

Enhancing healthcare with intelligent environments: Integrating medical knowledge into GPT for advanced medical personal chatbots

Primož Kocuvan ^{a,*}, Matic Zadobovšek ^b and Matjaž Gams ^c

^a *Jožef Stefan Institute, Ljubljana, Slovenia*

^b *University of Ljubljana, Ljubljana, Slovenia*

^c *Jožef Stefan Institute, Ljubljana, Slovenia*

ORCID: <https://orcid.org/0000-0002-5747-0711>

Received 4 July 2024

Accepted 19 August 2024

Abstract. ChatGPT has shown high performance in medical diagnosis, with various enhancement strategies proposed. However, national-level applications remain limited. This study explores integrating a personal medical chatbot into home environments nationwide, using knowledge from the Insieme platform, a robust electronic and mobile health system developed through an Italian-Slovenian project. This integration provides verified medical information, online support from healthcare professionals, and interactions with a virtual assistant powered by advanced natural language processing technologies. Detailed presentations of the Insieme platform and HomeDoctor application introduce innovative solutions for smart city ecosystems, potentially transforming national healthcare by enhancing patient care and optimizing workflows. The approach is based on deploying ChatGPT within the information and knowledge from the Insieme platform, using word embeddings and vector databases for efficient data retrieval and context-aware responses. This integration aims to reduce the burden on healthcare professionals, particularly in areas with workforce shortages, by providing 24/7 accessible and accurate medical information. The national-scale chatbot integration supports multilingual interactions, ensuring accessibility for non-native speakers, using country-specific knowledge. Preliminary studies show high user satisfaction and positive healthcare impacts, demonstrating the potential of integrating advanced AI technologies into national healthcare infrastructures and offering a blueprint for future medical smart-city developments.

Keywords: Virtual assistants, vector databases, word embeddings, GPT-4, natural language processing

1. Introduction

According to the WHO report “Health and Care Workforce in Europe: Time to Act,” all countries in the European region are facing challenges due to population, health and care workforce ageing [36]. Many countries are already facing shortages of medical workers, while on the other hand, the quality of artificial intelligence (AI) systems in

* Corresponding author. E-mail: primoz.kocuvan@ijs.si.

medical fields, such as diagnosis, is improving rapidly and consistently. Even in China, between 1998 and 2016 patient visits per physician increased by 135 percent, and inpatient admissions per physician rose by 184 percent. Similar significant increases demonstrate a dramatic rise in the workload of physicians in health institutions worldwide, highlighting the extent of overload among medical staff [11]. AI technology generally still cannot replace medical professionals directly, but it can help reduce their workload [36].

Besides the quality of replies, one of the key features of systems like ChatGPT is their ability to understand and process complex queries in natural language, thus making it more intuitive for users without medical training. By providing immediate, AI-driven responses, ChatGPT can effectively bridge the gap in primary healthcare information. Continuing from the initial integrations of ChatGPT with medical knowledge, we delve deeper into the specific applications and benefits of this technology by integrating national medical knowledge into GPT.

In general, the advantages of integrating LLMs into electronic health platforms are:

- Better accessibility – Patients can ask their medical questions at any time, regardless of the day or hour.
- Efficient data processing – LLMs can process data quickly and efficiently. For example, each patient has a unique medical history. LLMs can learn and understand a patient's medical history, including prescribed medications, illnesses, and allergies. This information can be beneficial for doctors. LLMs can answer medical questions for non-native speakers.

The disadvantages are:

- Data security – LLMs might have access to each patient's medical records, which can be risky in the event of potential data breaches.
- Cost of OpenAI API – If ChatGPT is used via API for a large number of patients, it can be costly for the healthcare provider. A healthcare provider using an open-source LLM requires appropriate computer infrastructure, which also incurs costs.

In addition to integrating GPT with national medical knowledge, we explore other aspects like the use of ChatGPT in patient education and health literacy improvement. By offering personalized, easy-to-understand explanations of medical conditions, treatments, and health tips, the platform may play a significant role in empowering patients to take charge of their health and providing medical information on national services. This aspect is crucial in home and preventive medicine, where informed patients are more likely to engage in correct healthcare and health-promoting activities and adhere to treatment plans [32].

Another aspect covered in this paper is the potential of ChatGPT in assisting healthcare professionals. The AI can serve as a support tool for doctors and nurses, providing quick access to medical literature, drug information, and case studies, thus enhancing the quality of care provided to patients [8]. This can be particularly beneficial in high-pressure environments where time and resources are limited, ensuring that healthcare providers have the information they need at their fingertips.

Finally, we present the results of preliminary studies conducted to assess the effectiveness of ChatGPT in the Insieme platform for a home personal doctor application, i.e., using mobile phones and home sensors connected. These studies focus on user satisfaction, the accuracy of information provided, and the impact on healthcare outcomes [26]. The results indicate a promising future for AI in healthcare, with potential applications extending beyond the current scope of the Insieme platform. We conclude by discussing future developments, including the integration of more advanced AI capabilities and expanding the reach of the platform to more users and healthcare settings. This exploration sets the stage for a new era in digital health, where AI becomes a fundamental component in delivering patient-centred, efficient, and accessible healthcare national services.

In summary, this work investigates the possibilities of using a large language model (LLM) integrated with a dedicated e-health platform to inform patients about healthcare-related topics that can be used in smart-city societies. The structure of the paper is as follows: After Section 1 Introduction, Section 2 provides Background and Section 3 provides an overview of related works. A medical Insieme platform representing the basic additional knowledge for our integrated system is described in Section 4. The major contribution of this study is presented in Section 5, followed by Section 6, Conclusions.

2. Background

Integrating LLMs into healthcare platforms has been explored in various ways, focusing on different aspects of healthcare delivery and administration. One way of using generative artificial intelligence is to support the mental health of patients. Many applications in the fields of psychology and psychiatry, such as chatbots, help patients cope with depression, anxiety, and other mental health issues. For example, VOS: Mental Health Therapy [33] and Wysa: Anxiety Therapy Chatbot [37] are among the many available options.

GPT-4, first introduced in March 2023 and later upgraded to Omni, represents a substantial advancement in natural language processing and AI in general. It is particularly effective in answering questions, generating text, and translating languages. Compared to earlier models, newer versions of GPT-4 offer improved reliability, accuracy, and better management of user commands, such as specifying the style of generated responses. Extensive testing [24] across various exams and knowledge tests from different fields has shown that GPT-4 often achieves results comparable to those of humans. It also demonstrated strong performance in several medical tasks [5,7,27–29]. Despite the advantages of using ChatGPT and its high accuracy in clinical decision-making [13], it may generate outputs that lack context, accuracy, and understanding of the nuances of medical sciences and language and might be biased [6]. However, to the date of the paper submission, the authors have not found any national medical application of GPT that operates 24/7 for all citizens.

3. ChatGPT in medicine

3.1. Soft skills and empathy

Our experiment began with several tests utilizing open-source AI chatbots, such as the Bot for waiting queues and the JSI assistant, that were both developed at JSI and integrated with the Insieme platform. Unfortunately, these chatbots did not achieve the desired level of performance. Comparisons with the default ChatGPT-4, even without additional knowledge from the Insieme platform, did not show any significant advantages. Consequently, we proceeded to integrate GPT-4 with the knowledge base of the Insieme platform.

The primary motivation for this project stemmed from reports indicating that when compared to human responses, even the default version of GPT-4 generated quite elaborate answers with high quality and empathy [2], as illustrated at the ophthalmology exam, and in question-answering as presented in Fig. 1.

Table 1 illustrates the superior performance of GPT-4 compared to other LLMs in the US Medical Licensing Examination (USMLE) [3] tests conducted in Spring 2024, as documented in the Master thesis by Dragan Gostimirović (in Slovene language).

