Application of ChatGPT in the Tourism Domain: Potential Structures and Challenges

Orkun Mehmet Kılıçlıoğlu
IT and R&D Center
SeturTech
İstanbul, Türkiye
orkun.kilicliogl u@setur.com.tr

Şuayb Talha Özçelik
[1]*Dept. of Computer Engr., Işık University*
[2]*IT and R&D Center, SeturTech*
Istanbul, Türkiye
[1]219DCS8111@isik.edu.tr,
[2]talha.ozcelik@setur.com.tr

Meltem Turhan Yöndem
IT and R&D Center
SeturTech
İstanbul, Türkiye
meltem.yondem@setur.com.tr

2023 4th International Informatics and Software Engineering Conference (IISEC) | 979-8-3503-1803-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/IISEC59749.2023.10390989

***Abstract*—The tourism industry stands out as a sector where effective customer communication significantly influences sales and customer satisfaction. The recent shift from traditional natural language processing methodologies to state-of-the-art deep learning and transformer-based models has revolutionized the development of Conversational AI tools. These tools can provide comprehensive information about a company's product portfolio, enhancing customer engagement and decision-making. One potential Conversational AI application can be developed with ChatGPT. In this study, we explore the potential of using ChatGPT, a cutting-edge Conversational AI, in the context of Setur's products and services, focusing on two distinct scenarios: intention recognition and response generation. We incorporate Setur-specific data, including hotel information and annual catalogs. Our research aims to present potential structures and strategies for utilizing Language Model-based systems, particularly ChatGPT, in the tourism domain. We investigate the advantages and disadvantages of three different architectures and evaluate whether a restrictive or more independent model would be suitable for our application. Despite the impressive performance of Large Language Models (LLMs) in generating human-like dialogues, their end-to-end application faces limitations, such as system prompt constraints, fine-tuning challenges, and model unavailability. Moreover, semantic search fails to deliver satisfactory performance when searching filters that require clear answers. To address these issues, we propose a hybrid approach that employs external interventions, the assignment of different GPT agents according to intent analysis, and traditional methods at specific junctures, which will facilitate the integration of domain knowledge into these systems.**

Keywords*—*ChatGPT, travel assistant, tourism, large language models

I. Introduction

The tourism industry stands out as a sector where effective customer communication significantly influences sales and customer satisfaction [1, 2]. While conventional communication channels with agencies remain prevalent, the digital landscape is witnessing a preference for chatbots and collaborative frameworks involving live sales representatives [3]. The Conversational AI can provide very detailed information about the entire product portfolio in seconds. If you have enough computational processing power, human-level service can be provided to everyone trying to receive assistance.

The recent shift from traditional natural language processing methodologies, including grammar analysis, stemming, part-of-speech tagging, spell-checking, and dependency parsing, to contemporary advancements powered by deep learning and transformer-based models is changing the way of developing Conversational AI tools. We will try to imitate a sales representative. Users can ask questions on different kinds of areas, and if it is not related to tourism, we need to detect and prevent it. We will design a system to understand what the user wants and provide the correct information directly on that subject.

The Generative Pre-trained Transformer (GPT) series is a pre-trained model on a wide range of text data [4]. GPT-1, the first iteration, was introduced in 2018 with 110 billion parameters [5]. It demonstrated significant success in language modeling and text completion tasks, showcasing its potential in the domain of natural language processing. Subsequently, GPT-2 was introduced in 2019, boasting a substantial increase in capacity with 1.5 billion parameters [6]. This model was noted for its ability to generate human-like text in some cases. The following year, GPT-3 was introduced in 2020 with 175 billion parameters [7]. This latest iteration further advanced the capabilities of its predecessors, demonstrating remarkable performance across a wide range of natural language understanding and generation tasks. To facilitate the integration of GPT-3 into various applications, an API was made available to developers, enabling them to harness the model's capabilities for their specific use cases. The GPT-3.5 model is an adaptation of the original GPT-3 model [8]. It is aimed to reduce toxic outcomes and strengthen dialogues. GPT-4 was announced on March 14, 2023, aiming to be more reliable and creative than GPT-3 [4]. After the release of GPT-4, the GPT-3.5 and GPT-3.5 Turbo (optimized for chat usage) regained prominence due to their cost-effectiveness and the limitations associated with GPT-4, particularly in terms of tokens per minute constraints [9]. Consequently, the development of GPT is a significant milestone in the fields of artificial intelligence and natural language processing [10].

