**University Name:** Kafr El-Sheikh University

**Faculty:** Faculty of Computers and Information

**Department:** Information Systems

**Academic Year:** 2024-2025

**Project Title:** Ollama LLMs Integration (Chat AI )

**Supervisor(s):**

* Dr. Reda Mabrouk
* Prof. Abdullah Rashwan

**Team Members:**

**Name ID Department**

1. Ibrahim Ahmed Ibrahim 1618120210100339 CS
2. Ahmed Abdulrahim Hamed 1618120210100038 CS
3. Esraa Ali Sultan 1618120210100255 IS
4. Asmaa Ahmed Ayad 1618120210100365 IS
5. Basant Mohamed Selima 1618120210100366 CS
6. Khaled Mohamed Abdul-Hafez 1618120210100082 IS
7. Khoulod Khaled Bakr 1618120210100115 IS

**Project Objectives:**

* **Enable Advanced Text Processing**: Implement natural language processing (NLP) capabilities using the Llama models to enhance text generation, summarization, and contextual understanding.
* **Seamless API Integration**: Integrate Ollama's Llama2 API into an existing application with a robust back-end system, ensuring smooth and efficient communication between the front-end and the AI modeAutomate operations like employee scheduling, payroll, and performance tracking for efficient cinema management.
* **Optimize Response Time**: Minimize latency in API calls and ensure that AI-generated responses are delivered in real-time, maintaining a user-friendly experience.
* **Develop Personalized Recommendations**: Use Ollama’s ability to understand user preferences and interactions to build personalized recommendation systems, tailoring responses and suggestions based on individual user behavior and context.
* **Support Multi-Turn Conversations**: Implement multi-turn dialogue management using Ollama, enabling the application to maintain conversation states across multiple interactions and provide coherent, follow-up responses.

**Project Description:**

This project is designed to integrate the full capabilities of Ollama's suite of language models into a .NET environment, supporting all available models like Llama2, GPT-style models, and more. The application aims to provide advanced language processing functionalities such as text generation, summarization, and sentiment analysis, while maintaining support for multi-turn conversations to ensure a natural and coherent dialogue experience.

One of the core features is the inclusion of embedding models coupled with a vector database, enabling **Retrieval-Augmented Generation (RAG)**. This allows the app to perform complex, context-aware responses by referencing a user-uploaded knowledge stack, providing highly relevant and personalized outputs. By utilizing RAG, the app can read and understand user data from documents, PDFs, and other content, dynamically adapting its responses based on this knowledge.

The integration also prioritizes real-time performance, error handling, and scalability. The project is built using modular principles to accommodate future model updates and additional AI capabilities, making it a robust platform for businesses seeking to leverage AI for advanced, context-specific interactions.

**Tools and Technologies Used:**

* C#, Python
* .NET Core, ASP.NET Core
* **APIs**: Ollama API
* **Deployment** : Docker, Azure
* Vector Database
* Web Development (HTML, CSS, JavaScript)
* **Version Control**: Git/GitHub

**Project Phases and Timeline:**

| **Phase** | **Start Date** | **End Date** | **Description** |
| --- | --- | --- | --- |
| **Phase 1: Planning** | 01/11/2024 | 15/11/2024 | Initial planning, project scope definition, and requirements gathering. |
| **Phase 2: Research** | 16/11/2024 | 30/11/2024 | Research on Ollama models, vector databases, and RAG implementation strategies. |
| **Phase 3: Design** | 01/12/2024 | 16/12/2024 | Design system architecture, database schema, and API structure for model integration. |
| **Phase 4: Development** | 17/12/2024 | 31/02/2025 | Implement Ollama model integration, embedding models, and vector database features. |
| **Phase 5: Testing** | 01/03/2025 | 15/03/2025 | Conduct unit and integration testing, API performance optimization, and debugging. |
| **Phase 6: Final Report** | 16/03/2025 | 28/03/2025 | Prepare the final project report, documentation, and presentation for stakeholders. |

### **Expected Deliverables:**

* Fully functioning **Ollama-Integrated Language Model Application** with RAG and knowledge stack support.
* **Final Project Report** documenting implementation details and research findings.
* **PowerPoint Presentation** highlighting the project’s objectives, features, and results.

**Supervisor’s Approval:**

Dr. Ahmed Ali

A black text with a star

Description automatically generated

Date: 01/11/2023