

1. Tóm tắt nội dung tuần 8

Tuần 8 tập trung vào các khía cạnh **nâng cao của triển khai AI**, bao gồm hệ thống đa tác nhân (**Multi-Agent Systems**), tối ưu hóa **LLMs** với **RAG**, triển khai **serverless AI**, quản lý dữ liệu vector, và tối ưu hóa giao diện người dùng AI. Các nội dung chính được chia theo từng ngày như sau:

Ngày 1-2: Xây dựng hệ thống AI đa tác nhân & tối ưu hóa mô hình AI

- Triển khai **Multi-Agent Systems (MAS)** để xử lý nhiệm vụ phức tạp bằng cách kết hợp nhiều mô hình AI.
- Tích hợp **RAG (Retrieval-Augmented Generation)** để cải thiện khả năng truy xuất thông tin của LLMs.
- Tạo **Vector Stores** với **ChromaDB** giúp lưu trữ và truy vấn dữ liệu hiệu quả hơn.
- Ứng dụng AI trong tìm kiếm giao dịch (**Deal Finding**) và dự đoán giá cả (**Pricing AI**).
- So sánh **Fine-Tuning** và **Multi-Agent Systems** để tối ưu hiệu suất của mô hình AI.

Ngày 3: Cải thiện quy trình làm việc AI (AI Workflows) & Tích hợp dữ liệu

- Tối ưu hóa **đầu ra có cấu trúc** với **Pydantic & BaseModel** để AI tạo dữ liệu rõ ràng hơn.
- Xây dựng hệ thống AI **thu thập dữ liệu từ RSS Feeds** để tự động chọn lọc thông tin hữu ích.
- Cải thiện **độ chính xác của AI** thông qua **Prompt Engineering** và tối ưu hóa truy vấn.
- Thiết kế **quy trình làm việc cho AI đa tác nhân (Multi-Agent AI Workflows)** giúp AI hoạt động linh hoạt hơn.

Ngày 4: Agentic AI & Tích hợp hệ thống thông minh

- Giới thiệu **5 đặc điểm** của **Agentic AI**: Tự động hóa (**Autonomy**), Lập kế hoạch (**Planning**), Bộ nhớ (**Memory**), Thích ứng (**Adaptability**), và Hợp tác (**Collaboration**).
- Tích hợp **Pushover API** để gửi thông báo tự động trong hệ thống AI.
- Xây dựng **Planning Agent** giúp AI tự động lên kế hoạch và quản lý quy trình làm việc.
- Phát triển **framework kết nối LLMs với Python** để AI thực hiện nhiệm vụ linh hoạt.

Ngày 5: Hoàn thiện hệ thống AI & Tối ưu hóa giao diện người dùng

- Xây dựng giao diện AI bằng **Gradio** giúp người dùng tương tác dễ dàng với hệ thống.
- Cải thiện giao diện bằng **real-time logging** để theo dõi hiệu suất AI trực tiếp.

- Phân tích **hiệu suất AI Agents** để đánh giá khả năng hoạt động của hệ thống.
- **Tổng kết hành trình 8 tuần học về LLM Engineering**, từ xây dựng mô hình đến triển khai thực tế.

2. Các từ khóa quan trọng có trong tuần 8

Hệ thống & Mô hình AI

- **Multi-Agent Systems:** Hệ thống AI gồm nhiều tác nhân phối hợp để xử lý nhiệm vụ phức tạp.
- **Agentic AI:** AI có khả năng tự động hóa, lập kế hoạch, ghi nhớ và điều chỉnh hành vi theo môi trường.
- **LLM (Large Language Model):** Mô hình ngôn ngữ lớn, nền tảng cho AI hiện đại.
- **LLAMA:** Một dòng mô hình AI tiên tiến do Meta phát triển.
- **RAG (Retrieval-Augmented Generation):** Phương pháp kết hợp mô hình ngôn ngữ với truy xuất dữ liệu bên ngoài để nâng cao chất lượng phản hồi.
- **Fine-Tuning:** Kỹ thuật tinh chỉnh mô hình AI trên tập dữ liệu cụ thể để tối ưu hóa hiệu suất.
- **Planning Agent:** Tác nhân AI có khả năng lập kế hoạch và quản lý quy trình làm việc.
- **AI Workflows:** Quy trình tự động hóa các tác vụ AI giúp tối ưu hóa hiệu suất hoạt động.

