BOARD INFINITY

COURSE 2 OF 3 · AI AGENTS: FROM FOUNDATIONS TO APPLICATIONS SPECIALIZATION

RAG Systems & Agentic Workflows with Pinecone and LangGraph

Full course syllabus - every module, video lesson, reading and assignment, from embeddings and vector search to stateful agents you can trace and debug.

Board Infinity
Instructor · Offered on Coursera

4	26	14	2 weeks
Modules	Video lessons	Assignments	At 10 hours a week

INSIDE THE COURSE

01	Fundamentals of Retrieval-Augmented Generation (RAG)	4 hours to complete
02	Vector Databases and Semantic Search	5 hours to complete
03	Conversational Agents and Workflows	6 hours to complete
04	Optimization, Debugging, and Observability	4 hours to complete

Enroll on Coursera

https://www.coursera.org/learn/rag-systems-agentic-workflows-with-pinecone-and-langgraph

01 ABOUT THE COURSE

Retrieve, reason and respond.

This course equips you to build AI systems that retrieve, reason and respond using real-world data. You start with Retrieval-Augmented Generation - connecting LLMs to external knowledge with embeddings - then build vector indexes and semantic search across Chroma, Pinecone and FAISS.

Next you design stateful agents with LangGraph, memory and retrieval-driven Q&A, then debug, trace and optimize them - 26 videos, 10 readings and 14 assignments in all. Intermediate level; comfort with Python assumed.

WHO IT'S FOR

AI developers
Building LLM applications that need external knowledge.

GenAI practitioners
Moving from basic prompting to retrieval and agents.

Python developers
Ready to build practical RAG and agent systems.

AI engineers
Working on vector search, memory and observability.

WHAT YOU'LL LEARN

01 **Build end-to-end RAG pipelines** Module 1
Connect LLMs with external knowledge sources using embeddings and vector search

02 **Design and query vector databases** Module 2
Use ChromaDB, Pinecone and FAISS for context-aware semantic retrieval

03 **Create conversational agents** Module 3
Multi-step agents with session memory, state management and retrieval-driven Q&A

04 **Optimize and debug agent workflows** Module 4
Use observability tools like LangFuse, W&B and OpenTelemetry to improve agents

02 SYLLABUS

Four modules, from retrieval to agents.

2 weeks to complete at 10 hours a week. Flexible schedule - learn at your own pace.

01 Fundamentals of Retrieval-Augmented Generation (RAG)

RAG architecture, embeddings and a first pipeline

4 hours
6 videos

BY THE END OF THIS MODULE YOU CAN

- Explain the concept and workflow of RAG systems
- Break down the RAG architecture, from query to grounded answer
- Use embeddings and similarity search to retrieve relevant context
- Build a simple RAG pipeline, evaluate its responses and avoid common pitfalls

CONTENTS: 6 videos (63 min) · 2 readings (60 min) · 3 assignments (120 min) · 1 plugin (5 min)

TOPICS: Introduction to RAG · RAG Architecture · Embedding and Similarity Search · Evaluating RAG Responses

You'll leave with: a working RAG pipeline that grounds an LLM's answers in your own data.

02 Vector Databases and Semantic Search

FAISS, Chroma and Pinecone, hybrid search and chunking

5 hours
7 videos

BY THE END OF THIS MODULE YOU CAN

- Build vector indexes using Chroma or Pinecone
- Compare FAISS, Chroma and Pinecone and their performance tradeoffs
- Optimize queries and combine vector and keyword search
- Ingest and chunk data, and index a product knowledge base in a lab

CONTENTS: 7 videos (83 min) · 3 readings (90 min) · 4 assignments (150 min)

TOPICS: Building a Vector Store · Semantic Search and Querying · Performance Optimization and Scaling

You'll leave with: a searchable vector index, and the judgement to pick the right vector database for the job.

03 Conversational Agents and Workflows

Agentic workflows, LangGraph, memory and multi-step agents

6 hours
7 videos

BY THE END OF THIS MODULE YOU CAN

- ✓ Design state-based conversational workflows
- ✓ Explain agentic workflows and compare agent architectures
- ✓ Build workflows and agent architectures with LangGraph
- ✓ Add long-term memory and build retrieval-aware multi-step agents

CONTENTS 7 videos (99 min) · 3 readings (90 min) · 4 assignments (150 min)

TOPICS Designing Conversational Flow · Context Management and Memory · Building a Q&A Chatbot

You'll leave with: a stateful, memory-aware Q&A agent that retrieves before it answers.

04 Optimization, Debugging, and Observability

Tracing, reranking, query transformation and optimization

4 hours
6 videos

BY THE END OF THIS MODULE YOU CAN

- ✓ Diagnose agent performance issues using observability tools
- ✓ Set up tracing with LangSmith, or LangFuse as an alternative
- ✓ Improve outputs with query transformation and reranking
- ✓ Optimize a full pipeline, worked through as a case study

CONTENTS 6 videos (81 min) · 2 readings (60 min) · 3 assignments (120 min)

TOPICS Debugging RAG Pipelines · Observability and Tracing Tools · Query Transformation & Reranking

You'll leave with: traced, measurable agent pipelines - and a method for making them faster and more accurate.

03 TOOLS & SKILLS

The stack you build with.

Pinecone, Chroma & FAISS

Vector databases

Vector indexes, semantic search, hybrid search and chunking strategies.

In this course: Module 2, including a product knowledge base lab.

LangChain & LangGraph

Orchestration

Building pipelines, state-based workflows and agent architectures.

In this course: Module 3, from workflow design to memory agents.

LangSmith, LangFuse & OpenTelemetry

Observability

Tracing, monitoring and debugging agent runs end to end.

In this course: Module 4, with a pipeline optimization case study.

Python & embeddings

Foundation

Python code, embedding models and similarity search behind every pipeline.

In this course: Module 1, then used throughout.

SKILLS YOU LEAVE WITH

Retrieval

- Retrieval-augmented generation
- Embeddings
- Vector databases
- Large language modeling

Agents

- Generative AI agents
- Agentic systems
- Agentic workflows
- Memory management

Production

- Debugging
- Performance tuning
- AI workflows
- LangChain

04 CERTIFICATE & CAREER

Finish with a shareable certificate.

Add it to LinkedIn
Add this credential to your LinkedIn profile, resume, or CV.

Share it anywhere
Share it on social media and in your performance review.

Earned, not attended
Awarded on completion of the practice and graded quizzes in all four modules.

Course 2 of 3
Part of the AI Agents: From Foundations to Applications Specialization by Board Infinity.

ROLES THAT USE THESE SKILLS

AI Engineer · GenAI Developer · LLM Application Developer · Machine Learning Engineer · AI Solutions Architect



Board Infinity

Instructor · Offered on Coursera

Intermediate level · Taught in English · 9 languages available · Flexible schedule

Ready to build agents that know things?

Enrollment, lessons, grading and the certificate all happen on Coursera.

Enroll on Coursera

Course page:
coursera.org/learn/rag-systems-agentic-workflows-with-pinecone-and-langgraph

Want a coach alongside the course?

Board Infinity 1:1 coaching is a separate, optional Board Infinity offering - it is not part of the Coursera course.
Details: https://www.boardinfinity.com/career-coaching/

Board Infinity is an independent education provider. Coursera and the Coursera logo are trademarks of Coursera, Inc. This independent educational resource is for informational and educational purposes only and is not affiliated with, endorsed by, sponsored by, or officially associated with any company, organization, or certification body unless explicitly stated. All trademarks are the property of their respective owners.