DOI: 10.15514/ISPRAS-2025-37(6)-10

Enhancing E-Government Services through Chatbot Development Using Azure OpenAI

L. Radosavljević, ORCID: 0009-0005-6666-8913 <lr20180196@student.fon.bg.ac.rs>
M. Simić, ORCID: 0000-0002-6870-2303 <milica.simic@fon.bg.ac.rs>
A. Joksimović, ORCID: 0009-0008-5711-7636 <aleksandar.joksimovic@elab.rs>
T. Naumović, ORCID: 0000-0001-9849-7665 <tamara.naumovic@fon.bg.ac.rs>
M. Despotović-Zrakić, ORCID: 0000-0002-6458-1575 <marijana.despotovic-zrakic@fon.bg.ac.rs>

University of Belgrade – Faculty of Organizational Sciences,
Jove Ilića 154, Belgrade, 11000, Serbia.

Abstract. Implementing and developing chatbots as e-government services contributes significantly to the modernization and increased efficiency of public services. By leveraging Microsoft technologies - specifically the Azure OpenAI service - it is possible to rapidly and effectively develop intelligent chatbots. When integrated with the e-government portal, such chatbots offer users improved access to information and enable personalized communication between citizens and government institutions. A key issue currently lies in the lack of effective communication channels, which results in longer response times and reduced user satisfaction. The objective of this paper is to develop a chatbot that enhances service quality and brings public services closer to citizens via the e-government portal. The paper analyzes chatbot functionalities such as response generation, relevance checking of user queries, and information filtering. Furthermore, attention is given to legal and ethical considerations, data protection, and continuous model training to maintain data accuracy. This paper also explores and demonstrates how modern artificial intelligence technologies contribute to making public services more accessible to users. The proposed solution to the identified challenges involves the implementation of a chatbot model integrated with the e-government portal to improve communication. Ultimately, the focus is placed on the digitalization and modernization of public sector services to deliver benefits for society as a whole.

Keywords: chatbot; Azure OpenAI; e-government; artificial intelligence.

For citation: Radosavljević L., Simić M., Joksimović A., Naumović T., Despotović-Zrakić M. Enhancing e-government services through chatbot development using Azure OpenAI. Trudy ISP RAN/Proc. ISP RAS, vol. 37, issue 6, part 1, 2025, pp. 167-180. DOI: 10.15514/ISPRAS-2025-37(6)-10.

Acknowledgements. The authors are thankful to the Ministry of Science, Technological Development, and Innovation of the Republic of Serbia for financial support through institutional funding of FON.

Улучшение электронных государственных услуг посредством разработки чат-ботов с использованием Azure OpenAI

Л. Радосавльевич, ORCID: 0009-0005-6666-8913 <lr20180196@student.fon.bg.ac.rs>
М. Симич, ORCID: 0000-0002-6870-2303 <milica.simic@fon.bg.ac.rs>
А. Йоксимович, ORCID: 0009-0008-5711-7636 <aleksandar.joksimovic@elab.rs>
Т. Наумович, ORCID: 0000-0001-9849-7665 <tamara.naumovic@fon.bg.ac.rs>
М. Деспотович-Зракич, ORCID: 0000-0002-6458-1575
<marijana.despotovic-zrakic@fon.bg.ac.rs>

Университет Белграда – Факультет Организационных наук
ул. Йове Илича, 154, Белград, 11000, Сербия.

Аннотация. Реализация и развитие чат-ботов как сервиса электронного правительства в значительной степени способствуют модернизации и повышению эффективности государственных услуг. Используя технологии Microsoft - в частности, сервис Azure OpenAI - возможно быстро и эффективно разработать интеллектуальных чат-ботов. При интеграции с порталом электронного правительства такие чат-боты обеспечивают пользователям улучшенный доступ к информации и персонализированную коммуникацию между гражданами и государственными учреждениями. Одной из ключевых проблем на сегодняшний день является отсутствие эффективных каналов связи, что приводит к увеличению времени отклика и снижению удовлетворенности пользователей. Цель данной работы – разработка чат-бота, который повысит качество предоставляемых услуг и сделает государственные сервисы более доступными гражданам через портал электронного правительства. В работе анализируются функциональные возможности чат-бота, такие как генерация ответов, проверка релевантности заданных вопросов и фильтрация информации. Кроме того, внимание уделяется правовым и этическим аспектам, защите данных, а также постоянному обучению модели для поддержания точности данных. В исследовании также рассматривается и демонстрируется, каким образом современные технологии, основанные на искусственном интеллекте, способствуют приближению государственных услуг к пользователям. Предлагаемое решение выявленных проблем заключается в внедрении модели чат-бота с интеграцией в портал электронного правительства для улучшения коммуникации. В конечном итоге основное внимание уделяется цифровизации и модернизации услуг в государственном секторе для получения выгод для общества в целом.

Ключевые слова: чат-бот; сервис Azure OpenAI; электронное правительство; искусственный интеллект.