We can use ChatGPT in two different scenarios: the first is to understand the intention of the question, and the second is to generate the answer. We want to direct people to Setur's products and services, so we have to provide Setur-specific data to this system. We will give hotel information from our portfolio and previously created annual catalogs. We will

979-8-3503-1803-6/23/$31.00 ©2023 IEEE

present potential structures that researchers who will use Large Language Model (LLM) based systems (especially ChatGPT) benefit from our findings. We will analyze whether we should follow a strategy that restricts or makes the model more independent. We tried three different architectures, each with its advantages and disadvantages.

In summary, this study will describe the processes of developing an LLM conversational AI solution specific to the tourism sector. A comprehensive analysis of the advantages and disadvantages inherent in transformer-based models is presented.

II. Data Description and Preprocessing

While developing an LLM-based conversational AI tool with industry knowledge, the first major challenge was how to integrate our own data into this pre-trained model. We needed to incorporate our data in a way that preserved the model's pre-existing knowledge, while also allowing us to selectively enable or disable different aspects of the model's knowledge base as needed.

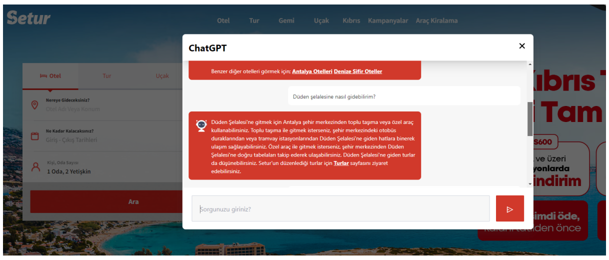
The user is not limited to ask questions about the products that Setur provides. They can ask questions from a tourist perspective: "How can I go to Duden waterfall?" or "What activities can we do in Antalya?". We allow ChatGPT to answer such questions based on its own knowledge. We ensure that ChatGPT responds to questions about the services and products provided by Setur completely based on the information that Setur provides. Fig. 1 presents example screenshots of the conversations obtained during our trials, which demonstrate the traveler assistant's ability to engage in informative and contextually appropriate conversations.

We started our development with GPT-3.5 and GPT-4. For both models, since you could not fit all your data in a system prompt or instant message (OpenAI limitations [11]), we had to pass a filter before putting the big data into GPT. Every filtering in this section is vital to the success of the system, as it will determine the data that GPT will process.

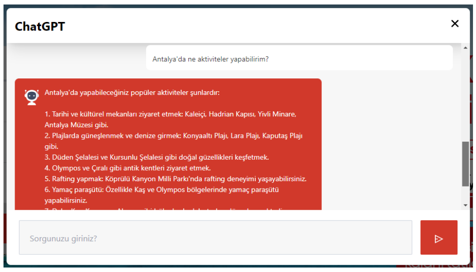
We initially gathered an extensive range of data from various sources, including hotel metadata, catalogs, pricing details, promotional campaign datasets provided by Setur [12], as well as third-party platforms such as TripAdvisor [13], which we have integrated. This diverse dataset has been systematically organized and consolidated into a database, as depicted in Fig. 2. We divided the data into chunks such as location, campaigns, features, summary description to allow GPT to process faster.