These results align with the findings of [2], which evaluated the performance of GPT-4 and ChatGPT in the USMLE soft skills test, covering areas such as professionalism, communication, and ethics. The study revealed that GPT-4 outperformed both ChatGPT and human participants, achieving a 90% accuracy rate, while ChatGPT demonstrated a 62.5% accuracy rate.

In addition to GPT-4, several other AI applications, such as question-answering methods, have provided reasonable medical information [35]. However, GPT-4 seems to outperform the competition, particularly in terms of its generality and generative capabilities [5,18,19]. This superiority in generating contextually relevant and coherent responses has solidified (a version of) GPT-4 as the foundation for our virtual assistant initiative.

AI vs human doctors [2]: ChatGPT Outperforms Physicians in Providing High-Quality, Empathetic Healthcare Advice – SciTechDaily: This article reports on a study that compared the quality and empathy of the responses of ChatGPT and physicians to real-world health questions. The study found that healthcare professionals preferred AI responses to those of physicians 79 % of the time, citing higher quality and empathy. In terms of the USMLE tests, GPT achieved over 90 % accuracy.

Comparison of GPT-3.5, GPT-4, and human user performance on a clinical decision support system – Nature [12]: This article reports on a study that compared the performance of GPT-3.5, GPT-4, and human users on a clinical decision support system (CDSS) that provides diagnosis and treatment recommendations for eye diseases. The study found that GPT-4 achieved the highest accuracy and efficiency, followed by GPT-3.5 and human users.



Fig. 1. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. Source: JAMA Internal Medicine, 183, April 2023.

Table 1

USMLE tests on several LLMs in Spring 2024. Source: Master thesis of Dragan Gostimirović

GPT-3.5	Bard	Copilot	GPT-4
74.71	56.32	87.36	90.80

The conclusion drawn from these studies suggests that we are progressing towards developing and implementing artificial intelligence medical systems equipped with soft skills and empathy, performing at a level close to human physicians.

3.2. Personal data processing

Convention 108, established in 1981, had a huge impact not just in Europe but globally. The Convention set forth fundamental principles for the protection of personal data and the handling of special categories of data, principles that have since been adopted universally. Notably, several countries outside of Europe have ratified the Convention, highlighting its global influence [9]. The principles enshrined in Convention 108 are echoed in modern data protection laws, such as the General Data Protection Regulation (GDPR) of the European Union.

Under the GDPR, health data is classified as a special category, which receives increased protection due to its sensitive nature. Article 9 of the GDPR explicitly outlines the conditions under which the processing of health data is either prohibited or allowed, ensuring that such data is handled with the utmost care. This legal framework not only governs data handling within the EU but also imposes strict regulations on the transfer of personal data to third countries or international organizations. These rules are designed to safeguard personal data and ensure that its protection is not compromised once it leaves the EU's jurisdiction [16].

While basic GPT already enables quality performance, several improvements are being introduced. In this study, we enabled the user to inquire about data from our Insieme platform and other relevant medical documents. In this way, the language model is separated from the knowledge base, allowing the user to communicate with the given documents, and use only information found within the supplied documents to generate the answer, ensuring the most relevant response for the user. With this approach, it is easy to add new sources of information and adapt the model for specific tasks without training the existing model, which is time-consuming and computationally demanding.

Furthermore, users can provide as much information as they wish, and all data is erased after the session ends. The process is described in more detail below.

4. Insieme

The Insieme platform was selected for integration with GPT, with its main Web page presented in Fig. 2 and other pages (example of cancer) in Fig. 3, and Fig. 4. The medical and medical-related knowledge in Insieme was recently introduced in collaboration with Slovenian and Italian partners as part of the cross-border ISE-EMH project [25]. It is equipped with a user-friendly interface that allows users to easily access helpful healthcare information from a single website, either by manual search or by the Insieme search. Insieme is the successor to the national Electronic and Mobile Health (eHealth) project, which involved collaboration among 15 partners. Additionally, the platform has been influenced by insights gained from examining EU healthcare platforms, particularly those focused on elderly care [17].

The Insieme platform is built upon three core ideas:

- Gather all medical and medical-related national and global information relevant to an average user needing medical help.
- Provide similar information as “Dr. Google,” but only those that are intensively verified by medical experts and concentrated on the needs and possibilities of the local population.
- The insieme information and knowledge should be human- and computer-readable.

In our informal tests, an average user of Insieme was able to find most of the relevant basic medical information within a few minutes in approximately 90 percent of the cases, while in around 10 percent of the cases, users failed to do so. This information includes details on applications, services, medications, and more. Informal comparisons have also shown that, for the majority of queries, users either found significantly less relevant information in the same amount of time or required approximately 5–10 times longer to locate most of the relevant information using

The screenshot shows the homepage of the Insieme platform. At the top, there is a yellow navigation bar with links for 'HELP', 'MORE', and 'LOGIN', along with language selection options: 'SLO-ENG', 'SLO', 'ITA', and 'ITA-ENG'. On the left side, a dark blue sidebar contains a 'List of services' menu with categories such as 'Ageing', 'Applications', 'Dermatology', 'Oncology', 'Hormonal disorders', 'Mental illness and cognitive impairment', 'Related platforms', 'Respiratory diseases', 'Digestive tract and metabolism', 'Infections, coronavirus', 'Neurological diseases', 'Urology', 'Gynecology', and 'Pediatrics'. The main content area is white and features a 'Welcome to Insieme, the Electronic and Mobile Health platform!' heading. Below this, a brief description states: 'The Insieme platform was developed within the cross-border project ISE-EMH and connects Italian and Slovenian partners. On this platform you can get online information and human support in the field of electronic and mobile health.' A search bar with a 'SEARCH' button and an 'Advanced search' link is provided. The 'Online human help' section includes a description: 'Use live chat to contact medical professionals, ask questions, and get medical advice from qualified health specialists.' At the bottom of this section, it shows 'Call centers' and 'Active users' with a status message: 'Currently there is no active contacts'.

Fig. 2. The Insieme platform webpage is available at <https://www.ise-emh.com>.

Interreg
ITALIA-SLOVENIJA
ERDF

HELP MORE SLO-ENG SLO ITA ITA-ENG LOGIN

List of services

- Ageing
- Applications
- Dermatology
- Oncology
 - Rehabilitation
 - General Information
 - Prevention
 - Urogenital Cancers
 - Gastrointestinal Cancers
 - Other Cancers
 - Skin Cancers
 - Gynecologic Cancers
- Hormonal disorders
- Mental illness and cognitive impairment
- Related platforms
- Respiratory diseases

Gastrointestinal Cancers

Stomach (Gastric) Cancer

Stomach cancer is a type of cancer localized in the wall of the stomach. Usually, people aged between 60 and 80 are affected. It is the 3rd leading cause of death due to cancer worldwide, which might be attributed to its late diagnosis and treatment. The symptoms appear relatively late in the course of the disease, when there is a high chance that the cancer cells have already spread. The most common symptoms include feeling full after eating small amounts of food, heartburn, indigestion, nausea, stomach pain, unintentional weight loss and an enlarged painless lymph node in the supraclavicular region (Virchow's node) or the left underarm area. If you experience these symptoms, you might want to see your general practitioner, who will refer you to an oncologist or gastroenterologist. It is important to keep in mind that the aforementioned symptoms are not necessarily connected to a cancerous formation but may be a sign of some other disease of the stomach.

