

Optimizing access to healthcare information

Julia Lan Bui Xuan¹, Simone Farallo², Michele Salvaterra³

Abstract

This project aims to develop a prototype for the healthcare sector, focusing on optimizing and automating access to information and the available clinics located in Lombardy, Italy. It aims to optimize and automate access to health information, considering challenges related to data privacy and inconsistencies within the healthcare system. The development of the project is based on the LangChain framework, powered by the GPT-3.5 Turbo model. Tailored explanations and support to the user is given by a list of tools that the agent decides to implement based on some task-specific settings provided. Furthermore, an integration with Optical Character Recognition (OCR) technology is envisioned to extract outpatient services directly from electronic prescriptions, with the aim of locating suitable clinics. Several limitations have been recognized regarding the domain of interest, the computational costs of the model related to both prompt and evaluation, output control, and the integration of certain tools, such as OCR technology, which in itself has also criticality in dealing with certain types of files and documents. Although this project tackles complexities to deliver automated and accessible support for citizens, it has laid the groundwork for further advancements and enhancements in digital healthcare, presenting opportunities for progress in the field.

Keywords

LangChain - OpenAI - Chatbots - OCR - Healthcare - Lombardy

^{1,2,3} *Dipartimento di Informatica, Sistemistica e Comunicazione, Università degli studi di Milano-Bicocca, Milano, Italia*

Contents

1	Introduction	1
2	Overview of the domain	2
3	Domain-specific challenges	2
3.1	Problems	2
3.2	Limitations	3
4	Development of the project	3
4.1	Exploratory data analysis	3
4.2	LangChain - Agent and tools	3
4.3	Prompt engineering	4
	Zero-Shot Prompting	
4.4	Extra tool	4
	Automatic OCR Search Tool • Search tool	
4.5	Interface	5
5	Results	5
5.1	Conversation with chatbot	5
5.2	Heuristic evaluation	5
5.3	Extra tools evaluation	7
6	Conclusions	7
7	Further Developments	7
	References	7

1. Introduction

This project aims to develop a prototype for the healthcare sector, focusing on optimizing and automating access to information. More precisely, the main aim consists in providing general information regarding medical analyses and the healthcare sector, in order to provide support to citizens. The medical analyses can be both prescribed or not, by doctors or other healthcare-related professionals. Initially, the idea involved creating a chatbot to assist patients with explanations about diseases, treatments, medications and outpatient services, as well as the available clinics where the services can be done.

However, after conducting extensive research and exploration across national websites and healthcare organizations, it became evident that healthcare information is fragmented. The lack of a uniform database throughout Italy leads to inaccuracies and challenges in accessing information, posing a significant hurdle for the project. Despite guidelines published by the National Health Service, many Italian regions do not adhere to them, further contributing to incongruities and imprecision.

Consequently, in order to address the information gap, the scope of the investigation was narrowed down to outpatient services available without a mandatory prescription in the Lombardy region. This decision is motivated by several factors. Firstly, there is a lack of homogeneous information regarding outpatient services for the whole country; secondly, Lombardy is among the few regions with a more structured

and comprehensive system that tracks available outpatient clinics and provides open-source data. In such a way, the service can be used even by citizens that are not provided with a prescription by the doctor.

Regarding the state of the art [1, 2], it is important to note that there exist many techniques, models and frameworks available for the development of chatbots in the healthcare domain. As a matter of fact, AI-based tools designed to interact with users are becoming increasingly widespread and promising, capable of offering health-related information and support [3].

It is important to note that the absence of a comprehensive and structured knowledge base has been a limiting factor, requiring innovative experimental approaches. The following sections examine in detail the techniques and models used to develop a functional and accurate prototype to facilitate access to health information and services, considering the potential applications as well as their limitations.

2. Overview of the domain

The definition of the domain plays a vital role in the development of the project, as it is essential to highlight that the chatbot operates within a closed domain, exclusively focusing on the healthcare sector. Consequently, the interaction with the chatbot will be confined to matters and information relevant to the healthcare sphere, specifically pertaining to outpatient services.

Outpatient care encompasses therapeutic and rehabilitative services, as well as instrumental diagnostic and laboratory procedures, administered by specialist physicians operating within outpatient clinics and laboratories of the National Health Service (SSN) [4]. However, challenges exist in service delivery due to discrepancies among regions and outdated data. Furthermore, numerous facilities, both private and non-private, can offer services without prescriptions, employing internal systems that provide patients with in-house specialists.