III. Methodology

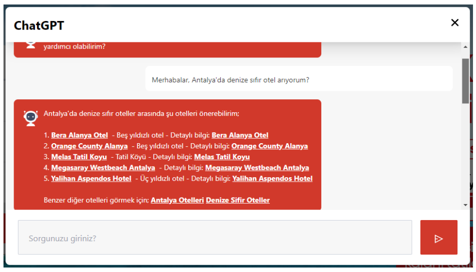
We developed three different models and aimed to create a better model by correcting the deficiencies of the previous model in each version. We compared the results and costs of cloud services to find out the most efficient and cost-effective solution. The advantages and disadvantages of each method are detailed below in the relevant section.



(a)



(b)



(c)

Fig. 1: Example conversations from the Setur Traveler Assistant

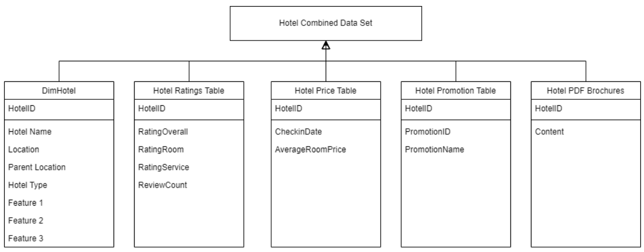


Fig. 2: Data model for latest version.

A. Method 1: pdf data and no code

In the first version, we proceeded by taking all hotel information from the hotel catalog. If the semantic search found a similarity with the query in these PDFs, we would direct the relevant hotel pages to GPT and answer the questions on the GPT side as illustrated in Fig. 3. The problem with this model is that if the query contains a very general keyword, we have to limit the number of incoming pages at a certain point, and we can not necessarily redirect other pages containing answers to GPT. We could not modify the similarity search, so it received irrelevant data. This causes ChatGPT to produce irrelevant or inaccurate responses. That's why we decided to abandon Azure's cognitive search structure.

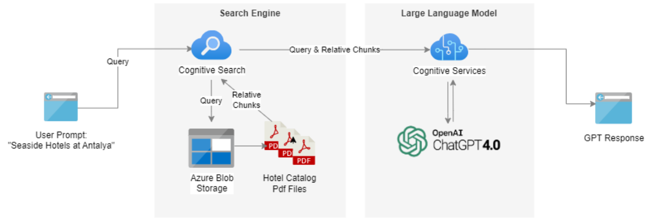


Fig. 3: Sequence diagram showing flow of POC version.

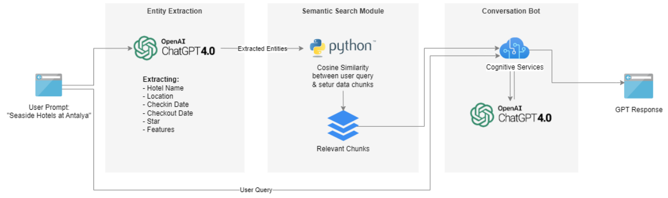


Fig. 4: Sequence diagram showing flow of V1 version including entity extraction and using data chunks instead of pdf files.

Technology Kit:

- GPT-3.5 model
- Azure Cognitive Search
- Power Apps

Advantages:

- Rapid Development
- Easy Adaptation

Disadvantages:

- Restricted Data
- Unable to keep chat history
- Bot reply takes a long time
- Not interfering with the filtering of relevant information, not allowing manipulation
- Difficulty integrating different data sources

B. Method 2: Structured Data and GPT-4

In the second method, we combined the data in PDFs with other data and turned them into XML files format and increased the number of files by dividing this information on a hotel basis. We created 8 data categories for each hotel. These categories include information such as description, location, campaign, features, summary and price.

For instance, when questions pertained solely to the campaign, we enhanced the model's efficiency by retrieving only the relevant information in the campaign line rather than giving an exhaustive list of hotel data as illustrated in Fig. 4. This allowed for the transfer of more comprehensive responses to GPT, resulting in a significant improvement in the model's performance compared to the initial design. Our experiments, conducted using the GPT-4 model, demonstrated substantial advancements in areas such as context limitation, interpretation, and adherence to the provided dataset.