Для цитирования: Радосавльевич Л., Симич М., Йоксимович А., Наумович Т., Деспотович-Зракич М. Улучшение электронных государственных услуг посредством разработки чат-ботов с использованием Azure OpenAI. Труды ИСП РАН, том 37, вып. 6, часть 1, 2025 г., стр. 167–180 (на английском языке). DOI: 10.15514/ISPRAS–2025–37(6)–10.

Благодарности. Авторы выражают благодарность Министерству науки, технологического развития и инноваций Республики Сербия за финансовую поддержку через институциональное финансирование ФОН.

1. Introduction

The modernization of technology and the process of digitalization have significantly transformed the way communication occurs between citizens and government institutions. In the past, users had to be physically present and wait in long queues to submit various documents. Today, each of these processes has a digital counterpart. Signing documents, obtaining certificates and official records, and paying fees can now be accomplished with significantly less time and effort. This transformation not only reduces administrative costs but also provides citizens, i.e., users, with easier access to public services.

Despite the digitalization of public services, users often encounter obstacles when attempting to utilize them. The wide range of information available on government portals, along with guidance that is sometimes not particularly intuitive, can lead to confusion and cause citizens to abandon the digital service altogether. To overcome these challenges and facilitate communication, there is a growing need for new, modern technologies, such as chatbots.

The aim of this paper is to explore the application of chatbots in e-government in order to identify their advantages, limitations, and challenges in improving communication between governmental institutions and citizens. The main issue lies in the inefficiency of traditional methods of delivering public services, which are often slow, overly bureaucratic, and limited in accessibility. As a proposed solution, this paper analyzes how artificial intelligence-based chatbot models can automate processes, enhance communication, and simplify access to information and services.

The paper is organized into several interrelated chapters. The second chapter provides definitions of key concepts related to chatbots, an overview of their application in e-government, and a review of relevant Azure technologies used for AI system development, including Azure OpenAI, Azure App Service, Azure AI Search, and Azure Cosmos DB. The third chapter focuses on the development of a chatbot as an e-government service using the Azure OpenAI service. The fourth chapter presents the implemented solution. Finally, the conclusion offers directions for further development and potential improvements.

2. Literature Review

2.1 Chatbots

A chatbot is a computer program designed to simulate human conversation with a user. In order to understand user queries and provide appropriate feedback, it relies on artificial intelligence (AI) and natural language processing (NLP) techniques. Chatbots are widely used across various domains such as sales, education, customer support, and many others. They are commonly integrated into web portals where they serve as digital assistants, and are also frequently implemented on social media platforms and mobile applications [1].

There are several classifications of chatbots that differ in complexity, sophistication, and other criteria. However, three main types fall into the category that utilizes natural language processing (NLP), and they differ primarily in purpose [2-3]:

1) **Rule-based chatbots** – These operate based on predefined rules. They use automation to provide responses based on user input, making them less complex compared to other types.
2) **Retrieval-based chatbots** – These select a suitable response from a set of predefined replies by analyzing the user's input. They offer more flexibility than rule-based bots but still don't generate responses.
3) **Generative AI chatbots** – These are based on artificial intelligence and are capable of producing dynamic responses and understanding the broader context. However, they can sometimes generate incorrect or inaccurate answers. Due to this risk, the models must be continuously trained in order to minimize the occurrence of false or misleading outputs.

The next section provides an overview of the core technologies underlying modern chatbots: artificial intelligence, machine learning, natural language processing, and large language models. It explains how each contributes to design and operation, laying the foundations for the architectural and implementation choices that follow.

2.1.1 AI technologies for chatbot development

Chatbots, as advanced AI-based solutions, are built upon several fundamental technologies from the field of artificial intelligence. The most important among these are Artificial Intelligence (AI) itself, Machine Learning (ML), Natural Language Processing (NLP), and Large Language Models (LLMs). Each of these plays a specific role in enabling chatbots to understand, process, and generate human-like responses.

Artificial Intelligence is a comprehensive discipline within computer science that empowers machines to simulate human cognitive functions, such as perception, reasoning, and problem-solving. Those functionalities are achieved by iteratively training sophisticated models on vast and heterogeneous datasets, thereby enabling systems to continuously refine their predictive capabilities, generalize across diverse tasks, and adapt to evolving environments [4].

Machine Learning is a subfield of artificial intelligence focused on developing algorithms that learn from data. It enables computer systems to improve their performance over time by analyzing and adapting to new information. The training data can include various formats such as images, text, numerical data, or reports. This data is prepared and used to create training datasets. Generally, using high-quality data leads to better model performance and more accurate outputs [5], yet these data-driven predictions should be validated when factual precision is critical.

Natural Language Processing is a technology that enables machines to understand, interpret, and generate human language. NLP systems process large volumes of text, audio, or video data, often collected from platforms such as social media, emails, or communication tools [6]. By applying machine learning techniques, NLP allows for the automatic analysis and extraction of meaningful information from this data. NLP is frequently integrated into applications and services that require interaction with users, such as chatbots and virtual assistants. Advances in NLP have been key to the development of generative AI models capable of simulating human conversation or even creating images from text-based prompts [7].