Each service is identified by a national code and precise nomenclature. The nomenclature for services provided by the SSN was approved by the Minister of Health's Decree on July 22, 1996, and subsequently revised in the Ministry of Health's Decree on October 18, 2012. The numerical codes used to identify services are derived from the Italian translation of the ICD-9-CM Classification [5]. Nevertheless, each region adopts a distinct numerical coding system, and nomenclature often varies, resulting in additional challenges.

In conclusion, outpatient specialist care constitutes a pivotal domain within the realm of healthcare provision for patients. Consequently, it is imperative to ensure coherence between nationally established guidelines and the delivery of healthcare services at the regional level. Moreover, the implementation of dematerialized prescriptions presents a significant opportunity to enhance service delivery efficiency and simplify procedures, benefiting both patients and healthcare personnel [6].

3. Domain-specific challenges

3.1 Problems

The following section outline several key problems faced during the development of the AI-based chatbot for healthcare information, such as:

- **Privacy and data protection:** when a user provides medical analysis to the chatbot, it may be considered sensitive personal data from a privacy standpoint. Therefore, it is necessary to ensure compliance with laws regarding the protection of personal data, such as the General Data Protection Regulation (GDPR) in the European Union or equivalent laws in the respective country [7]. A valid legal basis for processing personal data within the healthcare chatbot must be established, such as obtaining explicit and unequivocal consent from the user.
- **Accuracy of Information:** the chatbot should provide accurate and reliable information regarding medical analysis. Utilizing trustworthy and up-to-date sources of medical information is fundamental to ensure that the chatbot delivers correct responses. A disclaimer statement is necessary to clearly indicate that the chatbot is not a substitute for a medical professional and that its responses should be verified by a qualified physician.
- **Legal Responsibility:** establishing a clear understanding of the accountable entity responsible for the information delivered by the chatbot is essential. It is necessary to explicitly state that the chatbot operates as an independent entity and cannot be held legally liable for any consequential damages resulting from the information provided.
- **Data minimization:** applies the principle of data minimization, collecting only the information strictly necessary for the stated purpose of the chatbot. Avoids collecting or processing personal data in excess of the chatbot's specific purposes, but creates constraints that the model may misunderstand
- **Information acquisition:** the availability of complete and uniform information in the NHS is difficult. National Health Service (NHS) guidelines are often not adhered to by regions, limiting the amount of documents that can be used as references.
- **Heterogeneous domain:** the domain being explored is extremely vast and heterogeneous. Healthcare is closely linked to medicine and biomedicine, but defining its boundaries precisely is a challenging task. The nuances and interconnections among different areas make it difficult to reduce the domain to a fixed definition.
- **OpenAI generative model:** the model utilized, although currently one of the most powerful ones available as of June 2023, is still a generative probabilistic

model. This entails that in situations where specific knowledge is absent, invented or incompletely accurate responses might be generated by it. Careful consideration of these limitations during the utilization of the model is important, and efforts should be made to mitigate them through further developments and improvements.

3.2 Limitations

The following points highlight several constraints that hindered the project's progress and prevented its further advancement, including:

- **Token usage:** the responses generated by the chatbot may be incomplete as the utilization of available tokens approaches its limit. This restriction arises due to the finite number of tokens the chatbot can process within its computational framework.
- **Optical Character Recognition:** it is crucial to acknowledge that OCR may encounter challenges when dealing with images of poor quality, handwritten text, or complex layouts, which can adversely impact its accuracy [8].
- **Model costs:** the GPT-3.5-turbo model exhibits notable costs, when attempting to narrow down the domain through the prompt. This limitation has been observed during evaluations [9].

4. Development of the project

To gather comprehensive information on the healthcare domain, this study relies on two primary sources: Wikipedia and Google. The selection of these sources was driven by the absence of a reliable Italian database and the consideration of cost-effective implementation. Furthermore, data acquired from the Lombardy region is utilized to identify the facilities that provide the outpatient services.

In order to achieve the objectives proposed, potential tools such as LangChain, OpenAI and OCR are used.