Quản lý Dữ liệu & Tối ưu hóa AI

- **Vector Stores:** Hệ thống lưu trữ vector giúp AI truy xuất dữ liệu nhanh chóng và chính xác hơn.
- **ChromaDB:** Hệ thống cơ sở dữ liệu vector hỗ trợ truy vấn nhanh trong AI.
- **Embedding Models:** Mô hình AI dùng để biến đổi dữ liệu thành vector giúp truy vấn hiệu quả.
- **Knowledge Graphs:** Mô hình biểu diễn dữ liệu giúp AI hiểu mối quan hệ giữa các thực thể.
- **Structured Data Extraction:** Trích xuất dữ liệu có cấu trúc từ văn bản để AI xử lý tốt hơn.
- **AI Memory Management:** Quản lý bộ nhớ trong AI để duy trì thông tin ngữ cảnh dài hạn.

Triển khai & Hạ tầng AI

- **Serverless AI:** Triển khai AI trên cloud mà không cần quản lý hạ tầng.
- **Cloud Computing:** Điện toán đám mây giúp lưu trữ và xử lý dữ liệu AI hiệu quả.
- **Serverless Architecture (Modal):** Kiến trúc không máy chủ giúp AI vận hành linh hoạt hơn.
- **Scalable AI Systems:** Hệ thống AI có khả năng mở rộng dễ dàng khi nhu cầu tăng cao.

- **Automated Workflows:** Tự động hóa quy trình làm việc AI giúp tăng hiệu suất.
- **Business AI Scaling:** Mở rộng ứng dụng AI trong doanh nghiệp để tối ưu hóa hiệu quả kinh doanh.

Giao diện & Tương tác Người dùng

- **Gradio & Streamlit:** Công cụ tạo giao diện UI cho AI, giúp người dùng tương tác dễ dàng hơn.
- **Real-Time Logging:** Theo dõi hoạt động AI theo thời gian thực để đánh giá hiệu suất.
- **AI UX Optimization:** Cải thiện trải nghiệm người dùng khi tương tác với AI.
- **Công cụ & API Hỗ trợ**
- **Pydantic & BaseModel:** Công cụ hỗ trợ tạo cấu trúc dữ liệu đầu ra cho AI một cách rõ ràng.
- **Pushover API:** API dùng để gửi thông báo tự động trong hệ thống AI.
- **Logging System:** Hệ thống ghi log giúp theo dõi và phân tích hiệu suất AI.
- **Prompt Engineering:** Kỹ thuật thiết kế prompt giúp AI hiểu và phản hồi tốt hơn.
- **Python:** Ngôn ngữ lập trình chính được sử dụng trong các mô hình AI.

3. Các công nghệ được đề cập trong tuần 8

Mô hình AI & Kỹ thuật

- **LLM (Large Language Model)** – Mô hình ngôn ngữ lớn, nền tảng cho AI hiện đại.
- **LLAMA** – Dòng mô hình AI tiên tiến do Meta phát triển.
- **Multi-Agent Systems** – Hệ thống AI gồm nhiều tác nhân phối hợp để xử lý nhiệm vụ phức tạp.
- **Agentic AI** – AI có khả năng tự động hóa, lập kế hoạch, ghi nhớ và điều chỉnh hành vi theo môi trường.
- **RAG (Retrieval-Augmented Generation)** – Kết hợp AI ngôn ngữ với truy xuất dữ liệu bên ngoài để nâng cao chất lượng phản hồi.

Nền tảng & Triển khai

- **Cloud Computing** – Điện toán đám mây giúp lưu trữ và xử lý dữ liệu AI hiệu quả.
- **Serverless Architecture (Modal)** – Kiến trúc AI không máy chủ giúp triển khai linh hoạt.
- **ChromaDB (Vector Stores)** – Hệ thống cơ sở dữ liệu vector hỗ trợ truy vấn nhanh trong AI.

Ngôn ngữ & Công cụ

- **Python** – Ngôn ngữ lập trình chính cho các mô hình AI.
- **Pydantic & BaseModel** – Công cụ hỗ trợ định dạng dữ liệu đầu ra của AI.

- **Gradio & Streamlit** – Công cụ tạo giao diện giúp người dùng tương tác với AI dễ dàng hơn.
- **Pushover API** – API dùng để gửi thông báo tự động trong hệ thống AI.
- **Logging System** – Hệ thống ghi log giúp theo dõi và phân tích hiệu suất AI.

Ứng dụng & Hệ thống

- **Automated Workflows** – Tự động hóa quy trình làm việc AI giúp tăng hiệu suất.
- **Business AI Scaling** – Mở rộng AI trong doanh nghiệp để tối ưu hóa hiệu quả.
- **Real-Time Log Visualization** – Theo dõi hoạt động AI theo thời gian thực.
- **AI Performance Monitoring** – Giám sát hiệu suất hệ thống AI để tối ưu vận hành.