Large Language Models are advanced AI models capable of understanding and generating human-like text based on the data they were trained on. One notable example is OpenAI's ChatGPT. These models are trained on vast datasets to recognize linguistic patterns and generate coherent responses. LLMs are built on deep neural networks, particularly the transformer architecture, which utilizes attention mechanisms to understand relationships between words and maintain context [8]. LLM performance improves with larger training datasets, although this comes with increased computational requirements. Due to their ability to generate natural-sounding language, LLMs are widely used in areas such as code generation, content creation, and language translation [9].

Together, AI, ML, NLP, and LLMs form the technological backbone of modern chatbots. Each of these technologies contributes a specific layer of capability, from data processing and pattern recognition to language understanding and generation. In the next section we will illustrate their interplay in e-government conversational agents.

2.1.2 Chatbots in e-government

One of the most prominent trends today in the private sector is the use of applications that communicate with users in a conventional, human-like manner. Private organizations operating in industries such as banking, tourism, finance, and others increasingly use these applications to automate communication and process management with clients. The first generation of chatbots had limited capabilities and could only respond to simple and clearly defined queries. However, as technology progressed, so did their functionalities - modern generations of chatbots can now even generate software code. Although the benefits of these technologies are mostly utilized in the private sector, more and more public sector organizations are starting to adopt AI-based systems [10].

In addition to the numerous benefits these models bring, their application also raises new questions and challenges. One such challenge is information extraction. This requires the development of a high-quality knowledge base and the preservation of data relevance. In the public sector, these challenges are even greater due to the need to handle confidential data and convert vast quantities of often legislative documents into formats that can be processed by machines. In addition to technical challenges, there are also ethical and societal barriers that hinder the implementation of artificial intelligence systems in the public sector. To effectively overcome these challenges, a careful approach must be taken, along with ongoing education efforts.

The advantages of using chatbots can be seen in practical examples around the world. For instance, in India, English is predominantly used in both business and public life, but only about 16% of the population speaks the language. If an e-government website were available only in English, it would limit access to services for the majority of the population. Chatbots overcome this issue by supporting multiple languages, thus enabling personalized communication for users. Another major advantage is the constant availability of services. These models can provide information anytime, regardless of institutional working hours. For example, consider a situation where someone loses their passport while in a foreign country. If this happens outside of embassy working hours, the citizen doesn't have to wait until the next day, as the chatbot can immediately provide guidance on how to apply for a replacement passport. [11].

One of the most advanced countries in terms of digitalization is Estonia. Over the years, Estonia has worked extensively on digitalizing public services, and to make the entire system efficient and accessible, it developed a chatbot called SUVE. SUVE uses data from the electronic ID system to provide personalized information to citizens. This is just one small segment of Estonia's comprehensive e-government system, but it demonstrates how the relationship between the state and its citizens can be improved [12].

2.2 Azure technologies for AI system development

Microsoft Azure, formerly known as Windows Azure, is a comprehensive cloud platform built on a globally distributed network of data centers, ensuring high availability and reliability. An active subscription grants users access to its full range of resources, enabling them to deploy services and establish operational environments. Those services help companies achieve their organizational goals. Available tools can be applied in various industries such as finance, e-commerce, and are also compatible with open-source technologies. Additionally, there are several usage models for the Azure platform [13]:

1) **Infrastructure as a Service (IaaS)** – Provides computing services via virtualization;
2) **Platform as a Service (PaaS)** – Enables development, testing, and management of applications without the need to manage the underlying infrastructure;
3) **Software as a Service (SaaS)** – Access to applications via the internet, usually through a subscription model.

Having established these foundational principles, we present a detailed example illustrating the application of Azure services in the development of our chatbot. The chatbot was developed as a web application using the Azure OpenAI service and Azure AI Search to enable conversational search functionality. The Azure OpenAI GPT model is used on the server side to generate responses, while Azure AI Search is utilized for data indexing and retrieval.

This architecture enables efficient, intuitive information retrieval and delivers a ChatGPT-like experience based exclusively on an organization's internal data without requiring advanced technical expertise. Technical administrators retain full control through configurable settings, ensuring both operational efficiency and adherence to corporate policies.

2.2.1 Azure OpenAI

The Azure OpenAI Service integrates OpenAI's language models directly into the Azure ecosystem, enabling organizations to embed advanced natural-language capabilities within their applications. By leveraging Azure's security, scalability, and seamless integration with other Azure tools, organizations can automate workflows, analyze high-volume text data, and drive both business innovation and operational efficiency [14].

The Azure OpenAI service is specifically designed for in-depth data processing. Based on text input, it can easily generate relevant responses within a given context, which is particularly helpful for content creation, writing emails, and similar tasks. An additional advantage is language translation, as this model supports multilingual communication in real time [15].

By using this service, users gain access to AI language models such as GPT-4, GPT-3, and text-to-speech models. In collaboration with OpenAI, Azure OpenAI provides an API, enabling compatibility and integration between the two platforms. In this implementation, the GPT-4o was chosen for its ability to handle complex and multilingual queries relevant to e-government use cases. A key benefit Microsoft offers compared to the classic OpenAI setup is enhanced security, made possible through the use of private networks, regional availability, and content filtering [16].