Oesophageal Cancer

Oesophageal cancer is a rare type of cancer, where malignant cells form anywhere in the oesophagus. Men after 50 years of age are most commonly affected. Other risk factors include: smoking, alcohol consumption, obesity and gastroesophageal reflux disease (GERD). The most common symptom of oesophageal cancer is difficulty swallowing, which may be accompanied by pain. Other symptoms include food coming back up before reaching the stomach (regurgitation), feeling sick (nausea) or being sick (vomiting), weight loss, hoarseness, cough and enlarged regional lymph nodes. If you experience the aforementioned symptoms that persist, it is important to get it checked by your general practitioner. The symptoms can be caused by other conditions, but it is important for your general practitioner to do further tests and/or refer you to a specialist if necessary.

Colorectal cancer

Colorectal cancer is one of the most common malignancies and the second most deadly cancer globally. It commonly affects people aged over 60. The early stages of the disease are often asymptomatic, that is why early screening tests are recommended. When symptoms appear, they may vary and include the following: a change in the bowel habits, such as diarrhea or constipation, rectal bleeding or blood in your stool, abdominal discomfort (cramps, gas or pain), weakness, fatigue and unexplained weight loss. If persistent symptoms appear, it is important to visit your general practitioner or gastroenterologist.

Pancreatic Cancer

Pancreatic cancer is a rare type of cancer that accounts for 3% of all cancers. Typically, it occurs after the age of 65 and affects both sexes equally. In the beginning stages, the pancreatic cancer tends to be silent and painless as it grows. The symptoms usually appear not specific – severe back pain is among the first. The majority of the symptoms appear when the disease has already progressed. Other symptoms include: loss of appetite and weight loss, jaundice, light stools, dark urine, itchiness and tiredness. Unfortunately, pancreatic cancer usually shows no symptoms in the beginning stages, and it is therefore diagnosed too later when it is difficult to treat. If these symptoms persist, it is important to make an appointment with your general practitioner or an oncologist or a gastroenterologist.

Liver Cancer

Liver cancer is marked by the presence of cancer cells in the liver. The most common type of liver cancer is the hepatocellular carcinoma that commonly begins in the cirrhotic liver. The most common risk factors include viral infections, alcohol consumption, exposure to carcinogens and metabolic disorders (i.e., Wilson's disease, hemochromatosis). Symptoms usually appear as the disease progresses, with nausea, weight loss, early satiety, enlarged liver palpable under the right costal arch, fluid buildup in the abdomen, itching and jaundice being the most common signs of the disease. Should you experience some of these symptoms, ask for a medical help from your general practitioner, who may refer you to a specialist of oncology or gastroenterology.

Biliary Tract Cancer

Biliary tract cancers encompass malignant neoplasms of the gallbladder and the biliary duct. They have a poor prognosis as most of them are found in the advanced stages of the disease. The symptoms are usually unspecific, they include anorexia, nausea, abdominal pain, weight loss, jaundice (yellowing of the eye-whites and the skin) with itching. A fifth of the cases are found accidentally when undergoing a biliary tract surgery. If you experience these symptoms, you might want to see your general practitioner, who will refer you to an oncologist or gastroenterologist. It is important to keep in mind that the aforementioned symptoms are not necessarily connected to a cancerous formation but may be a sign of some other disease of the biliary tract.

Fig. 3. Manual search on the Insieme platform – list of gastrointestinal cancer services.

List of services

- Ageing
- Applications
- Dermatology
- Oncology
 - Rehabilitation
 - General Information
 - Prevention
 - Urogenital Cancers
 - Gastrointestinal Cancers
 - Other Cancers
 - Skin Cancers
 - Gynecologic Cancers
- Hormonal disorders
- Mental illness and cognitive impairment
- Related platforms
- Respiratory diseases
- Digestive tract and metabolism
- Infections, coronavirus

Oesophageal Cancer

Oesophageal cancer is a rare type of cancer, where malignant cells form anywhere in the oesophagus. Men after 50 years of age are most commonly affected. Other risk factors include: smoking, alcohol consumption, obesity and gastroesophageal reflux disease (GERD). The most common symptom of oesophageal cancer is difficulty swallowing, which may be accompanied by pain. Other symptoms include food coming back up before reaching the stomach (regurgitation), feeling sick (nausea) or being sick (vomiting), weight loss, hoarseness, cough and enlarged regional lymph nodes. If you experience the aforementioned symptoms that persist, it is important to get it checked by your general practitioner. The symptoms can be caused by other conditions, but it is important for your general practitioner to do further tests and/or refer you to a specialist if necessary.

Professional help

- Ospedale Classificato Villa Salus, Via Terraglio, 114 - 30174 - Venezia (VE), +39 0412906411
- Ospedale dell'Angelo Di Mestre, Via Piacca, 11 - 30174 - Venezia (VE), 0412607111
- Ospedale di Dolo, Riviera 29 Aprile, 2 - 30031 - Dolo (VE), 0415133111
- Ospedale di Venezia, Calle Castello, 6777 - 30122 - Venezia (VE), 0415294111
- P.O. di San Daniele, Via Trento Trieste, 33 - 33038 - San Daniele del Friuli (UD), 04329491
- P.O. di San Daniele - Tolmezzo, Via Giobatta Morgagni, 20 - 33028 - Tolmezzo (UD), 0433 488213
- P.O. Santa Maria della Misericordia, Piazzale Santa Maria della Misericordia, 15 - 33100 - Udine (UD), 0434 223522
- Centro Riferimento Oncologico, Via Gallini Franco, 2 - 33081 - Aviano (PN), +390434659111
- Policlinico San Giorgio S.p.a., Via Agostino Gemelli, 10 - 33170 - Pordenone (PN), 0434519111
- ULSS 3 Serenissima Sede di Chioggia, Via Madonna Marina, 500 - 30015 - Chioggia (VE), 0415534111
- P.O. Santa Maria degli Angeli di Pordenone, Via Montereale, 24 - 33170 - Pordenone (PN), +390434399111
- AS FO - Azienda Sanitaria Friuli Occidentale, Via Della Vecchia Ceramica, 1 - 33170 - Pordenone (PN), 0434369111
- Ospedale Civile di Palmanova, Via Natisone, 11 - 33057 - PALMANOVA (UD), 0432 921111
- Ospedale di Cattinara, Strada di Fiume, 447 - 34169 - Trieste (TS), 0403991111
- Ospedale di Gorizia, Via Fatebenefratelli, 34 - 34170 - Gorizia (GO), 04815921
- Ospedale di Latisana, Via Sabbionera, 45 - 33053 - Latisana (UD), 0431529111
- Ospedale di Monfalcone, Via Galvani, 1 - 34074 - Monfalcone (GO), 04814971
- Ospedale di San Vito al Tagliamento, Via Savorgnano, 2 - 33078 - San Vito al Tagliamento (PN), 0434841200
- Casa di Cura "Salus, Via Bonaparte Napoleone, 4 - 34123 - Trieste (TS), 040 317111

Articles

- World J Gastrointest Oncol.: K.J Nagler, M Schreier, S Misra. Oesophageal cancer: A review of epidemiology, pathogenesis, staging workup and treatment modalities

Patient's story

- Oesophageal Cancer Patient Stories

Video

- Barret's Esophagus (Esophageal Cancer). Tom's Story
- What is oesophageal cancer?

Additional information

- cancer.gov: Esophageal Cancer
- macmillan.org.uk: Oesophageal Cancer
- ASCOanswers Esophageal Cancer (fact sheet)

Symptoms: enlarged lymph nodes, pain, weight loss, difficulty swallowing, vomiting blood, hoarseness, coughing up blood

Fig. 4. An example of core information in the Insieme platform – the esophageal cancer service.

traditional web search engines. In the latter case, we indicated which additional information should be provided, making these tests more about finding effective approaches rather than formal objective evaluations.