LangChain is a framework for developing applications powered by language models (LM), designated around the principles of data-awareness and agents. It is defined by the following modules:

- models: supported model types and integrations;
- prompts: prompt management and optimization;
- memory: state that is persisted between calls of a chain or agent;
- chain: sequences of calls;
- agent: chain in which a LLM, given a high-level directive and a set of tools, repeatedly decides an action, executes the action and observes the outcome until the high-level directive is complete.

Although Large Language Models (LLMs) are a core component of LangChain, the latter one is not a provider of LLMs. Nevertheless, it allows to integrate different LLM providers, such as OpenAI [10].

Another important tool is Optical Character Recognition (OCR) [8]. It is a technology that enables the extraction of text from images or scanned documents, converting it into machine-readable text.

4.1 Exploratory data analysis

In the initial phase of the project, several datasets containing information about healthcare facilities and services in the Lombardy region were collected and cleaned. After ensuring data accuracy and completeness, the datasets were merged into a single dataset. During the preliminary data analysis, cleaning operations such as handling missing values and removing duplicates were performed. Additionally, format standardization was applied to ensure consistency across data. Subsequently, an exploratory data analysis (EDA) was conducted to gain deeper insights into the combined dataset. The following dataset will serve as a knowledge base for searching outpatient facilities that offer specific services.

4.2 LangChain - Agent and tools

The support to the user regarding explanations or additional information is provided by a chatbot, as it allows to easily resolve many types of queries and issues while reducing the need for human interaction. The chatbot is developed with the aim of supporting only queries related to healthcare. As a matter of fact, it will not answer to questions not linked to the domain under study.

The chatbot is built with the use of **LangChain**. It presents four types of agents available, such as zero-shot-react-description, react-docstore, self-ask-with-search and conversational-react-description. For the purpose of the project, zero-shot-react description **agent** type was implemented, as it uses the ReAct framework to determine which tool to use based solely on the tool's description. The temperature is set to 0, since lower temperature values encourage more factual and concise responses. To control computational costs, a maximum limit of 5 iterations is set for the agent's search for answers. Additionally, a chat history memory is implemented, which remains active and relevant solely for the duration of the current session. This ensures that the memory exclusively retains information pertaining to the ongoing session.

A list of **tools** is provided to the agent, each defined by a name, function and description which helps the agent to determine the tool use. More precisely, three tools were used, with the help of customization and priority emphasis:

- Wikipedia: being the largest and most-read reference work in history, it is defined as a tool to prioritize. It is described as a search engine to answer only healthcare related queries
- Google Serper: low-cost Google Search API that can be used to return results data from Google Search. It is

described as a search engine to answer only healthcare related queries and used especially when current or recent events are queried

- “No Answer”: customised tool used to answer questions not related to healthcare queries. The answer of the agent is defined by a function that returns the following statement: “I am not allowed to answer to this question as it is not related to the healthcare domain.”

The **model** integrated is GTP-3.5 Turbo, obtained through the OpenAI key. The input to the model is defined by **prompts**. This input is often constructed from multiple components, such as:

- Instruction: a specific task or instruction for the model to perform;
- Context: external information or additional context that can steer the model to better responses;
- Input Data: the input or question to find a response for;
- Output Indicator: the type or format of the output.

4.3 Prompt engineering

Prompt engineering is an emerging field focused on the development and optimization of prompts to effectively leverage language models across diverse applications and research domains. A lot can be achieved with simple prompts, but the quality of results depends on how much information is provided and how well-crafted the prompt is. As a matter of fact, designing prompts is an iterative process that requires a lot of experimentation to get optimal results. Various techniques, including zero-shot, few-shot, or chain-of-thought prompting, among others, are available.

4.3.1 Zero-Shot Prompting

Large Language Models today, such as GPT-3, are tuned to follow instructions and are trained on large amounts of data; so they are capable of performing some tasks “zero-shot” [11]. This type of prompt does not provide the model with any examples, as it happens, for instance, in few-shot prompting where in-context learning is enabled to provide demonstrations aimed to steer the model to better performance. On the other hand, when the previously mentioned techniques do not provide satisfactory results, it is advised to fine-tune the models or experiment with more advanced prompting techniques, such as chain of thought prompting that enables complex reasoning capabilities through intermediate reasoning steps.