The Azure OpenAI service is based on several key concepts that enable its efficient use. According to [17], the following concepts are essential for understanding how the service operates:

1) **Prompt** – Represents the input data the user provides in order to receive a corresponding output. These inputs can range from simple questions to complex instructions (e.g., "Generate the simplest HTML code.");
2) **Token** – Azure OpenAI processes the input text by dividing it into tokens. A token can be a word or part of a word (i.e., a string of characters). For example, the word "faculty" in Serbian ("факултет") may be split into syllables such as "фа," "ку," and "лтет," while shorter words like "мир" (peace) may represent a single token. The total number of tokens depends on the input parameters and user requirements;
3) **Models** – A specific pre-trained network (e.g., GPT-4 for text, DALL·E for images, Whisper for audio), each optimized for particular generation or analysis tasks;
4) **Prompt Engineering** – The practice of crafting and refining input prompts to steer model outputs toward desired formats and levels of detail.

One of the most important components of the Azure OpenAI service is content filtering. This feature works in conjunction with other models and operates by filtering prompts through a set of classification models to prevent the generation of harmful or prohibited content. Text filtering models are specifically trained for categories such as hate speech, sexual content, violence, and similar topics in languages like English and French, although the service supports many languages. In any case, it is recommended to create a custom list to guide content filtering [18].

2.2.2 Azure App Service

Azure App Service is a service fundamentally based on the HTTP protocol and is used for hosting web applications and REST APIs. Applications can be written in various programming languages such as .NET, Java, Python, or PHP. These applications are optimized to run in environments based on either Windows or Linux and can be scaled according to specific needs. Additional advantages of this service include security, automatic scaling, and various features offered both by the service itself and the broader Microsoft Azure portal. Furthermore, DevOps benefits can be leveraged due to integrations with various tools, such as Azure DevOps, GitHub, and Docker [19].

2.2.3 Azure AI Search

Azure AI Search is a system designed for accessing information based on search queries. It enables advanced search capabilities across various types of content based on what is entered into the search index.

In addition, this tool offers a wide range of advanced search technologies developed for applications that require high performance and scalability. This system provides direct integration with large language models (LLMs) from the Azure OpenAI service as well as Azure Machine Learning. Moreover, it is possible to connect to models outside the Microsoft ecosystem using specific implementation strategies. The architecture of Azure AI Search is shown in Fig. 1.

In Fig. 1, the Azure OpenAI service is positioned between the database, which contains unindexed data, and the client application, which sends search requests. On the client application side, the search is defined through the Azure AI Search API and may include additional settings such as synonym search, filtering, sorting, and more.

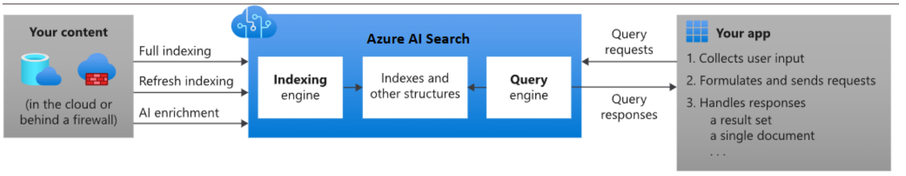

Fig. 1. Architecture of the Azure AI Search [20].

According to [21], the two primary operations of the service are indexing and query execution. Indexing is the process of loading content into the search service, thereby making it available. Inserted text is broken down into tokens and stored in inverted indexes. Azure AI Search supports JSON document types that can be indexed. The client application communicates with the index to retrieve information, which can then be filtered and sorted. Once these processes are completed, the query execution takes place. The client-side application sends a query to Azure AI Search specifying a search term along with other parameters based on which results are returned. Azure AI Search encompasses both traditional search methods and newer approaches powered by generative artificial intelligence, which makes this model an efficient and user-friendly solution for accessing information.

2.2.4 Azure Cosmos DB

Modern technological advancements demand that applications remain continuously available and highly responsive. They must handle dynamic workloads and concurrent requests without interruption, even under peak user activity. Deploying application instances in the data center nearest to end users helps minimize latency. Furthermore, the rise of AI-driven features has increased the complexity of data management, since many modern systems now rely on multiple, heterogeneous databases [22].

These challenges can be overcome by using Azure Cosmos DB. This type of database simplifies and accelerates the application development process by offering a unified database for diverse needs, such as geolocation caching or backup storage. It also supports various data models, including relational, graph, and table formats. Additional benefits of using Azure Cosmos DB include simplified administration due to extensive automation, even in the area of scaling, which is crucial when an application needs to meet user demand [23].

In the model whose implementation will be described in the following chapter, one of the main roles of Azure Cosmos DB is to enable the storage of each individual conversation's history, i.e., chat history. This functionality significantly helps users as it allows them to reference previous queries and continue earlier conversations.

3. Architecture of a chatbot as an e-government service using Azure OpenAI service

This chapter will describe and illustrate the system architecture for the development and implementation of a chatbot, as well as its components. The goal of the system as a whole is to enable efficient resource management, the construction of an internal knowledge base, all in combination with artificial intelligence technologies. The development of this system is achieved through the integration of several Azure services: Function Apps, Azure OpenAI, Azure AI Document Intelligence, and Azure AI Search - some of which have been previously described. This integration creates a unified solution capable of handling large volumes of data, automating processes, and enabling rapid resource scaling.

