2024

CIC summer internship

School of ITCS
AI program
Karim Mohamed AbdElSalam

IBM | Smart Village, Building No. B144

Contents

Brief description about IBM

IBM Egypt is a subsidiary of IBM, established to provide advanced technology solutions and services in Egypt. It focuses on helping businesses and organizations across various sectors with digital transformation through innovative technologies such as cloud computing, artificial intelligence, data analytics, and cybersecurity. IBM Egypt collaborates with clients to develop and implement tailored solutions that enhance efficiency, competitiveness, and growth. The company also invests in local talent and contributes to the development of the IT industry in Egypt through training programs, partnerships with educational institutions, and initiatives that foster innovation and entrepreneurship.

Brief description about Client innovation center

IBM's Client Innovation Center (CIC) in Egypt aims to leverage IBM's global expertise and local talent to drive digital transformation for businesses in Egypt and the broader region. It also serves as a collaborative space where clients can work closely with IBM experts to co-create and implement cutting-edge solutions tailored to their specific needs.

Project description

The project aimed to enhance the travel experience by providing a comprehensive and user-friendly platform for exploring Egypt's cultural heritage, historical landmarks, and natural attractions. Utilizing IBM's advanced technologies, our team developed a solution that offers personalized recommendations, real-time updates, and seamless booking options. This experience allowed me to apply my technical skills in a professional setting while supporting the promotion of Egypt as a premier travel destination.

Agenda

Week 1: We were given various lectures on various technologies such as cloud computing, testing frameworks, mobile development, AI, Devops and many more. During this week each intern would be able to explore all fields if he/she wanted to switch.

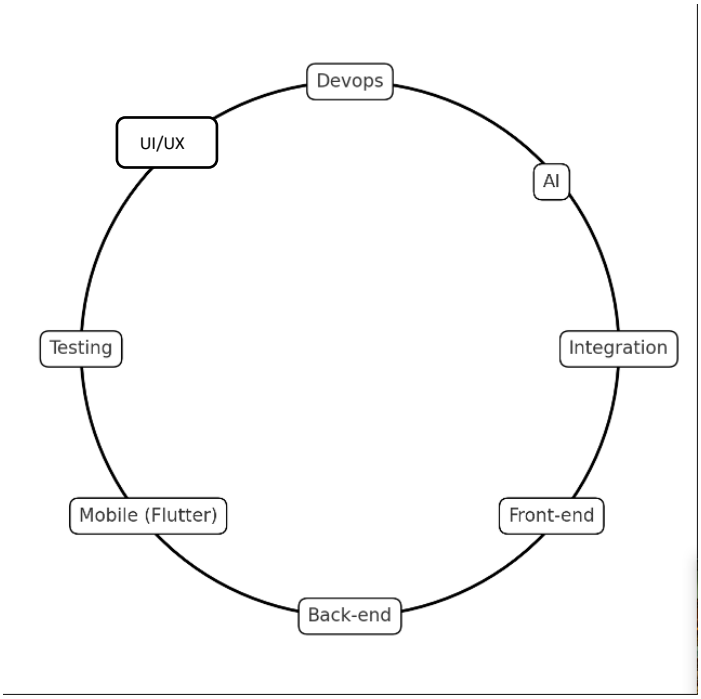
Week2: We participated in the IBM design thinking course in which we formulated ideas about our project. We were split into random teams each team would be given a persona. Each persona would require a list of features that would suit him/her

Week3: Each intern was assigned in 1 of the following teams: AI, devops, Integration, testing, mobile, front-end, back-end, UI/UX.

Specifically, I was in the AI team. We used the Agile format. We were required a demo of the application every week to showcase our work.

Subsequent weeks: Each intern would work on his/her part of the project with his team. Only team leaders would officially establish communications between teams to keep it professional.

Final week: We were assigned to showcases our work in front of IBM's general manager Marwa Abbas and CIC vice president Tamer Fahmy. Showcasing the application. Not individually as per the agile sprint meetings



List of teams

Picture of my design thinking team with our persona and ideas

Objectives

Tourist Assistance: The primary goal was to create a chatbot that could provide accurate and helpful responses to tourists' questions, ranging from historical facts to logistical information. Including information about hotels and places.

Trip Planning Module: In addition to answering questions, the chatbot included a trip planning module that could suggest itineraries, recommend attractions, and help with scheduling. The user would just specify his/her preferences and the chatbot would recommend a full trip planned for the user including stay, places to hang out, and even restaurants to eat at. Based on budget, preference, and time.

Text to speech model: The chatbot would be able to convert the output text to speech for the user.

Speech to text model: The user would be able to send an audio recording to the chatbot as if he's talking with a friend.