However, when we wanted to take this model live, we had problems at two points. We realized that we would be stuck with the simultaneous usage limits of these services we use from OpenAI. The fact that it answers only 120 questions per minute and token limits showed us that the services we take live will be limited when used very widely.

Technology kit:

- GPT-4 model
- Langchain

Advantages:

- Ability to interfere with filtering of relevant information
- Easy to integrate data sources
- Improve user experience by streaming bot response
- Increasing the answer quality in follow-up questions by keeping the chat history
- Keeping chat logs

Disadvantages:

- Limited quota (20k tokens & 120 api calls per minute)
- High price (input: $0.06/1K$, output: $0.12/1K$)

C. Method 3: Structured Data, Intent Detection and GPT 3.5 Turbo

In the third development method, we had to update the system to work with GPT 3.5 Turbo. When we switched to GPT 3.5, we saw that version 3.5 was insufficient in the issues where version 4 was quite successful, and we had to solve them in different ways.

First of all, the dataset we gave had become much more important than before. When we searched for a hotel by the sea at Antalya in the GPT-4, even if the wrong data came from the semantic, it was doing its own filtering correctly while giving the answer, but GPT-3.5 only brought the hotels that are on the seaside and not in that region. So now more work fell on the semantic search layer and GPT's source data needed to be error free.

Likewise, we have experienced that GPT-3.5 is behind GPT-4 in terms of going out of context. To solve all these, in Fig. 5, we have created a structure to determine the end user's intention and the necessary entities with a GPT-4 working with a simple system prompt that will not exceed the limits. After these intentions and entities were detected, our algorithm managed to perform a more accurate filtering in the semantic layer by calling the relevant services. By detecting 'Out of Context' situations with GPT-4 in the same way, we chose to send a fixed message sentence without leaving it to the interpretation of GPT.

With this version, where we can approach the success rates of GPT-4, we both provided a cost advantage and ensured that the data set coming to GPT was one step more accurate.

Even in the last version we came to, we encountered situations where semantic search was not the desired answer to the problem in the tourism sector. For example, when we searched for hotels with strollers and by the sea in the tests, we encountered situations such as being able to bring only hotels with strollers. This is due to the fact that we cannot assign weight to any word in semantic search. At this point, we continue our research by investigating hybrid methods combining semantic and lexicon searches. We think that if we can direct the search by weighting certain meaning groups in the lexicon search method, where the words are searched as they are, we can improve the results.

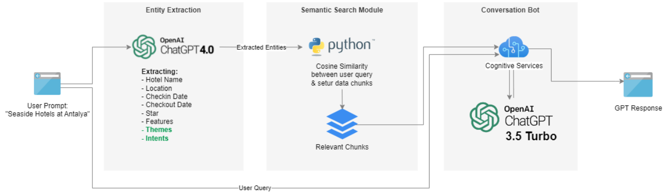


Fig. 5: Sequence diagram showing flow of V2 version including entity extraction, intent detection and using more detailed data chunks.

Technology kit:

- GPT3.5 16k context model
- Langchain

Advantages:

- High quota, responding to more requests (240k tokens & 240 api calls per minute)
- Lower price (input: $0.003/1K$, output: $0.004/1K$)
- Large context window

Disadvantages:

- Less capable model
- Filtering business rules in the code
- Manualization and stricter rules

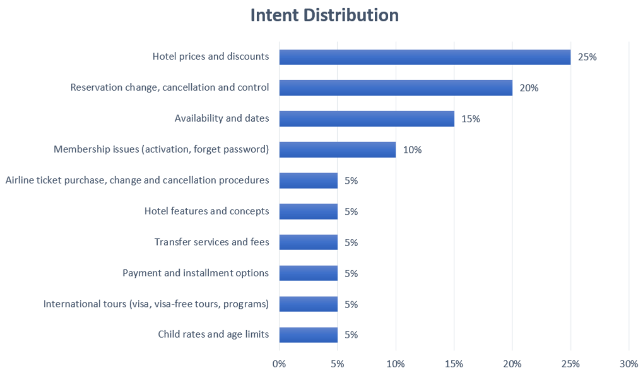


Fig. 6: Intent distribution of queies for 3 months test usage.