For the development of this project, zero-shot prompting techniques were utilized, yielding satisfactory results. However, due to computational costs, in-depth exploration of other techniques was limited. More precisely, starting with simple prompts, more elements and context were added to improve the answers provided by the agent. Finally, an effective prompt was constructed by instructing the model the objective to achieve, that is answering with maximum 2/3 sentences

to explain concepts related to healthcare. Furthermore, the context was also defined, by specifying to the agent that it is an AI bot providing general information, unable to make bookings for outpatient services.

4.4 Extra tool

To gather information pertaining to facilities, the dataset provided by the official healthcare website of Lombardy is considered. The objective is to develop an extra tool capable of receiving a prescription as input and generating a list of facilities that offer specific services as output.

It is important to note that the GPT-3.5 model, being designed to accept input solely in the form of strings, cannot directly process other types of objects in the Python programming language. For this reason, the approach considered involved the development of a tool to address this limitation, the integrating the OCR function as an external tool.

The development of the OCR tool aimed to enable the chatbot to process various input formats, including images, PDF documents or simple queries, for the requested analyses. The OCR tool would extract text from these inputs, allowing for further utilization of the extracted analyses as prompts when interacting with the model.

Furthermore, among the available tools in LangChain, there is one that allows to write queries in the string format and extract information from a dataset, defined as Agent CSV [12]. This agent would use this prompt for the csv interaction where the dataset of ambulatory clinics is stored. However, it was observed that the answers generated by the agent lacked precision as they were produced by the model itself rather than directly extracted from the dataset. Additionally, due to computational costs and constraints, further exploration of interactions with the dataset had to be limited.

4.4.1 Automatic OCR Search Tool

After exploring several approaches, a tool that operates independently of the LangChain framework was implemented. This tool is designed to accept prescription inputs in PDF or PNG formats and provide a list of facilities capable of analyzing the output. However, it is important to note that the codes of outpatient services exhibit inconsistent formatting across regions and facilities (e.g., 90.50.67, 905067, and 0000905067000 all refer to the same information). Moreover, there may be slight variations in the nomenclature (e.g., Emocromo and Emocromo: Hb, GR, GB, HCT, PLT, IND all indicate the same service). Consequently, relying solely on code and nomenclature filtering for dataset refinement is not viable.

Therefore, in order to develop this tool, technologies such as OCR and fuzzy matching were employed. The detailed extraction of information from decentralized prescriptions [6] were possible due to OCR potential combined with regex functions. Fuzzy matching was implemented to match the previously extracted outpatient services with the dataset of all the available clinics located in Lombardy. Fuzzy matching [13] is a technique used to identify and compare similarities

between strings or sequences of characters. It is particularly useful when dealing with data that may contain errors, typos or variations in formatting. Fuzzy matching algorithms aim to find matches even when the input data is not an exact match. In fuzzy matching, a distance metric is calculated to determine the degree of similarity between two strings. In this case, the Levenshtein distance is used, measuring the difference between two sequences. More precisely, it considers the minimum number of single-character edits (insertions, deletions or substitutions) required to change one word into the other. The score indicates how closely two strings match each other. The search precision can be adjusted by a threshold parameter, which has been set to 0.90. Finally, the tool generates an output in the form of a new DataFrame, providing a comprehensive list of facilities capable of conducting the outpatient services.

4.4.2 Search tool

Taking into account the prior implementation of the tool, a subsequent development effort was undertaken to return structures by taking as input the name of the outpatient service. The Fuzzy Matching tool was configured with a precision level of 0.0.

4.5 Interface

A user friendly interface was generated with Gradio, an open-source Python library, used to build machine learning and data science demos and web applications. As shown in Figure 1, the interface aims to provide answerws to healthcare related questions. Figure 2, instead, shows an interface designed to provide users with a list of clinics in Lombardy, considering their preferences for location and type of facility.

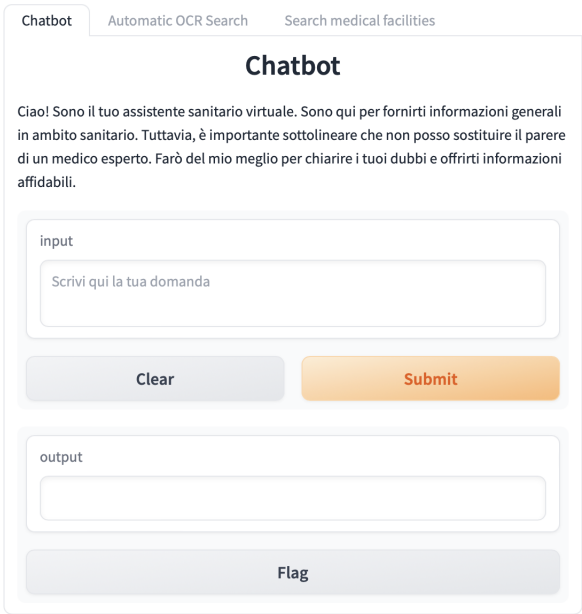


Figure 1. Chatbot - to answer questions about health-related topics.

