AI and ML Services -- Amazon Generative AI Service -- Amazon Bedrock -- Choice of Models -- Bedrock Playground -- Agents for Amazon Bedrock -- Amazon Bedrock Knowledge Bases -- Amazon Bedrock Data Automation -- Model Customization -- Foundation Model Evaluation -- Guardrails -- Pricing -- Example Use Case for Amazon Bedrock -- Amazon Q -- Amazon Q for Business -- Amazon Q for Developers -- AWS PartyRock -- AWS AI Services -- Amazon Comprehend -- Amazon Textract -- Amazon Transcribe -- Amazon Translate -- AWS ML Services -- Amazon SageMaker AI -- Amazon SageMaker Studio -- SageMaker JumpStart -- SageMaker Canvas -- SageMaker Studio Lab -- Pricing -- Summary -- Exam Essentials -- Review Questions -- Part III: Common GenAI Patterns -- Chapter 5: Model Selection and Prompt Engineering. Selecting the Right Foundation Model for Your Use Case -- Decision Framework for Selecting Foundation Models -- Model Cards and Documentation -- Modality Support and Integration -- Multilingual Capabilities -- Cost-Performance Ratio -- Latency, Infrastructure, and Scale -- Model Size and Performance Benchmarks -- The Effect of Inference Parameters on Model Responses -- Temperature -- Maximum Output Length -- Top-k Sampling -- Top-p (Nucleus) Sampling -- Stop Words and Stop Sequences -- Prompt Engineering -- Prompt Engineering Fundamentals -- Providing Clear Instructions -- Constraining Outputs -- Role-Playing -- Use Examples in Context -- Model-Specific Considerations -- Handling Long Context -- Thinking Step-by-Step -- Cost-Performance Trade-offs -- Understanding Risks and Limitations in Prompt Engineering -- Summary -- Exam Essentials -- Review Questions -- Chapter 6: Generative AI Applications with RAG and Agents -- Retrieval-Augmented Generation Workflow -- The Data Ingestion Phase -- The Retrieval and Generation Phase -- Amazon Bedrock Knowledge Bases -- Amazon Bedrock Data Automation -- How Amazon Knowledge Bases Work -- Knowledge Bases Data Ingestion Workflow -- Connecting to a Data Source -- Unstructured Document Parsing Strategy -- Chunking -- Embedding Model -- Vector Store --

Knowledge Base Retrieval and Generation Workflow -- Knowledge Base Retrieve API -- Knowledge Bases RetrieveAndGenerate API -- Guardrails with Knowledge Bases -- Evaluating RAG Workflows on Amazon Bedrock -- Amazon Bedrock Agents -- Amazon Bedrock Agents Components -- Agents in Action -- Preprocessing Prompt -- Orchestration Step -- Post-processing -- Multi-agent Collaboration -- Portfolio Assistant Agent -- Stock Data Researcher Agent -- Stock News Researcher Agent -- Financial Analyst Agent -- Summary -- Exam Essentials -- Review Questions.

Chapter 7: Model Customization and Evaluation -- Overview of Customization Techniques -- Pre-training Models: Building the Foundation -- Self-supervised Learning -- Data Selection and Collection -- Steps Involved in Pre-training a Model -- GPU Memory Considerations -- Floating-point Precision and Mixed-precision Training -- Distributed Training Frameworks -- Fine-tuning -- Continuous Pre-training -- PEFT and LoRA Fine-tuning -- AWS Services for Pre-training and Fine-tuning -- Amazon SageMaker AI -- SageMaker Training Jobs -- SageMaker HyperPod -- HyperPod Training Recipes -- Amazon Bedrock -- Data Processing -- Model Evaluation -- Model Evaluation for Foundation Models -- Evaluating Models for Business Objectives -- Summary -- Exam Essentials -- Review Questions -- Part IV: Bringing AI to Production -- Chapter 8: MLOps -- MLOps Phases -- Experimentation -- Repeatable Processes -- Building Scalable Systems -- Deploying to Production -- Model Monitoring -- MLOps Pipeline -- Data Collection -- Data Preprocessing -- Model Training -- Model Evaluation -- Model Deployment -- Model Monitoring -- Automating MLOps -- SageMaker Pipelines -- Fully Managed MLFlow on SageMaker -- AWS Step Functions for Orchestration -- Apache Airflow for Workflow Scheduling -- Continuous Integration and Continuous Delivery (CI/CD -- Infrastructure as Code (IaC) for MLOps -- Integrating CI/CD and IaC with SageMaker Pipelines and Step Functions -- Reducing Technical Debt with CI/CD, Pipelines, and IaC -- SageMaker Inference -- Advanced Deployment Scenarios -- Inference Optimizations for Large Language Models -- Model Distillation -- Quantization -- Pruning -- Tensor and Expert Parallelism -- Speculative Decoding -- Flash Attention -- Paged Attention -- Summary -- Exam Essentials -- Review Questions -- Chapter 9: Implementing Responsible AI with AWS Services.

Key Principles of Responsible AI -- ML Governance with SageMaker AI -- Amazon SageMaker Model Cards -- Amazon SageMaker Model Registry -- Amazon SageMaker Pipelines and Experiments -- Amazon SageMaker Clarify -- Evaluating Foundation Models Using SageMaker Clarify -- Model Quality Checks -- Data Quality Checks -- SageMaker Model Monitoring -- Amazon Bedrock Guardrails -- The ApplyGuardrail API -- Guardrail Policies -- Amazon Bedrock Evaluations -- Automated Model Evaluation -- LLM-as-a-Judge Evaluations -- Bedrock Knowledge Bases and RAG Evaluations -- Summary -- Exam Essentials -- Review Questions -- Chapter 10: AI Security, Governance, and Compliance -- Security of AI Systems -- Adversarial Tactics Against AI Systems -- An Overview of Adversarial Techniques for AI -- Phase 1: Reconnaissance Techniques: Gathering Intel -- Phase 2: Resource Acquisition Techniques: Building the Toolkit -- Phase 3: Execution Techniques: Crafting and Launching Attacks -- Advanced Tactics: Exploiting Large Language Models (LLMs) -- Prompt Injection -- LLM Jailbreaking -- Defensive Strategies: Protecting Against Adversarial Techniques -- The Generative AI Security Scoping Matrix -- Scope 1: Consumer Applications Utilizing Public Generative AI

Services -- Scope 2: Enterprise Applications Incorporating Generative AI Features -- Scope 3: Applications Built on Pre-Trained Models -- Scope 4: Fine-Tuned Models Customized with Organizational Data -- Scope 5: Self-Trained Models Developed from Scratch -- Data Governance Strategies -- Data Lifecycles in AI Systems -- Data Logging and Documentation -- Data Residency and Sovereignty Considerations -- Data Retention Policies for Generative AI Applications -- Understanding Data Retention in Generative AI -- Implementing Effective Data Retention Policies -- Data Retention on Amazon Bedrock -- Compliance and Regulatory Frameworks in AI. Compliance Requirements for AI Systems on AWS.

Sommario/riassunto

Quickly and intelligently prepare for the AIF-C01 exam and succeed in your first role as an AWS AI practitioner In AWS Certified AI Practitioner Study Guide: Foundational (AIF-C01) Exam, a team of veteran AWS and AI specialists walks you through an efficient and effective path to success on the challenging AIF-C01 exam.