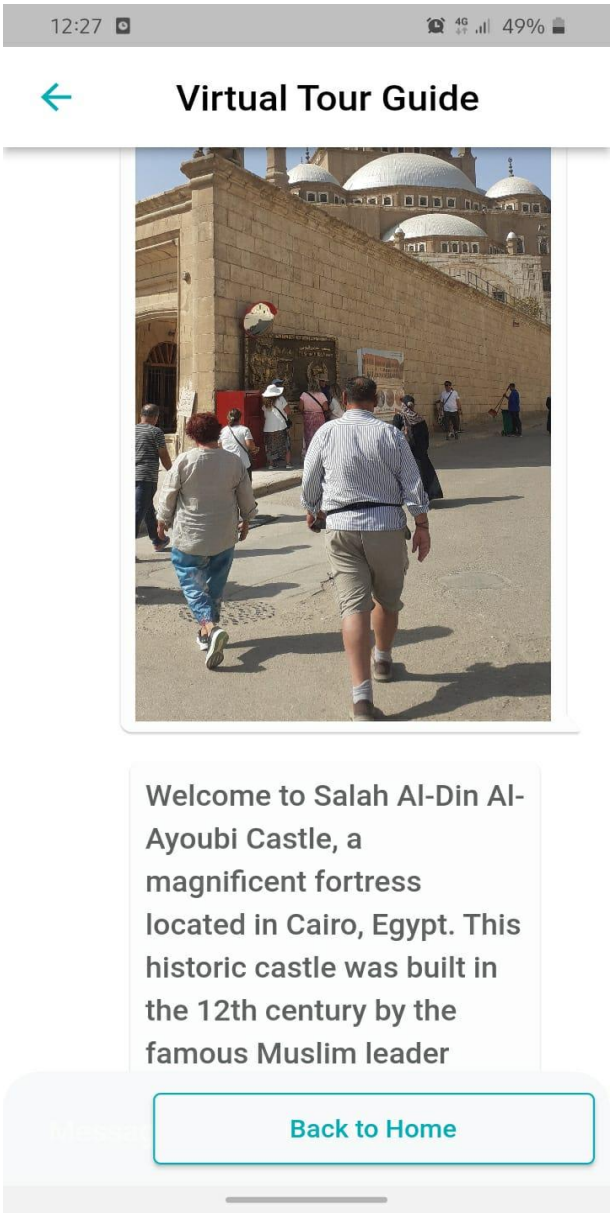
Virtual tour guide: The user would be able to take a picture and the chatbot would act as a tour guide for the user. Narrating the history of that place.

Note: the chatbot was able to support multiple languages

Anti-prompt injection layer: The user would not be able to hijack/manipulate the chatbot for malicious uses.

Adding sessions to the chatbot: the chatbot would be able to talk to multiple users at one time while holding the history of each user

Sample picture from the testing team while talking to the chatbot in Arabic



Sample picture of virtual tour guide

Knowledge/skills acquired

I gained extensive hands-on experience and developed a deep understanding of several advanced technologies and methodologies, which significantly enhanced my technical skill set.

LangChain Library

Learned the LangChain library in its entirety, enabling the creation of complex language models and workflows.

Large Language Models (LLMs)

Understanding LLMs: Gained insights into the inner workings of large language models, including their architecture, training methods, and practical applications.

Generative AI Hallucinations**:** Studied the phenomenon of generative AI hallucinations and implemented strategies to mitigate them, ensuring the chatbot provides accurate and reliable information.

Retrieval-Augmented Generation (RAG)

RAG Mechanisms: Learned how Retrieval-Augmented Generation works, enhancing the chatbot's ability to retrieve relevant information from external sources and databases.

Vector Databases: Acquired knowledge on how AI retrieves data from vector databases, improving the efficiency and accuracy of the information retrieval process.

Database Management

MongoDB**:** Gained practical experience with MongoDB, learning how to store and manage chatbot sessions and user data effectively.

Cloud Computing

Google Cloud: Utilized Google Cloud for training prompt injection, gaining expertise in cloud-based AI training and deployment.

API and Deployment Skills

REST APIs: Learned to use REST APIs for integrating various services and functionalities into the chatbot.

Docker: Acquired skills in deploying code using Docker, ensuring seamless and scalable application deployment.

Challenges faced

1-One of the major challenges was LLM hallucination. I would give the LLM detailed prompts on how to deal with the user expressing specific output parsers. But still, since LLMS are probabilistic, hallucination would occur in the most crucial parts. Info would be extracted incorrectly, API would be sent in the wrong format, and many more. I solved this issue by using chaining and prompt engineering.

