

Databricks Generative AI Engineer Associate Training

OEM: Apache • Duration: 3 Days (24 hrs) • Code: DBX-GENAI-A

COURSE MODULES & TOPICS

Section 1: Design Applications (14%)

- Design prompts that elicit specifically formatted responses
- Select model tasks appropriate for given business requirements
- Choose chain components for desired inputs and outputs
- Convert business goals into AI pipeline descriptions
- Define and sequence tools for multi-stage reasoning pipelines
- Apply Agent Bricks: Knowledge Assistant, Multiagent Supervisor, Information Extraction

Section 2: Data Preparation (14%)

- Apply chunking strategies based on document structure and model constraints
- Filter extraneous content from source documents
- Select appropriate Python packages for content extraction
- Write chunked text into Delta Lake tables in Unity Catalog
- Identify appropriate source documents for RAG applications
- Evaluate retrieval performance using tools and metrics
- Design advanced retrieval systems and understand re-ranking

Section 3: Application Development (30%)

- Select LangChain and similar orchestration tools
- Assess response quality and safety issues
- Augment prompts with contextual information (RAG)
- Implement LLM guardrails against unsafe outputs
- Select appropriate LLMs and embedding models from model hubs
- Utilize MLflow and Agent Framework for experiment tracking
- Compare evaluation vs. monitoring phases
- Enable multi-agent systems via Genie Spaces

Section 4: Assembling and Deploying Applications (22%)

- Code a chain using pyfunc model with pre- and post-processing
- Control resource access from model serving endpoints

- Create RAG applications with all necessary components
- Register models to Unity Catalog
- Create and query Vector Search indexes
- Serve LLM applications via Foundation Model APIs
- Apply CI/CD best practices for GenAI applications
- Integrate MCP servers; manage prompt lifecycle and version control

Section 5: Governance (8%)

- Use masking techniques as guardrails
- Select guardrail techniques against malicious inputs
- Address legal and licensing requirements for data sources
- Mitigate problematic text in source data

Section 6: Evaluation and Monitoring (12%)

- Select LLMs based on quantitative metrics
- Identify key monitoring metrics for deployed GenAI applications
- Evaluate agent performance via MLflow
- Use inference logging for RAG assessment
- Control LLM costs with Databricks features (AI Gateway, rate limiting)
- Apply custom Scorers for evaluation; incorporate SME feedback