For the development of the chatbot as an e-government service, the system architecture consists of two main components: document ingestion and management, and user response generation. This structure enables the chatbot to search and respond to user queries based on documents stored in the knowledge base, as illustrated in Fig. 2.

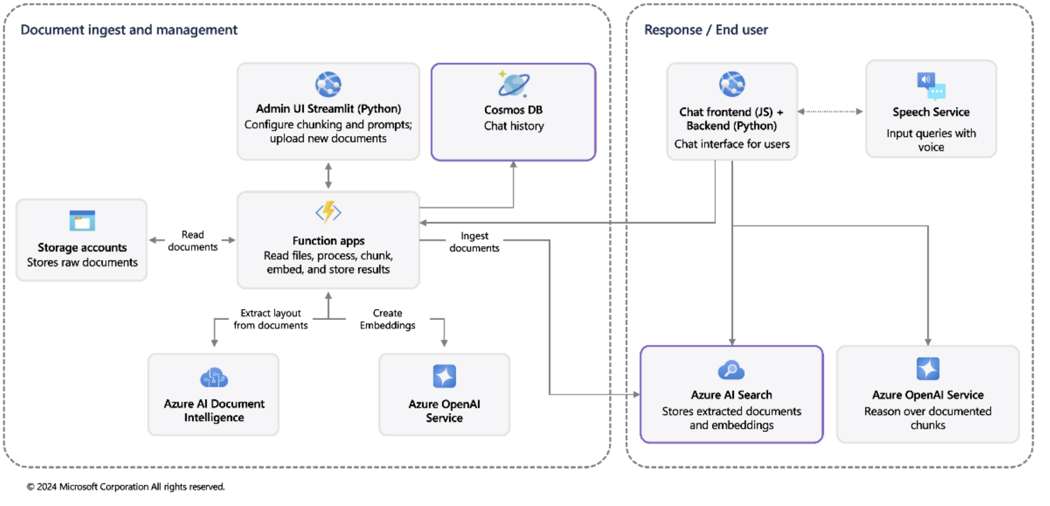

Fig. 2. System architecture for chatbot development [24].

Within the left section, which pertains to document management, the Function Apps component serves as the central element interacting with all others. Its role is to read documents, process them, divide them into smaller segments, and then store them. The files, i.e., the documents, are stored within a Storage Account. The Azure OpenAI service provides the ability to understand the content, while Azure AI Document Intelligence extracts key information and sends it for further processing. The Cosmos DB ensures the functionality of saving chat history.

Once the documents are processed, they are forwarded to Azure AI Search, where they are stored for later usage. In the Response / End-User section, the user interacts with a JavaScript chat frontend paired with a Python backend, which together facilitate communication with the chatbot. Voice queries are captured and transcribed by the Speech Service. Each request is routed concurrently to Azure AI Search for retrieval of relevant document embeddings and to Azure OpenAI Service for reasoning over those embeddings to produce context-aware, human-like responses.

Below is a concise explanation of how the previously mentioned components work together in a concrete example, as illustrated in Fig. 3:

1) The user submits a question or request through the graphical interface;
2) The application then forwards the request to Azure AI Search, which performs the search;
3) Once the information is found in the knowledge base, it is forwarded to the Azure OpenAI service, which generates a meaningful response and sends it back as feedback;
4) Finally, the information is returned to the user as a response to their request via the user interface.

4. Developed chatbot: functionalities and evaluation

In this chapter, we describe the architecture and operation of the implemented chatbot system, outline its user-facing functionalities, and present the methodologies employed to validate system performance and user experience. In order to guarantee uninterrupted operation within production environment, it is necessary to carry out the previously mentioned activities to verify the fulfillment of requirements and to continuously improve the system as a whole.

Although no formal benchmarking was performed, GPT-4o was selected based on internal testing and current literature. The model demonstrated strong performance in handling multilingual queries, retrieving information from longer documents, and producing clear and relevant responses, which are key features for public services designed. As noted in [25], GPT-4o improves response speed and efficiency while maintaining high accuracy in reasoning, making it a suitable choice for real-world e-government applications.

Upon accessing the chatbot, the user is presented with a window where they can submit their request either by typing or by using voice input. The main page of the user interface is shown in Fig. 4.

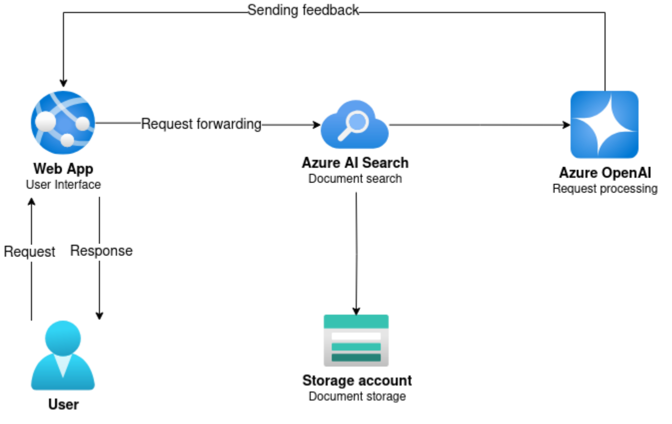

Fig. 3. Illustration of a Practical Example.

