MGMTMSA 440 Building LLM Powered Applications

Final Project

Multi-Agent LLMs For Wine Explorer

Project Report Summary

Group 2: Sen Yang, Cocoon Cao, Vicky Li, Darrel Wu

**Problem**

Selecting the perfect wine can be an overwhelming task, whether for a special occasion or simply as a daily indulgence. With thousands of options available, each varying in flavor profile, ratings, price, and reviews, the decision-making process becomes daunting for both novice and seasoned wine consumers. Traditional recommendation systems often rely on simplistic filters, such as type, price, and origin, which fall short of capturing nuanced user preferences. These systems tend to ignore factors like individual taste profiles or how a wine pairs with specific meals, leaving users unsatisfied and struggling to make an informed choice.

To tackle these challenges, our project aims to develop a comprehensive, **Multi-Agent LLMs Wine Recommendation System**. By integrating large language models (LLMs) and advanced techniques such as semantic search and AI agents, our system will offer personalized wine suggestions that go beyond traditional filters. This approach ensures that users receive highly tailored recommendations based on wine characteristics like reviews, ratings, expert insights, and even food pairing suggestions. Additionally, AI agents scrape educational content from YouTube channels to further enhance user knowledge and provide a holistic wine selection experience.

**Solution**

Our solution leverages the power of **LLMs** and **AI agents** to redefine how users explore and select wines. Using **semantic search** and **vector search techniques**, the system moves beyond keyword-based filtering by understanding the intent behind user queries. For example, if a user searches for “best red wine under $20,” the system not only finds matching wines by price but also ensures that the wines align with individual taste preferences and event needs.

The process begins by analyzing a vast dataset of wine reviews, ratings, and descriptions from Kaggle. Each wine summary is transformed into high-dimensional embeddings using the **SentenceTransformer** model. These embeddings capture the semantic meaning of each wine's description, which allows the system to retrieve the most relevant results based on user queries. Using **FAISS** (Facebook AI Similarity Search), the system stores these embeddings, enabling fast and efficient retrieval of similar wines based on **cosine similarity** measures.

One of the unique features of our solution is the integration of **AI Agents** to enhance the overall wine recommendation process. Once a wine is selected through our system, **Agent 3** takes over to provide additional educational content, such as information on the grape variety, optimal drinking methods, and suggested food pairings. This is achieved by **Agent 2**, the Video Analyst Agent, which scrapes relevant YouTube channels for expert reviews, educational videos, and pairing suggestions. This deeper layer of information ensures that users not only get a recommendation but also a well-rounded understanding of why that wine is an ideal choice for them.

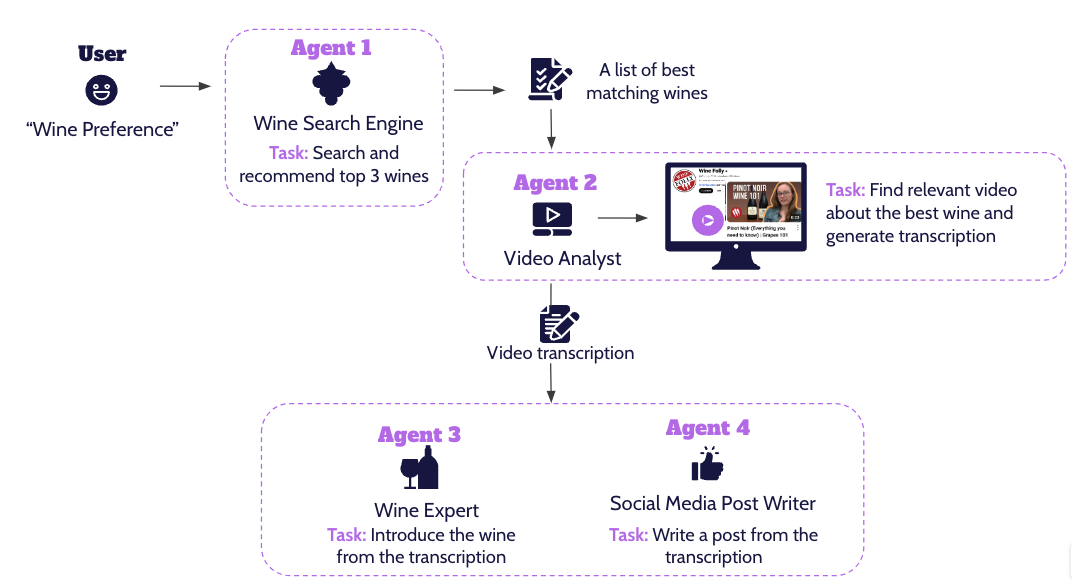
Moreover, our system automates social media engagement through **Agent 4**, the Social Media Post Writer Agent. This agent creates concise, engaging social media posts that summarize key attributes of the wine, such as flavor profile, region, and pairing ideas. Users can effortlessly share their wine discoveries with their network, making the experience more interactive and fun.