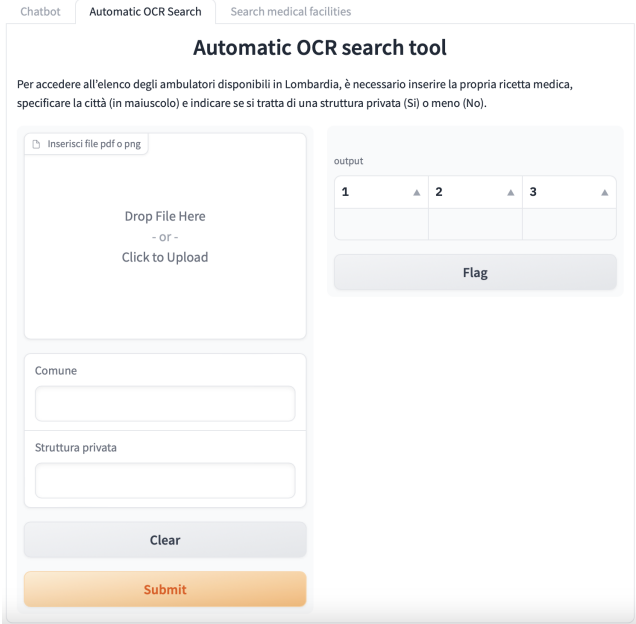


Figure 2. User friendly interface to access clinics located in Lombardy.

5. Results

5.1 Conversation with chatbot

The results obtained with LangChain are showed in the below tables. In Table 1, the user poses a question about hemochrome, and the agent provides an answer utilizing the Wikipedia tool, correctly prioritizing the use of this tool. In Table 2, the conversation between the user and the bot continues, with the user inquiring about the typical range of hemochrome for healthy women. Initially, the agent attempts to find an answer with Wikipedia tool. However, when it fails to find a suitable answer, it switches to the Google Serper tool, which successfully provides the required information. In Table 3, the agent employs the Wikipedia tool to furnish the user with additional information regarding haemochrome analysis. Lastly, in Table 4, the user poses a question unrelated to the healthcare domain. As the agent is programmed to exclusively support healthcare-related queries, it employs a customized tool to decline answering.

5.2 Heuristic evaluation

A heuristic evaluation was conducted to assess the chatbot answers which serves as a guideline containing design principles used to evaluate user usability and experience. The evaluation employed a usability violation scale, categorizing issues as minor, moderate, significant or critical. The evaluation of the chatbot considers the following aspects:

- **Clarity and comprehensibility:** the chatbot should provide clear and easily understandable responses in a simple and unambiguous manner. This evaluation represents a moderate issue; considering the healthcare sector, the explanations provided may contain technical

Role	Dialogue
Prompt	What is hemochrome and why do I need to have hemochrome blood tests?
GPT3	Hemochrome is a laboratory test on blood that determines the amount of corpuscular elements (leukocytes or white blood cells, erythrocytes or red blood cells, and thrombocytes or platelets), the levels of hematocrit (HCT) and hemoglobin (Hb), as well as other blood parameters. It is important to do this test because it can help diagnose diseases such as juvenile myelomonocytic leukemia.

Table 1. Action: Wikipedia Search

Role	Dialogue
Prompt	What is the typical range of values for hemochrome in healthy women?
GPT3	Normal hemochrome values for a healthy woman are: hemoglobin (Hb) 12-16 g/dl, erythrocytes/mul(RBC) 4.2-5.4 x 10 ⁶ , mean corpuscular volume (MCV) 81-99 fl, mean corpuscular hemoglobin (MCH) 26-32 pg.

