



# Mastering RAG: Building Knowledge-Aware Intelligent Systems with Large Language Models

## Introduction

As **Large Language Models (LLMs)** expand into ever more demanding application scenarios, the limits of pure generative approaches are becoming increasingly visible: outdated knowledge, hallucinations, and an inability to reason over private or proprietary data. **Retrieval-Augmented Generation (RAG)** has emerged as one of the most strategic answers to these limitations—combining the generative power of LLMs with real-time, contextual access to external and specialized knowledge bases. RAG is rapidly becoming a foundational architecture for production-grade AI systems, even though real-world implementations still face critical challenges around knowledge base design, retrieval-generation orchestration, latency, and dynamic data updates.

This expert-led training delivers a structured, end-to-end view of RAG: from its fundamental concepts and technical architecture to its construction process and deployment in real industry scenarios. It unpacks how the two core engines—**retrieval** and **generation**—work together to deliver accurate, up-to-date, and grounded responses, and walks through each stage of building a robust RAG pipeline. The course also explores how to iterate and improve RAG systems through evaluation metrics, and showcases concrete case studies in medical diagnosis, financial analysis, and intelligent customer service. By the end, participants will be equipped to design, build, and operationalize RAG-powered systems that deliver real business value.

## Course Objectives

By the end of this course, learners will be able to:

1. Understand the foundational concepts and technical architecture of RAG.
2. Master the end-to-end process of designing and building RAG systems on top of large models.
3. Identify high-value RAG application scenarios and best practices across industries.

## Target Trainees

AI architects, engineers, technology professionals, product managers, and IT support staff working with or planning to deploy large model-based applications, intelligent assistants, or knowledge-intensive AI systems.

## **Duration**

1 hour 20 min

## **Course Contents**

Module 1 : RAG Foundations: Core Concepts and High-Impact Application Scenarios

Module 2 : Inside the Engine: Key Technologies Powering Retrieval-Augmented Generation

Module 3 : From Blueprint to Production: Building and Iterating a Real-World RAG System