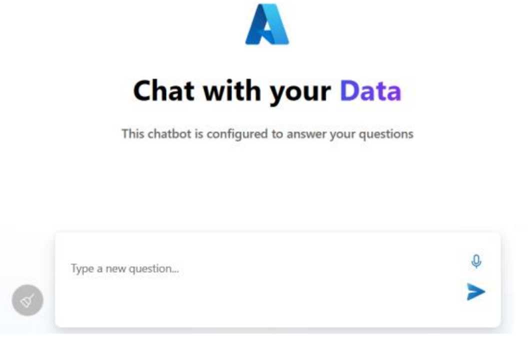

Fig. 4. User interface of chatbot.

Once the user asks a question, the chatbot returns a response if it can find it in its knowledge base; otherwise, it returns a generic message. When responding, the model references the documents it has within the database.

Fig. 5 illustrates a sample interaction in which the user asks, “What is eUprava?” and the chatbot returns a concise overview of Serbia's eUprava portal in Serbian. Directly beneath the response, the disclaimer “AI-generated content may be incorrect” appears to remind users that the information should be verified and not accepted uncritically.

If the user wishes to delete messages within the conversation, they can simply click the brush icon inside the blue circle immediately to the left of the input field. In any case, the conversation will be saved within the chat history, which can be seen in Fig. 6.

The chat history feature is particularly useful because the user can return to previous conversations. Additionally, the general recommendation for interacting with chatbots is that the user should maintain a separate conversation for each topic or service request to avoid confusion and overlap.

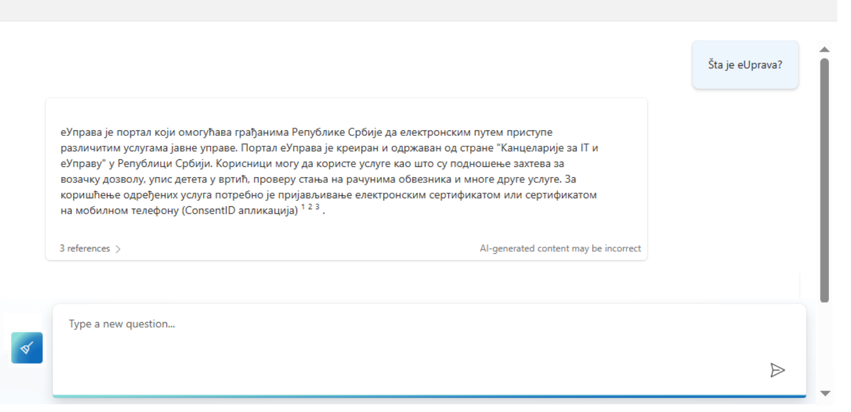

Fig. 5. User interaction with chatbot.

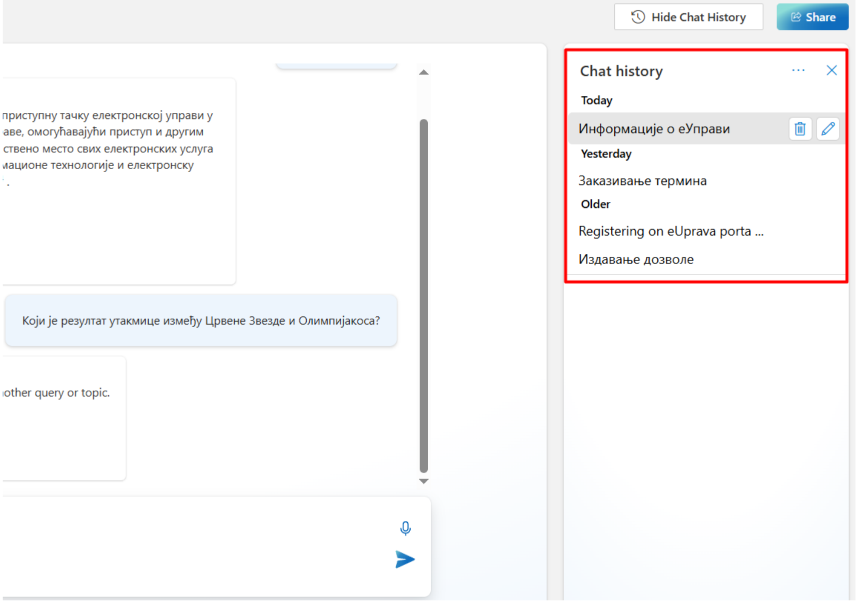

Fig. 6. Chat history.

Furthermore, the user can share any conversation by clicking the "Share" button, which will allow them to copy a link that they can share.