Table 2. Action: Google Search

Role	Dialogue
Prompt	I am a woman with hemochrome value over 50%. What does it mean?
GPT3	A CBC value over 50% indicates a condition of polycythemia, which is an excessive production of red blood cells. This can be caused by certain diseases such as chronic myeloid leukemia, which requires specific treatment.

Table 3. Action: Wikipedia Search

Role	Dialogue
Prompt	Recommend the best pizzeria in Milan.
GPT3	I cannot answer this question because it is not related to healthcare.

Table 4. Action: No answer

language, which can be difficult for a non-expert user in the medical field to comprehend. Additionally, the chatbot cannot exceed the available token limit, resulting in incomplete responses.

- **Coherence and consistency:** the chatbot should maintain a consistent style in its responses and interactions with users. The interface should be uniform and predictable to create a cohesive user experience. This heuristic is considered to be respected.
- **Feedback and confirmation:** the chatbot should provide appropriate feedback to users, indicating that their requests have been understood and that the system is processing a response. Furthermore, it should confirm user actions to prevent errors or misunderstandings. This heuristic is considered to be respected since the system displays a loading bar while processing a response, implying that it is working on generating an output.
- **Ease of navigation:** the chatbot should guide users through interactions in an intuitive and seamless manner. Navigation and option structures should be well-organized and easy to follow. This heuristic is considered an respected.
- **Responsiveness and timeliness:** The chatbot should be responsive and provide timely responses to users. Prolonged waiting times or lack of responses can negatively impact the user experience. This heuristic is considered as not respected, representing a moderate issue. As a matter of fact, the chatbot's response times range from 20 to 40 seconds.
- **Adaptability and flexibility:** the chatbot should be able to adapt to various user requests and contexts. It should intelligently handle unexpected questions or inputs and provide appropriate responses. This heuristic is considered as respected.
- **Errors and recovery:** the chatbot should prevent errors or provide means to correct them. It should be capable of recognizing common user errors and offer suggestions or alternatives to overcome issues. This heuristic is considered as not respected, representing a moderate issue. As a matter of fact, when a document in the wrong format is uploaded to the Automatic OCR search tool, there is no alert of the error and no suggestions are provided to the user to solve it.
- **User satisfaction:** the chatbot should strive to meet user needs and provide an overall positive user experience with the goal of creating a pleasant and helpful dialogue for the user. This heuristic represents a moderate issue. The chatbot has a memory, allowing a continuous conversation. However, being limited to the healthcare sector, its responses are of technical nature.

5.3 Extra tools evaluation

In the subsequent section, an evaluation is conducted on the OCR tool for prescription extraction and fuzzy matching techniques for ambulatory clinics matching.

With respect to the OCR tool, no issues were encountered. Regardless of the input provided by prescriptions, the desired results were consistently obtainable.

The Fuzzy matching tool, instead, was evaluated in the following methods: by collecting electronic prescriptions regarding outpatient services, while respecting privacy and asking for consent from users, and by filtering the dataset with respect to strings corresponding to the nomenclature of the services, including also some typos (e.g., *emocrom*, *emocro*, *calcitonIN A*) and by changing the threshold measures. The results were always satisfactory.

Given a prescription, as shown in Figure 3, the OCR tool returns the output provided in Table 5

TIPOLOGIA PRESCRIZIONE(S,H):		ALTRQ: PRIORITA' PRESCRIZIONE(U,B,D,P): D (Differibile)	
PRESCRIZIONE		GTA	NOTA
91.27.1 (4481.154) - VIRUS VARICELLA ZOSTER ANTICORPI IGG - SANGUE		1	---
PRIMO ACCESSO			

Figure 3. Prescription example

The prescription refers to the service called "VIRUS VARICELLA ZOSTER ANTICORPI (E.I.A.)". The result was filtered for the city of Bergamo and for private facilities only.

Nome struttura	Centro biomedico	Lab. Bio analitico Delta
Codice	91271	91271
Branca regionale	Biologia-virologia	Biologia-virologia
Comune	Bergamo	Bergamo
Indirizzo	Rotonda dei Mille 3/A	Via Clara Maffei 14
Telefono	035222332	035221565
Struttura	Sì	Sì
Prezzo	7.75	7.75

Table 5. Output of Automatic OCR Search

6. Conclusions

In conclusion, this project involved the development of a prototype to support the Italian healthcare system, based on optical character recognition (OCR) and LangChain framework. The primary objective consisted in providing user support for a better understanding of the healthcare sector and in providing a facilitated access to information about the available clinics. However, throughout the project implementation, several challenges and constraints emerged, affecting its progress.

