

Agentic AI for Data Engineering & Data/AI Platform Engineering

Duration: 5 Days

Description

Training Level: Advanced The use of artificial intelligence technologies in enterprise data platforms is evolving beyond traditional data processing and analytics applications toward AI agents that can interact with different systems and support complex data engineering processes. Agentic AI architectures bring together large language models (LLMs), enterprise data sources, API services, databases, and tools, offering new opportunities for data discovery, data quality control, metadata management, data lineage analysis, and the automation of operational processes. This training aims to enable participants to gain competence in designing, developing, orchestrating, integrating, evaluating, and deploying AI agents that can be used in enterprise data platforms into production environments. The program is structured by taking into account the data security, authorization, auditability, and operational continuity requirements of the banking sector. To develop participants' skills in designing, developing, managing, and securely operating AI agents that can be used in data engineering and enterprise Data/AI Platform Engineering processes in production environments. **Delegates will learn how to**

- Design Agentic AI architectures
- Develop Tool Calling and MCP integrations
- Create single-agent and multi-agent systems
- Integrate RAG and agent memory mechanisms
- Develop agents that interact with databases and API services
- Design agent-based solutions for data discovery, data quality, and metadata management
- Implement Human-in-the-Loop control mechanisms
- Evaluate and monitor agent performance
- Implement enterprise security, authorization, and data access policies
- Deploy Agentic AI applications to production environments

Audience

- Data Engineer / Senior Data Engineer
- Data Platform Engineer
- AI Platform Engineer

- ML Engineer / MLOps Engineer
- Data Architect / Solution Architect
- AI Engineer / LLM Engineer
- Technical teams responsible for data platforms and artificial intelligence infrastructures

Outline

1. Agentic AI Architecture & Foundations

- The concept of Agentic AI and its differences from traditional LLM applications
- Agentic AI architectural components
- Agent lifecycle
- Planning, Reasoning, and Acting approaches
- Comparison of single-agent and multi-agent architectures
- Agentic AI use cases in data engineering and data platforms

2. Tool Calling / Function Calling

- LLM Tool Calling and Function Calling mechanisms
- Tool definition and JSON Schema usage
- Structured Outputs
- Tool execution and result management
- Tool Calling error management
- Development of custom tools for data platforms
- Python-based Tool Calling applications

3. Model Context Protocol (MCP)

- MCP architecture and core components
- MCP Client and MCP Server concepts
- MCP Tools, Resources, and Prompts
- MCP Server development
- Making enterprise data sources accessible through MCP
- Database and API integrations with MCP
- MCP security and access controls
- Practice: Developing an MCP Server for a data platform

4. Agent Orchestration

- Agent orchestration architectures
- Workflow-based and dynamic agent systems
- State Management
- Agent execution lifecycle

- Retry, timeout, and error management
- LangGraph and similar orchestration approaches
- Long-running agent workflows
- Practice: Orchestration of a data engineering workflow with agents

5. Multi-Agent Architectures

- Architecture of Multi-Agent systems
- Supervisor and Worker Agent models
- Agent-to-Agent communication
- Task distribution and coordination
- Specialized agent design
- Error management in multi-agent systems
- Practice: A system consisting of Data Discovery, Data Quality, and Metadata agents

6. RAG + Agents

- Retrieval-Augmented Generation (RAG) architecture
- Agentic RAG approach
- Vector Database integration
- Semantic Search and Hybrid Search
- Retrieval strategies
- Evaluation of RAG results
- Integration with enterprise document and metadata sources
- Practice: Agentic RAG on an enterprise data catalog

7. Agent Memory

- Agent Memory concept
- Short-Term and Long-Term Memory
- Conversation Memory
- Persistent Memory
- Vector-based memory
- Memory Management strategies
- Data security and memory management
- Practice: A context-preserving data analysis agent

8. Agents Interacting with Databases & APIs

- Agent architectures that interact with databases
- Text-to-SQL approaches
- Generation and validation of SQL queries
- Database access security
- API Calling and service integrations