After testing the GPT-based travel assistant, we divided the requests from end users into intent groups using GPT in Fig. 6. Thus, we can observe which intent receives more requests. Similarly, analyzing and reporting these logs on a topic-based and semantic basis will be useful for improving the travel assistant in the future. In the test usage logs, questions were mostly asked about hotel prices, discounts, changing reservations, canceling, hotel search and quota.

IV. Conclusions & Potential Applications

In conclusion, although LLMs exhibit remarkable performance in generating human-like (sales representative) dialogues, their end-to-end application remains limited due to system prompt constraints, fine-tuning challenges, and the unavailability of specific models. Moreover, semantic search fails to deliver satisfactory performance when searching filters that require clear answers. To address these issues, we propose a hybrid approach that employs external interventions, the assignment of different GPT agents according to intent analysis, and traditional methods at specific junctures, which will facilitate the integration of domain knowledge into these systems. We strongly believe that the AI research community will develop comprehensive solutions to overcome these limitations, and the integration of such improvements into existing models will significantly advance the state-of-the-art in conversational AI systems, ultimately paving the way for more efficient edge applications.

In addition to the features of GPT, such as summarizing, interpreting, and making inferences, it can be widely used in online commerce sites, sales, or after-sales support points as the conversational AI solution mentioned in this article. It can even replace standard telephone agents in a short time by integrating speech-to-text and text-to-speech technologies during phone calls.

References

[1] Ahmed Hasanein, Abd El-Kafy, and Jermien Hussein. The impact of technology acceptance on the effectiveness of the electronic customer relationship management in hospitality and tourism industry: The mediating role of employees' satisfaction. *International Academic Journal Faculty of Tourism and Hotel Management*, 8(1):36–62, 2022.

[2] Bayad Jamal Ali, Bayar Gardi, Baban Jabbar Othman, Shahla Ali Ahmed, Nechirwan Burhan Ismael, Pshdar Abdalla Hamza, Hassan Mahmood Aziz, Bawan Yassin Sabir, Sarhang Sorguli, and Govand Anwar. Hotel service quality: The impact of service quality on customer satisfaction in hospitality. *Ali, BJ, Gardi, B., Othman, BJ, Ahmed, SA, Ismael, NB, Hamza, PA, Aziz, HM, Sabir, BY, Anwar, G.(2021). Hotel Service Quality: The Impact of Service Quality on Customer Satisfaction in Hospitality. International Journal of Engineering, Business and Management*, 5(3):14–28, 2021.

[3] Nagaraj Samala, Bharath Shashanka Katkam, Raja Shekhar Bellamkonda, and Raul Villamarin Rodriguez. Impact of ai and robotics in the tourism sector: a critical insight. *Journal of tourism futures*, 8(1):73–87, 2020.

[4] OpenAI. Gpt-4 technical report, 2023.

[5] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. Improving language understanding by generative pre-training. 2018.

[6] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

[7] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[8] OpenAI. Models - openai api. [Online]. Available: https://platform.openai.com/docs/models/overview.

[9] OpenAI. Pricing. [Online]. Available: https://openai.com/pricing.

[10] Jawid Ahmad Baktash and Mursal Dawodi. Gpt-4: A review on advancements and opportunities in natural language processing. *arXiv preprint arXiv:2305.03195*, 2023.

[11] Microsoft. Azure OpenAI Service quotas and limits. [Online]. Available: https://learn.microsoft.com/en-us/azure/ai-services/openai/quotas-limits.

[12] Setur. Setur Tourism Agency. [Online]. Available: https://www.setur.com.tr.

[13] Tripadvisor. Tripadvisor api. [Online]. Available: https://www.tripadvisor.com/developers.