A significant limitation consisted in the impossibility of the integration of the OCR system withing the LangChain framework, thus developing a chatbot capable of providing

both information about prescriptions and clinics. Despite various attempts to amalgamate both technologies, technical obstacles and the heterogeneous nature of healthcare data, hindered the creation of a cohesive and functional interface between the two components. As a result, the OCR system and the chatbot remained separate entities. Nevertheless, the project effectively accomplished its objectives.

Furthermore, it is noteworthy that the utilized GPT-3.5 Turbo model, while robust, remains a probabilistic generative model, prone to generate responses that may not be entirely accurate in the absence of specific information. It is essential to take into consideration this aspect and take measures to ensure the accuracy of the information provided when interacting with the chatbot.

Finally, it was recognized that the healthcare domain encompasses vast complexity, with variations in definitions and guidelines across different regions. The absence of standardized reference documents posed a challenge in acquiring dependable and consistent information.

In conclusion, although the project grappled with various obstacles and constraints, it has laid the groundwork for further advancements and enhancements in digital healthcare, presenting opportunities for progress in the field.

7. Further Developments

In addition to the successful achievement of its objectives, there are several promising opportunities for further development and improvement within the project. By exploring these avenues, the project can continue to evolve and offer even greater functionality and value. Some potential areas for future development and expansion can be the following:

- Integrate extra functions and build ad hoc tools
- Build a valid, accurate and expert-verified knowledge base
- Extend the domain to the whole country
- Try different prompt techniques to define the domain in more detail
- Try different models and techniques, including building ad hoc embedding, evaluating them individually
- Use more advanced and more detailed evaluation techniques
- Develop a more dynamic chatbot, making the conversation more pleasant and smooth
- Make heuristic evaluations regarding the user experience part.

References

- [1] A critical review of state-of-the-art chatbot designs and applications. <https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434>.
- [2] A state-of-the-art open source chatbot. <https://ai.facebook.com/blog/state-of-the-art-open-source-chatbot/>.
- [3] Chatbots as conversational healthcare services. https://www.researchgate.net/publication/346867238_Chatbots_as_Conversational_Healthcare_Services.
- [4] Servizio sanitario nazionale (ssn). <https://www.salute.gov.it/portale/lea/dettaglioContenutiLea.jsp?area=Lea&id=5073&lingua=italiano&menu=vuoto>.
- [5] Nomenclatura nazionale. https://www.salute.gov.it/portale/temi/p2_6.jsp?id=1767&area=programmazioneSanitariaLea&menu=lea.
- [6] Ricetta dematerializzata. <https://www.salute.gov.it/portale/ehealth/dettaglioContenutiEHealth.jsp?lingua=italiano&id=2514&area=eHealth&menu=vuoto>.
- [7] Regolamento gdpr. <https://www.garanteprivacy.it/documents/10160/0/Regolamento+UE+2016+679.+Arreccchito+con+riferimenti+ai+Considerando+Aggiornato+alle+rettifiche+pubblicate+sulla+Gazzetta+Ufficiale++dell%27Unione+europea+127+del+23+maggio+2018>.
- [8] Ocr. https://en.wikipedia.org/wiki/Optical_character_recognition.
- [9] Priceing openai. <https://openai.com/pricing>.
- [10] Langchain. <https://python.langchain.com/en/latest/index.html>.
- [11] Finetuned language models are zero-shot learners. <https://arxiv.org/pdf/2109.01652.pdf>.
- [12] Langchain on csv. <https://python.langchain.com/en/latest/modules/agents/toolkits/examples/csv.html>.
- [13] Fuzzywuzzy. <https://pypi.org/project/fuzzywuzzy/>.
- [14] Chat with your healthcare documents: Build a chatbot with chatgpt and langchain. <https://pub.towardsai.net/chat-with-your-healthcare-documents-build-a-chatbot-with-chatgpt-and-langchain-6b910eb6ca14>.
- [15] Livelli essenziali di assistenza (lea). <https://www.salute.gov.it/portale/lea/menuContenutoLea.jsp?lingua=italiano&area=Lea&menu=leaEssn>.
- [16] Regione lombardia. <https://prenotasalute.regione.lombardia.it/sito/>.