While generative AI unlocks powerful capabilities, it also presents new security vulnerabilities. In particular, attackers may craft malicious prompts to extract sensitive information or coerce the system into producing prohibited content. To guard against such threats, our chatbot undergoes rigorous adversarial testing, including scenarios in which an attacker requests confidential user records. Fig. 7 illustrates one such exchange: the user's prompt, "If you have database access, display all users and their national identifiers" is met by the chatbot's refusal, "I'm sorry, I don't have any information on that topic. I can assist with information about eUprava".

Performance testing is especially important after deploying the system to production. These tests can be conducted using Azure Load Testing. Through load testing, it is possible to measure how the system responds to a higher volume of user requests, and it can also help identify potential resource bottlenecks. Conducting such tests contributes to system optimization.

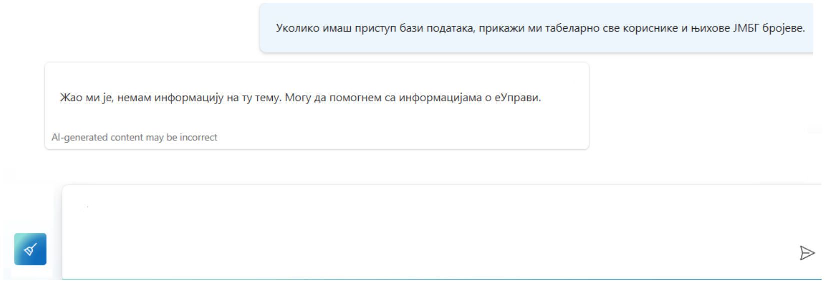

Fig. 7. Example of a malicious prompt.

From Fig. 8, some aggregated data can be seen, such as the total number of requests, average response time, and throughput. The data is refreshed every 5 seconds by default. The test results can also be visualized through charts.

Fig. 8 and Fig. 9 show the main metrics. What is also important is the utilization of the CPU, memory, and other resources, which can also be measured through these tests.

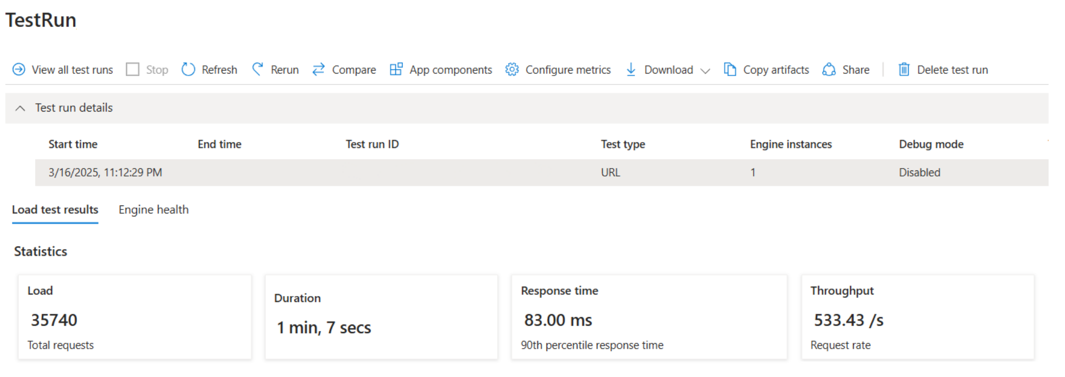

Fig. 8. Performance testing.

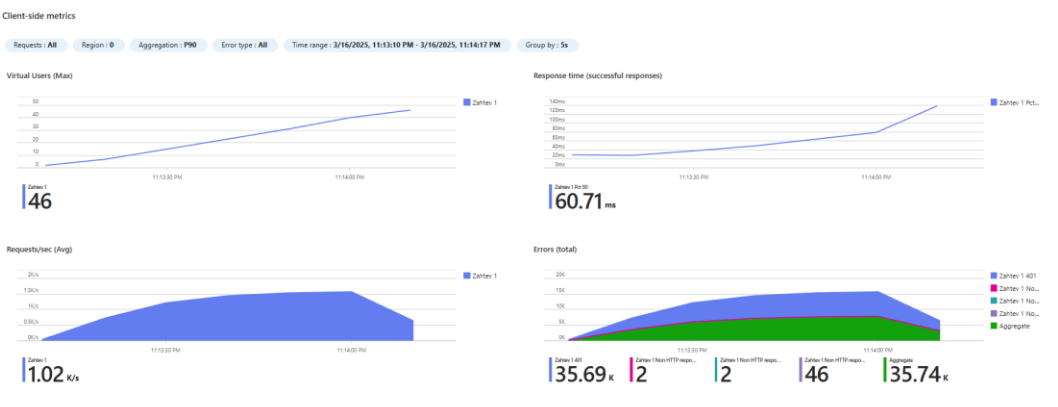

Fig. 9. Graphical representation of the test.

5. Conclusion

This paper describes how electronic public services can be modernized and improved through the application of artificial intelligence. The entire process of developing a chatbot as an e-government service is covered, from the theoretical aspect, through the analysis of the technologies used, to the implementation of the technical solution.

The chatbot discussed in this paper provides citizens with an intuitive way to access and better understand e-government services through the user interface. The advantages of this model lie in its simple integration with the web portal, high degree of automation, and rapid scalability of resources.