The main functionalities include several services to provide the needed information, e.g., the ability to search using a side menu bar or search function, online human assistance (live chat with a call center or healthcare expert), viewing health-related video content, and using a virtual assistant, which is presented as the central theme in the remainder of the article. All this content is available to users in four languages.

As presented in Fig. 2, on the left side of the main menu is a list of fields and services offered by the Insieme platform. This menu bar enables users to select different branches of medicine. Upon selection, it expands to show diseases and medical conditions associated with the chosen branch, along with information services related to the selected medical specialization. For example, by clicking on one of the medical fields, e.g. oncology, the user is redirected to the corresponding subpage. There or a step further, essential basic information about the course of a specific disease, symptoms, possible prevention, relevant institutions, relevant pictures and videos, relevant Web applications, and further actions are available, as presented in Fig. 4. There are also more links to external websites, enabling the user to acquire appropriate information and knowledge about the chosen disease. While most of the Insieme information can be found on the Web through search engines like Google, it is not easy to find it among hundreds of potential hits, also consisting of medically misleading information in contrast to the information gathered in the platform, carefully evaluated by medical experts in proper contexts.

Among other functions, online human assistance is also available to users. On the entrance page, there are lists of call centers and active doctors that can be contacted via the live Web chat integrated into the platform.

The Insieme platform offers several built-in assistants: queue assistants, IJS assistants, service search, virtual assistants for medicine, and links to other non-integrated assistants. For example, the queue assistant allows entering the name of the procedure or service, specifying the approximate urgency, when the chosen procedure is needed, and the desired region in Slovenia for performing the procedure. The medical assistant answers any user questions in the field of healthcare, providing appropriate guidelines and advice as a response.

5. Insieme-enriched HomeDoctor AI medical assistant

In this section, we present our HomeDoctor medical AI assistant, which integrates the Insieme platform with GPT.

5.1. Background

The preexisting virtual assistants on the Insieme platform were designed to answer health-related questions in the pre-GPT way. With the appearance of GPT-4, the existing assistants were upgraded with ChatGPT-type assistants- or, inversely, enriched GPT-4 with the Insieme platform. From the top GPT view, this assistant possesses general GPT-4 knowledge from the Web, as well as additional local information related to the Insieme project.

The first issue in using the top assistant is the large amount of information to be included, e.g., books and videos. Large language models typically have a limitation on how much input they can accept [22]. Therefore, it is crucial to provide the pre-prepared essential information to the large language model in the appropriate form. Key to this are word embeddings and vector databases.

5.2. Word embeddings and vector databases

Embeddings provide a method to represent words, sentences, or even entire documents. Calculating these embeddings requires appropriate models trained on extensive datasets, capable of identifying relationships between words by analyzing patterns within the data [15]. In this study, we utilized the text-embedding-ada-002 model offered by OpenAI. By generating a vector for each word, one can effectively capture the meaning of the text. These word embeddings can be represented in multidimensional spaces, where semantically similar words or sentences are positioned closely together. This allows the calculation of distances between vectors to identify semantically related words.

Vector databases store this information in the form of vectors, often referred to as word (vector) embeddings. This allows indexing and searching through a huge amount of unstructured data, such as images, raw text, or sensor data. Vector databases organize data using high-dimensional vectors, each dimension describing a specific characteristic of the data object it represents. Vector databases differ from traditional databases that store data in tabular form in that they return results based on similarity (traditional databases return exactly matching objects) [34]. Various measures, such as cosine similarity, are used to measure the similarity between vectors in vector space. These measures enable comparisons of vectors stored in the vector database and find those most similar to the user's input vector. They thus enable dealing with complex data and fast searching, which would otherwise cause difficulties for traditional databases.

For example, suppose there is a document to index. The model should enable the creation of word embeddings [23] stored in the selected vector database, and a reference to the document from which the embedding was created. Whenever a user sends a query, the same model is used to create embeddings — to find the most similar word embeddings in the vector database, which are linked to the original document where they were created due to the mentioned reference. The obtained documents can then be provided to the LLM, which will be used as context for generating a response.

5.3. Design and implementation based on LangChain

The HomeDoctor system is based on the LangChain library, which facilitates working with large language models. LLMs can efficiently perform a large number of different tasks, but it is to be expected that they will not be able to correctly answer questions from specialized fields, such as medicine, without specialized knowledge. LangChain enables the integration of models with knowledge of specific fields and awareness of data and conversation context. LangChain is a powerful tool that fills the gap between language models and domain knowledge, which is also why LangChain is increasingly used in applications that perform tasks related to natural language processing. LangChain includes numerous modules that help in development [1]:

- LLM: enables the use of the capabilities of a specific large language model.
- Chains: the main unit, as evident from the name LangChain, which combines multiple LLM calls. An example of this would be to first read the user's input, then use this input to compose a new input (prompt), which is given to the large language model, which then generates a response.
- Inputs, prompts: LangChain offers many different ways to change the input given to the language model. We can use prompt templates, where we precisely define the form of the input.
- Document loading modules: allow conversion of various types of data (PDF documents, HTML Web pages, image material) into text that can be processed.
- Agents: for applications where the sequence of calls is not predetermined, LangChain provides agents that can act based on inputs, instead of having a pre-determined sequence.

In general, LangChain's ability to integrate various data sources and formats enhances its utility across different domains. In educational settings, LangChain can facilitate the creation of interactive learning tools that adapt to students' needs by understanding the context of their questions and providing tailored responses. This personalized approach to education can significantly improve learning outcomes by addressing individual knowledge gaps and promoting active engagement [30].

In business environments, LangChain can streamline customer service operations by automating responses to frequently asked questions, processing complex queries, and escalating issues to human agents when necessary. By reducing the workload on customer service representatives and ensuring that customers receive timely and accurate information, businesses can improve customer satisfaction and operational efficiency [14].

LangChain also supports document summarization and data extraction, which are invaluable in legal and financial services where large volumes of text need to be analyzed and interpreted. By utilizing LangChain, these industries can enhance their data processing capabilities, ensuring that critical information is accurately captured and presented [10].

Overall, LangChain's comprehensive suite of tools and its ability to bridge the gap between general-purpose language models and domain-specific knowledge make it an indispensable asset for developing sophisticated NLP

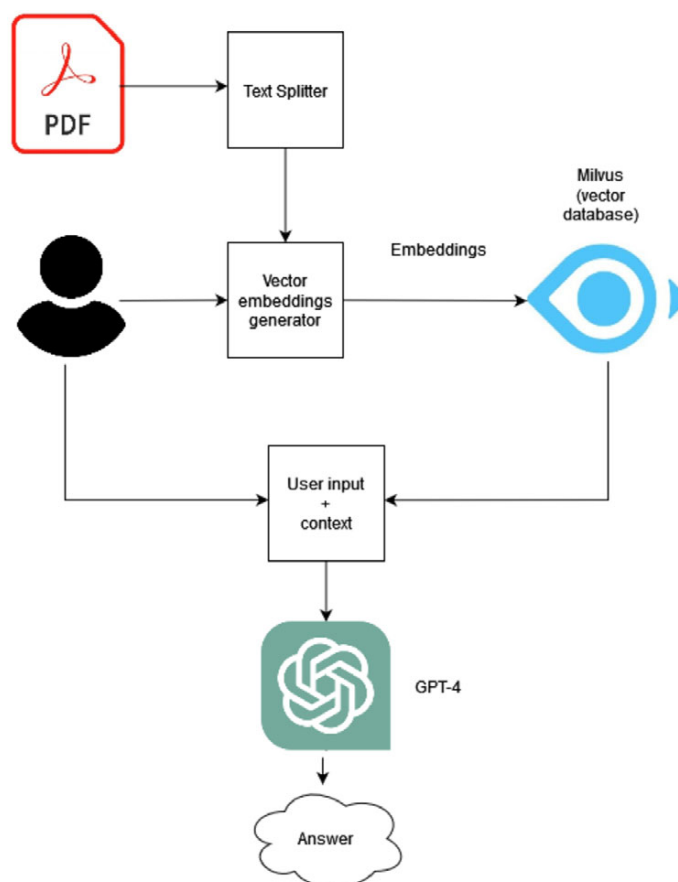


Fig. 5. Essential components of the HomeDoctor system from the LangChain perspective.

applications across various industries. As the demand for intelligent and context-aware conversational agents continues to grow, LangChain's role in advancing the capabilities of language models will become increasingly prominent [38].