- Authentication and Authorization
- Error management and data validation
- Practice: A data assistant using SQL and REST API

9. Data Discovery Agents

- Enterprise data discovery processes
- Automatic discovery of data sources
- Analysis of data schemas
- Data profiling
- Integration with data catalogs
- Data search through metadata
- Classification of data assets
- Practice: Data Discovery Agent development

10. Data Quality Agents

- Data quality dimensions
- Creation of data quality rules
- Data validation and consistency checks
- Anomaly detection
- Classification of data quality issues
- Automated quality reports
- Data quality improvement recommendations
- Practice: Data Quality Agent development

11. Metadata & Data Lineage Agents

- Metadata Management concepts
- Technical and Business Metadata
- Data Lineage concept
- Analysis of data flows and dependencies
- Integration with metadata catalogs
- Tracking data transformations
- Impact analysis (Impact Analysis)
- Practice: Metadata and Lineage Analysis Agent

12. Human-in-the-Loop

- Human-in-the-Loop architectures
- Human approval mechanisms
- Checkpoints for critical operations
- Approval Workflow design
- Agent execution pause/resume

- Manual intervention and error correction
- Controlled agent use in banking systems
- Practice: Human-approved data update process

13. Agent Evaluation

- Agent Evaluation approaches
- Measuring Tool Calling accuracy
- Task Completion and Success Rate
- RAG evaluation metrics
- Analysis of Hallucination and incorrect tool use
- Creation of test scenarios and benchmarks
- Regression Testing
- Practice: Agent Evaluation Pipeline

14. Observability

- Agent Observability concept
- Logging, Metrics, and Tracing
- Agent execution tracing
- Tool Calling monitoring
- Token usage and cost analysis
- Latency and performance measurement
- OpenTelemetry integration
- Monitoring dashboards
- Practice: Agent tracing and monitoring

15. Security & Governance

- Agentic AI security risks
- Prompt Injection
- Tool misuse and unauthorized operation risks
- Data leakage and sensitive data management
- Agent security policies
- Audit Logging
- AI Governance principles
- Enterprise risk and compliance requirements

16. Permissions & Data Access

- Authentication and Authorization
- Role-Based Access Control (RBAC)
- Attribute-Based Access Control (ABAC)
- Least Privilege approach

- Agent Identity
- Implementation of data access policies
- Secret Management
- Database and API authorizations
- Secure data access scenarios for banking systems

17. Production Deployment & LLMOps

- Production architecture of Agentic AI applications
- Containerization and Docker
- Deployment approaches on Kubernetes
- CI/CD processes
- Agent Versioning
- Prompt and model versioning
- Configuration Management
- Monitoring and Alerting
- Scalability and High Availability
- LLMOps lifecycle
- Production error management and operational processes

18. End-to-End Practice: Enterprise Data Engineering Agent Platform

- Definition of the enterprise data platform scenario
- Access to MCP-based data sources
- Data Discovery Agent development
- Data Quality Agent development
- Metadata / Lineage Agent development
- Multi-Agent orchestration
- Human-in-the-Loop approval mechanism
- Security and authorization controls
- Agent Evaluation and Observability integration
- Creation of the production deployment architecture

19. Recommended 5-Day Schedule

- Day 1: Agentic AI Architecture, Tool Calling, MCP
- Day 2: Agent Orchestration, Multi-Agent Architectures, RAG, Agent Memory
- Day 3: Database & API Agents, Data Discovery, Data Quality, Metadata & Lineage
- Day 4: Human-in-the-Loop, Agent Evaluation, Observability, Security & Governance, Permissions
- Day 5: Production Deployment, LLMOps, End-to-End Practice
- Note: If more time is requested for in-depth practice and production scenarios, the training duration can be extended to 6–7 days.

Prerequisites

- Experience in application development with Python
- Knowledge of SQL and relational databases
- Proficiency in REST API and JSON data formats
- Basic knowledge of data engineering and ETL/ELT processes
- Familiarity with LLM and Generative AI concepts
- Knowledge of Docker, Git, and basic cloud/platform technologies is recommended