BOARD INFINITY

COURSE 1 OF 3 · MANAGING AI SYSTEMS: DEVELOPMENT, DEPLOYMENT, AND GOVERNANCE

AI Systems Design: RAG Pipelines and LLM Architecture

Full course syllabus - every module, video lesson, reading and assignment, from framing the business problem to a working RAG prototype in Streamlit.

Board Infinity
Instructor · Offered on Coursera

4	36	15	2 weeks
Modules	Video lessons	Assignments	At 10 hours a week

INSIDE THE COURSE

01	Problem Formulation & AI Solution Scoping	4 hours to complete
02	The Modern AI Stack: Models, Orchestration, Vector Databases	5 hours to complete
03	Data Strategy for RAG and Fine-Tuning	4 hours to complete
04	Prototyping & Building the Company Brain	4 hours to complete

Enroll on Coursera

https://www.coursera.org/learn/ai-systems-design-rag-pipelines-and-llm-architecture

01 ABOUT THE COURSE

From business problem to working prototype.

This course teaches you to design and prototype enterprise-grade AI that creates real business value. You learn when to use predictive ML or generative AI, set success criteria across accuracy, latency, safety and cost, and diagram modern AI architectures - models, vector databases and orchestration.

You then prepare enterprise data with chunking and embedding strategies, compare RAG with fine-tuning, and build an end-to-end RAG prototype with a Streamlit UI - 36 videos, 12 readings and 15 assignments in all. Intermediate level; basic AI/ML familiarity helps.

WHO IT'S FOR

Engineers
Designing and building enterprise AI systems.

Product managers
Framing AI problems around business outcomes.

Technical program managers
Scoping, planning and delivering AI initiatives.

AI governance officers
Understanding how AI systems are built and tested.

WHAT YOU'LL LEARN

01 **Classify AI problems** Module 1
Apply decision frameworks to classify AI problems and evaluate build vs buy

02 **Build end-to-end RAG pipelines** Module 2
Use LangChain, ChromaDB and embedding models to retrieve and generate

03 **Design data strategies** Module 3
Chunking, embedding selection and retrieval optimization for enterprise data

04 **Prototype with Streamlit** Module 4
Create interactive AI prototypes for real-world business applications

02 SYLLABUS

Four modules, from problem to prototype.

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

01 Problem Formulation & AI Solution Scoping

ML vs GenAI, business metrics and probabilistic systems

4 hours
9 videos

BY THE END OF THIS MODULE YOU CAN

- Choose between predictive ML and generative AI with a decision framework
- Define KPIs, ROI and acceptance criteria for an AI initiative
- Translate business needs into AI requirements within real constraints
- Plan testing for non-deterministic outputs and explain AI behavior to stakeholders

CONTENTS: 9 videos (54 min) · 3 readings (90 min) · 3 assignments (90 min) · 1 plugin (5 min)

TOPICS: Framing AI Problems With Business Metrics · Deterministic Software vs Probabilistic AI · Failure Modes

You'll leave with: a scoped AI problem with clear success criteria - and a case for where AI fits and where it doesn't.

02 The Modern AI Stack: Models, Orchestration, Vector Databases

LLMs, vector stores, orchestrators and agent patterns

5 hours
9 videos

BY THE END OF THIS MODULE YOU CAN

- Identify the components of the AI stack: LLMs, vector stores and orchestrators
- Trace the data flow from ingestion to embedding, retrieval and generation
- Operate ChromaDB and FAISS with chunking and retrieval optimization
- Apply chains, tools and agent patterns, and implement a simple RAG chain

CONTENTS: 9 videos (77 min) · 3 readings (90 min) · 4 assignments (105 min)

TOPICS: Understanding the AI Stack · Vector Databases & Embeddings · LangChain / LlamaIndex

You'll leave with: a reference architecture for enterprise RAG, and a first working RAG chain.

03 Data Strategy for RAG and Fine-Tuning

Chunking, embeddings, RAG vs fine-tuning and data readiness

4 hours
9 videos

BY THE END OF THIS MODULE YOU CAN

- ✓ Set chunk size, overlap and structure, and choose an embedding model
- ✓ Measure retrieval accuracy
- ✓ Compare RAG and fine-tuning on cost, latency and data sensitivity
- ✓ Clean, normalize and tag enterprise data, and manage noise, PII and gaps

CONTENTS 9 videos (61 min) · 3 readings (90 min) · 4 assignments (105 min)

TOPICS Chunking & Embedding Strategies · RAG vs Fine-Tuning · Preparing Enterprise Data

You'll leave with: a data strategy for messy enterprise data - and the evidence to choose RAG or fine-tuning.

04 Prototyping & Building the Company Brain

End-to-end RAG, Streamlit UI, deployment and debugging

4 hours
9 videos

BY THE END OF THIS MODULE YOU CAN

- ✓ Implement ingestion, chunking and embedding, and validate retrieval quality
- ✓ Build a Streamlit UI and integrate the retrieval chain into it
- ✓ Deploy the MVP
- ✓ Find hallucinations, fix latency and improve grounding from test logs

CONTENTS 9 videos (59 min) · 3 readings (90 min) · 4 assignments (105 min)

TOPICS Building the Retrieval Pipeline · Streamlit for AI Prototypes · Testing & Refining the MVP

You'll leave with: a deployed "company brain" - an end-to-end RAG prototype with a Streamlit UI.

03 TOOLS & SKILLS

The stack you design with.

LLMs & embedding models

Models

Generative and predictive models, and the embedding models that power retrieval.

In this course: Modules 1 to 3, from paradigm choice to model selection.

ChromaDB & FAISS

Vector databases

Storing and searching embeddings, with chunking and retrieval optimization.

In this course: Module 2, then used in the Module 4 prototype.

LangChain & LlamaIndex

Orchestration

Chains, tools and agent patterns, and when to use each.

In this course: Module 2, including a simple RAG chain.

Streamlit

Prototype UI

An interactive front end for your RAG system, built in Python.

In this course: Module 4, through to the deployed MVP.

SKILLS YOU LEAVE WITH

Architecture

- Systems architecture
- Solution architecture
- AI orchestration
- LLM application

Data & retrieval

- Retrieval-augmented generation
- Embeddings
- Vector databases
- Unstructured data

Build & refine

- Fine-tuning
- Data preprocessing
- User interface (UI)
- Generative AI

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 assignments and quizzes in all four modules.

Course 1 of 3
Part of the Managing AI Systems: Development, Deployment, and Governance Specialization.

ROLES THE COURSE PREPARES YOU FOR

AI Solutions Architect · AI Product Manager · Technical Program Manager · AI Governance Officer · AI Engineer

Board Infinity
Instructor · Offered on Coursera

Intermediate level · Taught in English · Flexible schedule

Ready to design AI that holds up?

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

Enroll on Coursera

Course page: coursera.org/learn/ai-systems-design-rag-pipelines-and-llm-architecture

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 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.