By leveraging the power of LangChain, developers can create robust and versatile applications that harness the full potential of large language models while ensuring that these models can effectively navigate and respond to the intricacies of specialized knowledge domains [21]. This not only enhances the performance and reliability of NLP applications but also expands the horizons of what is possible with AI-driven language processing technologies [4].

In healthcare, LangChain can be used to create conversational agents that provide reliable information by accessing specialized medical databases and knowledge sources. This ensures that the answers provided by the model are accurate and up-to-date, mitigating the risk of misinformation [32]. The actual schema of the HomeDoctor system using LangChain is presented in Fig. 5:

- **Document Loader:** (left up in Fig. 5, loads a PDF file) This module allows for the easy uploading and pre-processing of various data types (e.g., PDF documents, HTML Web pages, images). The DirectoryLoader, a component of this module, enables the storage of documents used in a common directory.
- **Text Splitter:** This tool divides long text parts into smaller, semantically meaningful chunks. This step is crucial for maintaining the semantic integrity of the text while processing it.
- **Embeddings:** The Embedding class provides a standardized interface for generating word embeddings using models like OpenAI's text-embedding-ada-002. These embeddings convert text into vector representations, enabling semantic analysis and tasks such as semantic searching.

- **Vector Database:** Once the embeddings are generated, they are stored in a vector database. This database allows for efficient indexing and searching of large amounts of unstructured data based on similarity rather than exact matches [20].
- **Chains:** The core unit in LangChain, which combines multiple LLM calls. For instance, it can read the user's input, use it to generate a new input (prompt), and then generate a response.
- **Agents:** (left middle) These are used for applications where the sequence of calls is not predetermined. Agents can act based on inputs rather than following a predetermined sequence.
- **Prompts:** (User input + context in Fig. 5) LangChain offers various ways to modify the input given to the language model, such as prompt templates that precisely define the form of the input.

The first step in this development is uploading data into 'Documents', which are pieces of text. The Document Loader module in the LangChain tool simplifies this task and allows for easy uploading and preprocessing of data – DirectoryLoader enables storing all used documents in a common directory. This is followed by dividing the documents into smaller pieces – the text splitter allows for breaking long text parts into smaller, semantically meaningful chunks [31]. This task may seem simple, but it involves some complexity. The goal is to divide the text in a way that keeps semantically connected parts together, where 'semantic connectivity' depends on the type of text being processed. Text splitters divide the text into small pieces, often based on sentence boundaries. These small pieces are combined into larger pieces until they reach a certain size determined by a pre-defined function for measuring the size of the piece – when a piece reaches the desired size, it becomes an independent piece of text. Then a new piece is created with some overlap (chunk overlap) to maintain context between individual pieces.

This is followed by the generation of word embeddings, which play a key role in representing textual information. The Embedding class in the LangChain tool serves as a standardized interface for various embedding providers, including OpenAI. Through the generation of word embeddings, the text is converted into a vector representation, enabling semantic analysis and tasks such as semantic searching. All this is stored in the vector database as a new index using built-in methods enabling semantic searches over this object and retrieving documents relevant to the user's input. The obtained documents are then forwarded to the language model, which treats the documents as context for generating a response.

In Fig. 6, one can see the process of responding to user questions. ChatGPT fluently answers health questions by considering general knowledge, medical knowledge from the entire Web, and specific knowledge from the Insieme platform. The user can fluently change language from Slovenian to English and Italian, actually generating two versions of the English communication. All users are anonymous to prevent identification. The communication is through the platform calling for GPT-4 enriched by the Insieme platform information and knowledge.

The replies generated by the integrated system may initially appear similar to the standard GPT-4 responses. However, they are specifically tailored for the average local user, incorporating local information such as nearby institutions, and ensuring greater understandability. These responses are also verified by national medical experts, minimizing the likelihood of inaccuracies – particularly when the information is sourced directly from the Insieme platform. Nonetheless, hallucinations can still occur in responses where GPT relies on general knowledge instead of specific embedded knowledge sources. It is important to note that the specific knowledge sources contain only the most commonly used information, necessitating the use of GPT for about half of the responses. In our experiments, users often asked unusual questions.

The next phase is typically a discussion about which measures to take for prevention and cure, and where professional help can be found.

The verification and assurance of informative and appropriate responses were conducted by the authors of this paper, along with other department members, including AI researchers and medical experts, primarily in the Slovenian language. The integrated ChatGPT was tested with questions covering the knowledge contained in the Insieme platform to improve the system. Generated responses were critically evaluated, and corrections were made as necessary. Ensuring the accuracy of these answers is crucial, as misleading or inaccurate information could potentially worsen a user's health condition. It should be noted that formal extensive evaluations have yet to be performed.

The following two question-answer examples first demonstrate a typical GPT-style reply, followed by a response enriched with additional local information. This includes guidance on where a user with intimate health concerns can seek further testing and which web applications to utilize for elaborating on the issue.

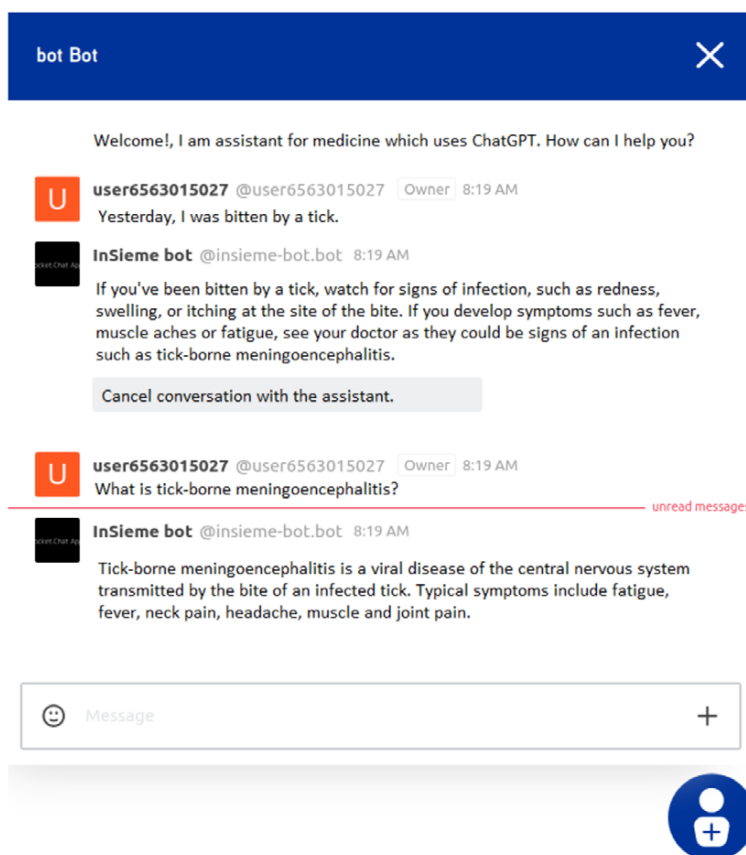


Fig. 6. An example of questions and answers from the integrated ChatGPT in the multilingual platform Insieme.