In total, our solution employs four distinct AI agents, each designed to perform a specific function:

1. **Wine Search Engine (Agent 1)**: Handles the initial query-to-dataset match.
2. **Video Analyst Agent (Agent 2)**: Scrapes YouTube content to extract additional insights about the wine and its background.
3. **Wine Expert Agent (Agent 3)**: Provides a concise summary of the wine’s background, including how to drink it and what food to pair it with.
4. **Social Media Post Writer Agent (Agent 4)**: Generates an engaging, shareable social media post that highlights the wine and invites friends to join in on the experience.

This multi-agent system not only recommends wines but also educates users, offering a rich, personalized experience that goes beyond traditional wine filtering.

**Technical implementation**



#### Key Technologies:

1. **Semantic Search:** The semantic search engine leverages multiple approaches to deliver personalized wine recommendations based on user queries. Natural Language Processing (NLP) plays a key role in transforming both user inputs and wine data into dense vector embeddings. These embeddings are generated using the all-MiniLM-L6-v2 model from the sentence\_transformers library, which captures the semantic meaning of text. During the process using RecursiveCharacterTextSplitter, the system breaks down wine descriptions into manageable chunks without losing context, enabling more precise embedding generation. Finally, to efficiently handle large-scale vector comparisons, the system stores embeddings in FAISS (Facebook AI Similarity Search), a robust tool designed for high-speed similarity searches over dense vectors.
2. **Multi-Agent System**: Rather than relying on a single LLM model, the system uses multiple specialized agents, each dedicated to a specific task. This modular approach allows the system to effectively automate complex workflows. By distributing tasks across agents, the system ensures both flexibility and scalability, allowing for easy adjustments or additions of new agents as needed.
3. **YouTube Content Scraping**: Tools like yt-dlp and youtube-transcript-api are used to gather real-time video content and transcripts from YouTube, allowing the system to provide expert insights beyond static datasets.
4. **LLM-Driven Content Generation**: Our project leverages **LLM** to generate structured, human-like content. The system uses “gpt-4o” model to distill complex information from the transcripts provided by Agent 2 into concise, user-friendly wine summaries (Agent 3) and engaging social media posts (Agent 4). This ensures high-quality, contextually relevant content that enhances user engagement and education.

#### Agent 1: Semantic Search (Wine Search Engine)

**Semantic Search** allows for more intelligent, context-aware retrieval of information by focusing on the meaning behind queries rather than exact keyword matches. It works by converting both queries and target data into embeddings, numerical representations that encapsulate the relationships between words and phrases. In this project, the search engine compares user query embeddings with wine embeddings generated through NLP transformer models like all-MiniLM-L6-v2, enabling more meaningful matches even when the exact words don’t overlap. By calculating cosine similarity between the query and stored wine embeddings, the system ranks the most relevant results, offering personalized recommendations. This approach ensures that users find results tailored to their preferences, as the engine prioritizes semantic understanding over simple text matching, making sure the input for later muti-agents stage.

Agent 2: Video Analyst (YouTube Content Scraper)

The **Video Analyst** agent gathers relevant video content from the designated YouTube channel and extracting transcripts for detailed analysis. By using tools like yt-dlp and youtube-transcript-api, this agent collects comprehensive insights about specific wines.

This agent allows the system to access real-time information and expert knowledge that isn’t available in static datasets. It identifies relevant videos based on the wine query, and extracts video transcripts for further processing. These transcripts are passed on to other agents for other tasks.

#### Agent 3: Wine Expert (LLM-Powered)

The **Wine Expert** agent uses the extracted transcript provided by Agent 2 to generate a detailed, informative introduction to the wine. This agent leverages “gpt-4o” LLM model to process the transcript data and craft structured wine summaries, including grape backgrounds, optimal drinking methods, and food pairings.

#### Agent 4: Social Media Post Writer (LLM-Powered)

Similar to agent 3, the **Social Media Post Writer** agent utilizes “gpt-4o” LLM model to process the extracted transcript and generate an appealing social media post with hashtags based on the requirement for the tasks.

**Appendix**

Code link: [Colab](https://colab.research.google.com/drive/1MAvXa86ObHaD3gky3hI8lCoS6zk99-kk#scrollTo=x-MV_wT9tkWd)