

Advanced RAG for LLMs Application

ระยะเวลาอบรม 3 วัน

หลักการและเหตุผล

หลักสูตรเชิงปฏิบัติการที่จะพาคุณเรียนรู้การใช้งาน Retrieval-Augmented Generation (RAG) สำหรับการประยุกต์ใช้กับ Large Language Models (LLMs) เรียนรู้ตั้งแต่พื้นฐานไปจนถึงการพัฒนา RAG ที่นำไปใช้งานได้จริง

รายละเอียดหลักสูตร

วันที่ 1: Essential of LLM & RAG

- LLM คืออะไร แตกต่างจาก ML model อย่างไร
- ทำความเข้าใจกลไกการทำงานภายใน Transformer Architecture
 - Embedding mechanism
 - Attention mechanism
- ความสามารถหลักของ LLM และ Application ที่เหมาะสม
- โครงสร้างพื้นฐานของการนำ LLM มาใช้ใน Application
- Workshop: การนำ LLM มาใช้อย่างง่าย
- Introduction to RAG
 - ความเป็นมาและความสำคัญ
 - ข้อจำกัดของ LLMs และการแก้ไขด้วย RAG
 - สถาปัตยกรรมพื้นฐาน

วันที่ 2: การจัดการข้อมูลใน RAG

- Document Processing
 - การเตรียมข้อมูลเอกสาร
 - เทคนิคการแบ่ง Chunk ที่เหมาะสม

Workshop: การเตรียมข้อมูลสำหรับ RAG

- Vector Database Basics
 - แนะนำ Vector Database
 - การเลือก Vector Database ที่เหมาะสม

Workshop: การใช้งาน Vector Database เบื้องต้น

วันที่ 3: การพัฒนาระบบ RAG ที่มีการสืบค้นข้อมูลจาก Vector DB

- Embeddings & Basic Retrieval
 - การสร้าง Embeddings
 - การเลือก Embedding Model
 - การค้นคืนข้อมูลพื้นฐาน

Workshop: พัฒนาระบบค้นคืนข้อมูล

- RAG Implementation
 - การเชื่อมต่อกับ LLMs
 - การสร้าง Prompt ที่เหมาะสม
 - การจัดการ Context

Workshop: พัฒนาระบบ RAG เบื้องต้น

- การออกแบบ Prompt สำหรับ RAG
 - Shot-based prompting techniques
 - Chain of Thought (CoT)
 - ReAct Prompting

Workshop: การออกแบบ Prompt ที่เหมาะสมสำหรับ RAG

- Basic Optimization
 - การปรับปรุงคุณภาพผลลัพธ์
 - การจัดการ Token Limit

Workshop: ปรับแต่งระบบ RAG

หลักสูตรนี้เหมาะสำหรับ

- นักพัฒนาซอฟต์แวร์ที่ต้องการเริ่มต้นทำงานกับ LLMs
- วิศวกรข้อมูลที่สนใจพัฒนาระบบ AI
- ผู้ที่ต้องการนำ RAG ไปประยุกต์ใช้กับงานจริง

พื้นฐานของผู้เข้าอบรม

- มีความรู้พื้นฐานการเขียนโปรแกรมภาษา Python

ประโยชน์ที่จะได้รับการอบรม

- เรียนรู้และเข้าใจความสามารถของ LLM และ Application ที่เหมาะสม
- เข้าใจหลักการทำงานของ RAG
- สามารถพัฒนาระบบ RAG ขั้นพื้นฐาน
- นำความรู้ไปประยุกต์ใช้กับงานจริงได้