The conversation (translated from Slovenian language) from Fig. 7:

User: *Yesterday I fell at soccer, and now my leg hurts.*

HomeDoctor: *I'm sorry to hear that you fell at soccer and that your leg hurts.*

To help you better, I need some additional information:

- *Where exactly does your leg hurt?*
- *Was there swelling or bruising at the site of the fall?*
- *Do you have trouble moving your leg?*
- *Have you noticed any other symptoms such as numbness or tingling in your leg?*

Please share more details about your symptoms and how bad the pain is. This will help me assess your condition.

Another example of communication is presented here:

User: *It hurts when I urinate. How can I find out if I have a sexual disease?*

HomeDoctor: If you have pain when urinating, it is possible that you have a sexually transmitted disease. Here are some steps to check this:

1. ****Symptoms**:** Pay attention to the following symptoms that may indicate a sexually transmitted disease:
 - Itching, - Discharge, - Rash, blisters or unusual bumps, - Blood in the urine – Pain when urinating (condensed actual output)

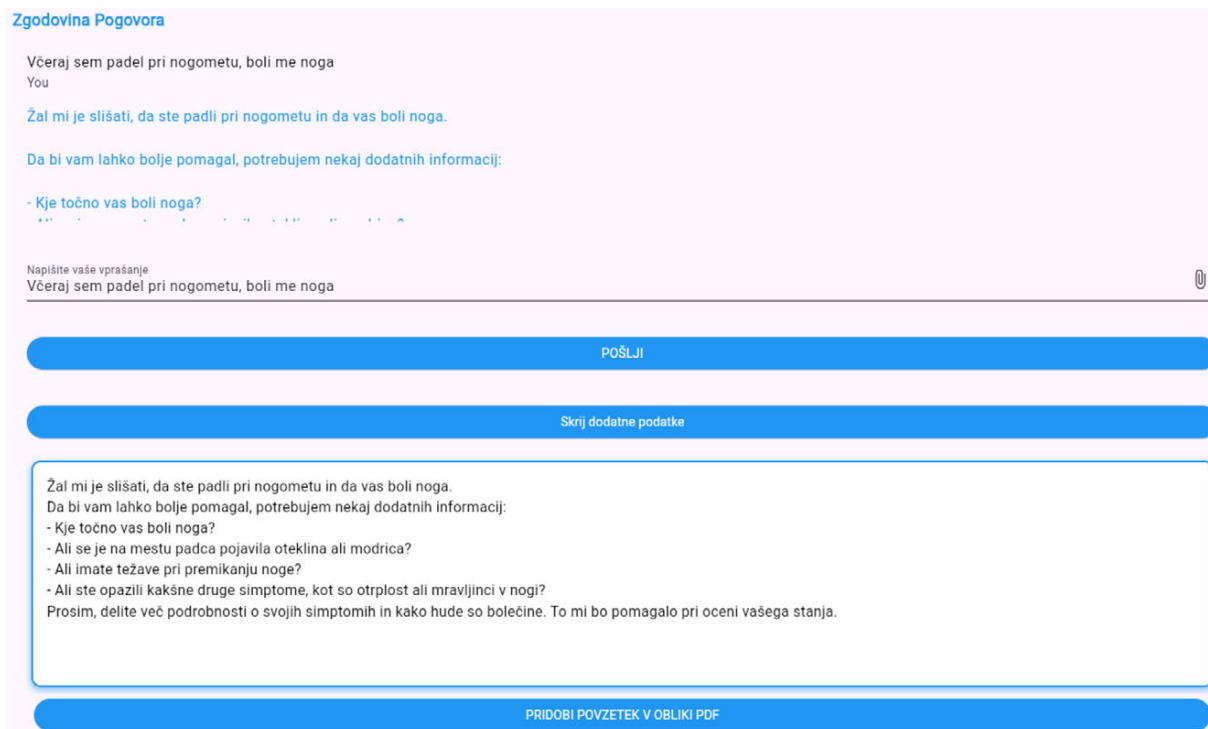


Fig. 7. An example of questions and answers from the stand-alone HomeDoctor application, an enhanced version of the Insieme chatbot (conversation in Slovenian language).

- ... ****Testing****: Sign up for STD testing. Here are useful links to login and test sites:
- [Application for testing] (<https://kajisces.si/testiranje/>)
 - [Test sites] (<https://kajisces.si/mesta/>)
- ... You can also use the ASPO app, which has pictures of symptoms, for more information and help: ASPO application (<https://aspo.mf.uni-lj.si/#/>)

The communication resembles a typical GPT-style response, but the second reply is enriched with links to local testing facilities and relevant web applications, allowing the user to further explore and address the issue without having to explain it to a human.

This communication is consistent across PCs, tablets, and mobile phones on major global platforms as HomeDoctor utilizes Flutter, an open-source UI software development toolkit created by Google. Flutter is used for building natively compiled applications for mobile (iOS, Android), Web, and desktop from a single codebase. It employs the Dart programming language and offers a rich set of pre-designed widgets and tools to create visually appealing and high-performance applications.

Concurrently, the design and implementation of the HomeDoctor represents an innovative approach to utilizing GPT as a software package for integration, which consequently introduces novel challenges concerning the integration and modification of knowledge. At the same time, it is fundamentally a software application dealing with data, adhering to the typical design-test-improve cycles.

The Insieme and HomeDoctor systems are still under development, and our evaluations thus far have been limited to informal tests, ranging from dozens to hundreds per person. The overall feedback suggests that users perceive the system as somewhat akin to GPT. However, when the differences are explicitly observed or demonstrated, users readily acknowledge and appreciate the distinctions.

In terms of performance, HomeDoctor responds within a matter of seconds, akin to GPT-4, with no significant differences in response time or performance observed. However, it is noteworthy that in recent months, there were

intermittent periods during which GPT-4 was non-operational, resulting in the HomeDoctor system also being rendered inactive.

HomeDoctor is designed to operate on any common computer, tablet, or mobile phone. While a small percentage of the population, particularly the elderly and those in rural areas, might not have access to such devices, it is estimated that over 95 percent of the Slovenian population, as well as the broader EU population, should be able to use HomeDoctor.

During the implementation, several challenges were encountered. One significant issue was managing the volume of information and ensuring semantic connectivity while splitting text. We addressed this by implementing chunk overlap techniques to maintain context. Another challenge was optimizing the search efficiency in the Vector Database, which we overcame by fine-tuning cosine similarity measures for better accuracy. Additionally, integrating the system with the Insieme platform required careful handling of medical data privacy, which was mitigated through strict adherence to GDPR regulations and local data storage solutions, giving the users responsibility and rendering the system not responsible. These troubleshooting steps were crucial in refining the system's performance and reliability.

5.4. Novelty and innovation in the Insieme platform and HomeDoctor application

The Insieme platform, along with its upgraded chatbot and the HomeDoctor application, represents a significant advancement in digital healthcare, to scale up to a national level to serve every citizen with access to mobile devices or computers. This initiative addresses a critical issue in Slovenia, where approximately 140,000 patients currently lack access to a personal doctor and all citizens face long waiting times for medical services. This situation places immense strain on an already overstretched medical workforce. By integrating generative AI, this project seeks to alleviate the burden on healthcare professionals while providing citizens with continuous access to high-quality medical advice. While the task might not seem scientific at first glance, its complexity lies in addressing nuanced issues such as ensuring accurate and contextually appropriate information, managing potential AI hallucinations, and balancing generality with local specialization. In addition, any mistake at this level is likely to be publicized and criticized in the media. These challenges present significant obstacles that must be resolved in this and future applications.

