AI for Data Engineers - 24 hours

Course description

This course bridges the worlds of data engineering and artificial intelligence, empowering data professionals to understand, design, and build AI-driven solutions across modern data ecosystems. Participants start with AI foundations, exploring core concepts, types of AI, and real enterprise use cases—from predictive models and generative AI to retrieval-augmented generation (RAG) and autonomous agents.
The course then focuses on prompt engineering and effective interaction with AI models using structured frameworks such as Role–Context–Task–Format, followed by practical applications. Building on this foundation, participants explore the role of Data Engineers in AI projects, including data preparation, feature engineering, and working with unstructured and multimodal data. The course includes hands-on experience with LangChain and RAG architectures for building enterprise chatbots, alongside critical topics such as security, governance, bias, and cost management.
Note: Tools and technologies may change due to the rapid evolution of Generative AI.Note: Please note that the tools and technologies covered in this course are subject to change, as the field of Generative AI is evolving at an exceptionally rapid pace.

What Participants Gain?

By the end of the course, participants will understand core AI concepts and types, and identify where AI creates business value in data platforms and analytics solutions. They will be able to design effective prompts using structured frameworks and apply them to real-world data engineering tasks. Participants will integrate AI into data pipelines, prepare and enrich data for AI use cases, and handle unstructured and multimodal inputs such as documents and images. They will design and implement RAG-based enterprise chatbots using LangChain, and apply best practices for security, governance, cost control, and responsible AI in production data environments. Graduates will leave with practical .tools to boost productivity using AI-powered developer tools

Target Audience

- Data Engineers looking to integrate AI into data pipelines and workflows
- BI Developers and Data Analysts expanding into AI-driven data systems
- Data professionals interested in AI for analytics, automation, and knowledge management
- Developers and architects building data infrastructure connected to Generative and Agentic AI

Prerequisites

- Experience with data engineering concepts and workflows
- Familiarity with databases, data pipelines, and basic programming
- Basic understanding of APIs and working with structured data
- No prior experience with AI or LLMs is required

Course Contents:

Module 1: Introduction to AI

- What is AI?
- AI Types (predictive, generative, RAG, agents)
- Where AI creates value for organizations
- Typical enterprise use cases (chatbots, document summarization, analytics)

Module 2: Prompt Engineering

- Why Prompt Engineering matters
- Instructional vs. Conversational prompts
- Role–Context–Task–Format framework
- Sample use cases
- Practice

Module 3: AI in the world of Data engineering

- Role of Data Engineers in AI projects
- Data preparation – collection, cleaning, feature engineering
- OCR for unstructured data (documents)
- Multimodal AI – text, image
- LangChain fundamentals and RAG process
- Building enterprise chatbots with AI
- Security & Governance in AI pipelines
- Challenges – data quality, bias, cost

Modeul 4 : AI Tools & Productivity for Data Engineers

- GitHub Copilot / Copilot Workspace – coding & PR assist
- Cursor / JetBrains AI – completions & refactor
- Notion AI / Confluence AI – docs & runbooks