2- Intent recognition, it was hard to recognize what module did the user want. Flight scheduler, trip planning, or hotel recommendations. I used a router chain to solve this issue that routes the user to a domain specific LLM. By analyzing each query to check if the user wants to switch context.

3- Adding sessions and memory to users, this was a hard task since it adds more tokens in the history of LLM. Thereby increasing hallucination and cost. I tackled this problem by using a buffer memory that picks up the last 5 queries so the LLM would not rack up memory and tokens for each user. Getting familiar with mongoDB was also challenging since I am always used to deal with SQL databases.

Theory vs market

During my university studies, I have been deeply immersed in the theoretical underpinnings of Artificial Intelligence, including the mathematical formulations of algorithms such as k-Nearest Neighbors (k-NN). This rigorous academic foundation has provided me with a thorough understanding of the principles and mechanics behind various AI techniques.

However, my experience in the job market has highlighted the importance of being able to apply these principles effectively in real-world scenarios. Employers highly value the ability to develop and deliver functional applications that meet business needs, often prioritizing practical implementation and a general understanding of the underlying technology over an in-depth theoretical knowledge. You don’t have to know how each mathematical formulas work.

This combination of strong theoretical background and practical application skills has enabled me to bridge the gap between academic learning and industry expectations. For instance, during my internship at IBM, I successfully developed a tourism application for Egypt, leveraging the ChatGPT-3.5 API to build a chatbot that assists tourists with questions and trip planning. This project underscored the importance of delivering solutions that work efficiently and meet user needs, demonstrating my capability to translate academic concepts into practical, impactful applications.

Experience and Observations

My internship at IBM provided a comprehensive and immersive experience in the field of artificial intelligence and software development. Working on the tourism application for Egypt allowed me to apply theoretical knowledge to real-world problems, enhancing both my technical skills and professional acumen. I observed a strong emphasis on collaboration, innovation, and the practical application of cutting-edge technologies, which significantly contributed to the success of the project and my personal growth.

Suggestions for Personal Improvement

Reflecting on my performance and experiences, I have identified several areas where I can improve to become a more effective and skilled professional:

1. **Time Management:** Focus on improving time management skills to balance multiple tasks and deadlines more efficiently.
2. **Continuous Learning:** Dedicate more time to continuous learning and staying updated with the latest advancements in AI and related technologies.
3. **Networking:** Actively seek opportunities to network and build professional relationships within the industry.
4. **Soft Skills Development:** Enhance soft skills, such as communication, teamwork, and problem-solving, to complement technical abilities and contribute to overall project success.
5. **Documentation Practices:** Improve my documentation practices to ensure that my work is easily understandable and maintainable by others.

Executive summary

During my internship at IBM, I was part of a dynamic AI team tasked with developing an innovative tourism application for Egypt. This project aimed to enhance the travel experience by providing a comprehensive and user-friendly platform for exploring Egypt's cultural heritage, historical landmarks, and natural attractions. Utilizing IBM's advanced technologies and AGILE methodologies, our team developed a solution that offers personalized recommendations, real-time updates, and seamless booking options.

The application was designed to guide and assist tourists by answering their questions and helping with trip planning. Key features of the chatbot included providing accurate and helpful responses to tourists' questions, suggesting itineraries, recommending attractions, and assisting with scheduling based on user preferences and constraints. The chatbot supported multiple languages to cater to a diverse audience. Additionally, the chatbot integrated text-to-speech and speech-to-text models, allowing for a more interactive user experience. It also featured a virtual tour guide that narrated the history of places based on pictures taken by the user, and an anti-prompt injection layer to prevent misuse.

Throughout the project, I learned and applied a variety of advanced technologies and methodologies. I mastered the LangChain library, gained insights into the workings of large language models (LLMs), and studied strategies to mitigate generative AI hallucinations. I learned how Retrieval-Augmented Generation (RAG) works and how AI retrieves data from vector databases. My experience with MongoDB taught me how to store and manage chatbot sessions, and I utilized Google Cloud for training prompt injections. I also acquired skills in using REST APIs and deploying code with Docker.

One of the major challenges I faced was dealing with LLM hallucinations, which I addressed through prompt engineering and chaining. Another challenge was intent recognition, which I resolved by using a router chain to direct users to domain specific LLMs. Managing user sessions and memory was also difficult, but I implemented a buffer memory system to handle it efficiently.

This internship provided a comprehensive learning experience, bridging the gap between theoretical knowledge and practical application. It underscored the importance of delivering functional applications that meet user needs and prepared me to contribute effectively to future AI projects. In the final week, we showcased the fully working application to the General Manager of IBM Egypt, demonstrating the project's success and the valuable skills I had acquired.

IBM Family picture

Certification