Key Innovations and Contributions National-level implementation: The Insieme platform's ambition to operate at a national level is a groundbreaking step in ensuring comprehensive healthcare access. By leveraging mobile devices and computers, the system aims to provide equitable healthcare services across Slovenia, particularly benefiting underserved populations. The recent AI progress makes this task feasible, yet far from straightforward, as demonstrated by the absence of such applications worldwide.

Alleviating healthcare workforce strain: The shortage of medical personnel in Slovenia, compounded by the lack of access to personal doctors for a significant portion of the population, has led to overworked and exhausted healthcare providers. The HomeDoctor application, powered by generative AI, aims to mitigate this issue by offering virtual consultations, automating administrative tasks, and providing decision support for healthcare professionals. This approach helps streamline operations and reduces the workload on medical staff, ultimately enhancing patient care and healthcare efficiency.

24/7 quality medical advice: One of the standout features of the HomeDoctor application is its ability to provide reliable medical advice around the clock. This continuous access to medical support ensures that citizens can receive timely assistance, manage their health conditions effectively, and prevent minor issues from escalating into serious health problems. The platform's capability to offer geographically and medically verified information further enhances the reliability and trustworthiness of the advice provided.

Enhanced accessibility: The application is designed to be user-friendly across various devices, including PCs, tablets, and mobile phones, on major worldwide platforms. This accessibility is particularly beneficial for individuals in remote areas or those with mobility challenges, reducing the need for in-person visits and making healthcare more inclusive.

Integration of verified knowledge sources: Unlike standard GPT implementations or Google, the HomeDoctor system integrates geographically and medically verified knowledge sources, ensuring that the information provided

is accurate and contextually relevant. This integration significantly reduces the risk of medical misinformation and enhances the quality of responses. In our experience, users ask all kinds of questions, even in medical sessions, where preventing hallucinations is beyond the scope of this study.

Support for Healthcare Professionals: The system is not only a tool for patients but also a valuable resource for healthcare professionals. By providing quick access to medical literature, drug information, and case studies, the AI supports doctors and nurses in making informed decisions, thus improving the quality of care.

Impact on Healthcare Outcomes: Preliminary studies have indicated high user satisfaction and positive impacts on healthcare outcomes. By facilitating virtual consultations and automating routine tasks, the HomeDoctor application helps improve healthcare delivery efficiency and patient outcomes.

Future Expansion: The vision for the Insieme platform extends beyond Slovenia, with plans to expand to other European Union (EU) countries. This expansion aims to deliver innovative healthcare solutions to a broader audience, thereby enhancing access to medical services and fostering community engagement across the region. Moreover, the primary goal of this study is to share experiences and insights in designing a national 24/7 AI home doctor system.

Multilingual Support: The system supports multiple languages, making it accessible to non-native speakers and ensuring that language barriers do not impede access to local healthcare information and services. This function is relevant, for example, for tourists.

6. Discussion and conclusion

This paper presents HomeDoctor, an innovative electronic and mobile health platform that integrates the conversational virtual assistant ChatGPT with added medical knowledge in the national home environment. The core innovation lies in the delicate and premeditated combination of GPT-4o with the Insieme platform and additional verified medical knowledge sources to provide comprehensive medical information most relevant for national use. This integration merges global and local knowledge, including local and global e-health mobile services. Rare medical cases or unrelated questions are managed by the core GPT-4o.

The prototype system demonstrates the ability to provide 24/7 online medical assistance, offering both standard ChatGPT responses and information from a vector database accessible through the Insieme platform. The implementation was evaluated locally, proving the system's effectiveness on common computer and mobile platforms via Flutter. This platform supports continuous, high-quality medical advice, combining generative AI with specialized medical knowledge.

The platform shows significant potential for further enhancements, particularly through integration with other health information systems and sources within Slovenia. This would enable the Insieme platform to become the most comprehensive national information source for users, providing 24/7 online quality medical information as a home doctor, and easing the burden on overburdened medical staff. Ongoing discussions with health institutions throughout Slovenia aim to facilitate this integration, broadening the platform's utility and reach. The timeline consists of the first publicly presented prototype in a couple of months, modifications according to the feedbacks, and then adding additional functionality.

Despite its promise, the system faces several challenges. Data privacy is a critical concern, particularly regarding sensitive personal medical data that should remain within the institution or at least within the country. One solution is to use local language models running on home computers or at national institutions. Local models, such as Llama 2 and Llama 3, were tested but proved less effective first in the English language and additionally highlighting the need for models trained in Slovenian. Nonetheless, the possibility of testing more advanced open-source models remains open, especially given the advantages of local installation. This approach ensures all data remains locally accessible, complying with legal requirements that prohibit sending formal medical data out of Slovenia without user consent. While the system forges forgetting after the end of the session, the data are sent to global servers. Sending data anonymously is one option. However, data privacy issues rely on users providing information at their own will and responsibility.

The roadmap for future work first consists of enhancing the natural language processing capabilities to better understand and respond to user queries. Next, we aim to expand the system's database to include a broader range of medical knowledge, which will involve continuous collaboration with healthcare professionals. Another critical

goal is to conduct extensive user studies to gather feedback and identify areas for improvement. Additionally, we plan to integrate advanced features such as predictive analytics for early disease detection and personalized health recommendations. These updates will be complemented by rigorous testing to validate the system's efficacy and reliability.

A critical challenge with GPT-4 communication is the dependence on input quality. Incorrect user input leads to unsatisfactory replies, a limitation also applicable to the training data. Ensuring high-quality input and contextual understanding is essential for the system's reliability. This issue is challenging to resolve, as most users are medically uneducated and may provide inaccurate medical data. Providing data from local sensors from local environments can eliminate a great part of this issue. One should note that modern mobile phones contain up to 20 sensors and there are hundreds of applications based on them, related to HomeDoctor. These services will be added to the HomeDoctor as another step in progress.

The primary aim of the platform is to introduce foundational ideas for modernizing the healthcare system using mobile phones and connected sensors to create an intelligent medical environment. This approach seeks to alleviate the burden on healthcare professionals while providing users with an effective, constantly available source of information based on the latest research findings. The system can serve as both a second opinion and an initial opinion for simple health issues, offering immediate, reliable support. Designing a national system involves addressing numerous orthogonal issues, ranging from computer-related to medical, media, social, ethical, and legal considerations.

The experiments conducted in this study confirm that generative artificial intelligence holds significant promise for improving national healthcare systems. By enriching the system with both general and local information, the platform can offer radical improvements in accessibility and quality of care. The insights and experiences gained from this project will serve as a valuable foundation for future developments and implementations in digital healthcare.

Acknowledgement

The Insieme platform was developed as part of the cross-border Interreg ISE-EMH project, which is funded by the Italy-Slovenia Cooperation Program from the European Regional Development Fund. Support was also provided by ARIS – Slovenian Research and Innovation Agency. We also thank members of the Department of Intelligent Systems and medical experts for providing info and testing the system.

Conflict of interest

The authors have no conflict of interest to report.

References

- [1] S. Agastya, LangChain: A comprehensive guide, 2023. Retrieved from https://www.langchain.com/docs/guides/langchain_comprehensive_guide.
- [2] J. Ayers, A. Poliak, M. Dredze, E. Leas, Z. Zhu, J. Kelley, D. Faix, A. Goodman, C. Longhurst, M. Hogarth and D. Smith, Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum, *JAMA Internal Medicine* **183** (2023). doi:10.1001/jamainternmed.2023.1838.
- [3] D. Brin, V. Sorin, A. Vaid et al., Comparing ChatGPT and GPT-4 performance in USMLE soft skill assessments, *Scientific Reports*. (2023). doi:10.1038/s41598-023-43436-9.
- [4] T.B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal and D. Amodei, Language models are few-shot learners, 2020. doi:10.48550/arXiv.2005.14165.
- [5] H. Castro, ChatGPT and healthcare. Independently published, 2023.
- [6] J. Dahmen, M.E. Kayaalp, M. Ollivier, A. Pareek, M.T. Hirschmann, J. Karlsson and P.W. Winkler, Artificial intelligence bot ChatGPT in medical research: The potential game changer as a double-edged sword, *Knee Surgery, Sports Traumatology, Arthroscopy* **31**(4) (2023), 1187–1189. doi:10.1007/s00167-023-07355-6.

- [7] T. Dave, S.A. Athaluri and S. Singh, ChatGPT in medicine: An overview of its applications, advantages, limitations, future prospects, and ethical considerations, *Frontiers in Artificial Intelligence* **22**(37) (2022). doi:10.3389/fraci.2023.1169595.
- [8] T. Davenport and R. Kalakota, The potential for artificial intelligence in healthcare, *Future Healthcare Journal* **6**(2) (2019), 94–98. doi:10.7861/futurehosp.6-2-94.
- [9] C. De Terwangne, Council of Europe convention 108+: A modernised international treaty for the protection of personal data, *Computer Law & Security Review* **40** (2021), 105497. doi:10.1016/j.clsr.2020.105497.
- [10] J. Devlin, M.W. Chang, K. Lee and K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: *Proceedings of NAACL-HLT*, 2019, pp. 4171–4186.
- [11] Y. Fu, D. Schwebel and G. Hu, Physicians' workloads in China: 1998–2016, *International Journal of Environmental Research and Public Health* (2018). doi:10.1038/s41598-023-43436-9.
- [12] N. Ghadiri, Comparison of GPT-3.5, GPT-4, and human user performance on a practice ophthalmology written examination and ChatGPT in ophthalmology: The dawn of a new era? *Eye* **37** (2023). doi:10.1038/s41433-023-02773-9.
- [13] T. Hirotsawa, Y. Harada, M. Yokose, T. Sakamoto, R. Kawamura and T. Shimizu, Diagnostic accuracy of differential-diagnosis lists generated by generative pretrained transformer 3 chatbot for clinical vignettes with common chief complaints: A pilot study, *International Journal of Environmental Research and Public Health* **20**(4) (2023), 3378. doi:10.3390/ijerph20043378.
- [14] M.H. Huang and R.T. Rust, Engaged to a robot? The role of AI in service, *Journal of Service Research* **23**(2) (2020), 155–172. doi:10.1177/1094670517752459.
- [15] Q. Jiao and S. Zhang, A brief survey of word embedding and its recent development, in: *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 2021, pp. 1697–1701. doi:10.1109/IAEAC50856.2021.9390956.
- [16] B.A. Juliussen, E. Kozyri, D. Johansen and J.P. Rui, The third country problem under the GDPR: Enhancing protection of data transfers with technology, *International Data Privacy Law* **13**(3) (2023), 225–243. doi:10.1093/idpl/ipad013.
- [17] J. Kolar and M. Gams, Integration of national healthcare platforms within the European Union: Challenges and opportunities, *Journal of Health Informatics Research* **7**(2) (2023), 123–145. doi:10.1007/s41666-023-00123-4.
- [18] P. Lee, C. Goldberg and I. Kohane, *The AI Revolution in Medicine: GPT-4 and Beyond*. The Medical Futurist, 2023.
- [19] B. Mesko, *Generative AI in Healthcare*. Pearson, 1st edn, 2023.
- [20] T. Mikolov, K. Chen, G. Corrado and J. Dean, Efficient estimation of word representations in vector space, 2013. doi:10.48550/arXiv.1301.3781.
- [21] O. Mishra, Using langchain for question answering on own data, 2023. Retrieved September 3, 2023, from <https://medium.com/@onkarmishra/using-langchain-for-question-answering-on-own-data-3af0a82789ed>.
- [22] H. Naveed, A.U. Khan, S. Qiu, M. Saqib, S. Anwar, M. Usman, N. Akhtar, N. Barnes and A. Mian, A comprehensive overview of large language models, *ACM Computing Surveys* **55**(3) (2023), 1–30. doi:10.1145/3448016.3457550.
- [23] A. Neelakantan, T. Xu, R. Puri, A. Radford, J.M. Han, J. Tworek, Q. Yuan, N. Tezak, J.W. Kim, C. Hallacy, J. Heidecke, P. Shyam, B. Power, T. Eloundou Nekoul, G. Sastry, G. Krueger, D. Schnurr, F. Petroski Such, K. Hsu, M. Thompson, T. Khan, T. Sherbakov, J. Jang, P. Welinder and L. Weng, Text and code embeddings by contrastive pre-training, 2022. doi:10.48550/arXiv.2201.10005.
- [24] OpenAI, GPT-4 Technical Report, 2023. doi:10.48550/arXiv.2303.08774.
- [25] I. Platform, 2023. Retrieved from <https://ise-emh.eu>.
- [26] R. Rajkomar et al., Scalable and accurate deep learning with electronic health records, *npj Digital Medicine* **1**(1) (2018), 1–10. doi:10.1038/s41746-018-0029-1.
- [27] P.P. Ray and P. Majumder, The potential of ChatGPT to transform healthcare and address ethical challenges in artificial intelligence-driven medicine, *Journal of Clinical Neurology* **19**(9) (2023), 509–511. doi:10.3988/jcn.2023.0158.
- [28] T. Reed, Can ChatGPT be used in healthcare. Axios, 2023. Retrieved January 10, 2024, from <https://www.axios.com/2023/11/29/chat-gpt-health-care-medicine-clinical-diagnosis>.
- [29] L. Rosencrance, 2024, 9 Uses of Generative AI in Healthcare. Techopedia. Retrieved from <https://www.techopedia.com/9-uses-of-generative-ai-in-healthcare>.
- [30] J. Seabrook, The algorithm will see you now: How AI is reinventing healthcare. The New Yorker, 2020. Retrieved from <https://www.newyorker.com/magazine/2020/04/06/the-algorithm-will-see-you-now>.
- [31] A.S.A. Sreeram and P. Jithendra, An effective query system using LLMs and LangChain, *International Journal of Engineering Research & Technology (IJERT)* **12**(06) (2023).
- [32] D.J. Topol, High-performance medicine: The convergence of human and artificial intelligence, *Nature Medicine* **25**(1) (2019), 44–56. doi:10.1038/s41591-018-0300-7.
- [33] VOS: Mental health AI therapy, 2024. Retrieved from <https://play.google.com/store/apps/details?id=com.vos.app>.
- [34] J. Wang, X. Yi, R. Guo, H. Jin, P. Xu, S. Li, X. Wang, X. Guo, C. Li, X. Xu, K. Yu, Y. Yuan, Y. Zou, J. Long, Y. Cai, Z. Li, Z. Zhang, Y. Mo, J. Gu, R. Jiang, Y. Wei and C. Xie, Milvus: A purpose-built vector data management system, in: *Proceedings of the 2021 International Conference on Management of Data*, Association for Computing Machinery, 2021, pp. 2614–2627. doi:10.1145/3448016.3457550.
- [35] A. Welivita, A survey of consumer health question answering systems, *AI Magazine* (2023). doi:10.1002/aaai.12140.
- [36] World Health Organization Health and care workforce in Europe: Time to act. World Health Organization, 2022.
- [37] Wysa: Anxiety, therapy chatbot, 2024. Retrieved from <https://play.google.com/store/apps/details?id=bot.touchkin>.
- [38] T. Zhang and Q. Yang, A survey on multi-task learning, *IEEE Transactions on Knowledge and Data Engineering* **34**(1) (2021), 1–16.