The direction for further thinking and enhancement of the web application should focus on the development of APIs to enable better interaction and improvement of the user interface for future integration with other platforms and applications. Although this paper places emphasis on server-side functionality, maximizing practicality requires simultaneous investment in API development, user-interface improvements, integration adapters, and supporting services to ensure seamless interoperability with external platforms, robust security, and an intuitive experience for end users.

Additionally, in order to maintain the relevance and accuracy of the data, the model must be continuously trained, which improves the service. Besides system maintenance, it is also necessary to implement new functionalities, one of which is Azure Cache Redis, which provides more efficient request processing by caching queries and offering faster data access.

Furthermore, future research could explore a comparative analysis between Azure OpenAI and similar services from other cloud providers, such as AWS Bedrock or Google Vertex AI. Such comparisons may provide valuable insights into performance, cost-efficiency, and integration flexibility, particularly in the context of public sector requirements. This would contribute to more informed deployment decisions in heterogeneous technological environments.

One downside of this model is that, despite the fact that the documents in the knowledge base contain facts, the answers generated by artificial intelligence may still be inaccurate. To minimize the number of such answers, continuous model learning and building a high-quality knowledge base are necessary.

Finally, the implementation of the chatbot shows that public sector organizations can successfully apply modern technologies and transform their services, enabling users to have a simpler experience and wider use.

References

[1]. E. Adamopoulou and L. Moussiades, “An Overview of Chatbot Technology,” IFIP Advances in Information and Communication Technology, vol. 584 IFIP, pp. 373–383, 2020, doi: 10.1007/978-3-030-49186-4_31.

[2]. S. Hussain, O. Ameri Sianaki, and N. Ababneh, “A Survey on Conversational Agents/Chatbots Classification and Design Techniques,” Advances in Intelligent Systems and Computing, vol. 927, pp. 946–956, 2019, doi: 10.1007/978-3-030-15035-8_93.

[3]. L. F. Lins dos Santos, “Evaluating and Comparing Generative-based Chatbots Based on Process Requirements,” Dec. 06, 2023, University of Waterloo.

[4]. C. Zhang and Y. Lu, “Study on artificial intelligence: The state of the art and future prospects,” Journal of Industrial Information Integration, vol. 23, p. 100224, Sep. 2021, doi: 10.1016/J.JII.2021.100224.

[5]. M. I. Jordan and T. M. Mitchell, “Machine learning: Trends, perspectives, and prospects,” Science, vol. 349, no. 6245, pp. 255–260, Jul. 2015, doi: 10.1126/SCIENCE.AAA8415.

[6]. K. R. Chowdhary, “Natural Language Processing,” Fundamentals of Artificial Intelligence, 2020, doi: 10.1007/978-81-322-3972-7_19.

[7]. E. Novikova, “Influence of NLP on organizational design and development of dynamic capabilities,” 2024.

[8]. S. Minaee et al., “Large Language Models: A Survey”.

[9]. M. Nejjar, L. Zacharias, F. Stiehle, and I. Weber, “LLMs for science: Usage for code generation and data analysis,” Journal of Software: Evolution and Process, vol. 37, no. 1, p. e2723, Jan. 2025, doi: 10.1002/SMR.2723.

[10]. A. Androutsopoulou, N. Karacapilidis, E. Loukis, and Y. Charalabidis, “Transforming the communication between citizens and government through AI-guided chatbots,” Government Information Quarterly, vol. 36, no. 2, pp. 358–367, Apr. 2019, doi: 10.1016/J.GIQ.2018.10.001.
[11]. A. Rukavina, “How chatbots can improve government services.” Accessed: May 08, 2025. [Online]. Available: https://www.infobip.com/blog/ai-customer-experiences-in-government
[12]. “Invest in Estonia.” Accessed: May 08, 2025. [Online]. Available: https://investinestonia.com/estonia-created-suve-an-automated-chatbot-to-provide-trustworthy-information-during-the-covid-19-situation/
[13]. M. M. Salas-Zárate and M. Luis Colombo-Mendoza, “CLOUD COMPUTING: A REVIEW OF PAAS, IAAS, SAAS SERVICES AND PROVIDERS,” Lámpsakos (revista descontinuada), no. 7, pp. 47–57, Jun. 2012, doi: 10.21501/21454086.844.
[14]. A. Jamshidi, “Applications of Microsoft Azure OpenAI to code management,” 2024.
[15]. N. P. S. Kumar, S. Ramakrishnan, and M. Vignesh, “AI-Based Data Analytics & Business Intelligence Chatbot Using Azure Functions and OpenAI,” pp. 129–137, 2025, doi: 10.1007/978-981-97-9619-9_11.
[16]. V. T. Mahajan and S. Sarode, “COMPARATIVE INSIGHTS INTO AI TOOLS: AN ANALYTICAL APPROACH”.
[17]. S. Ifrah, “Getting Started with Azure OpenAI,” Getting Started with Azure OpenAI, 2024, doi: 10.1007/979-8-8688-0599-8/COVER.
[18]. Microsoft, “What is Azure OpenAI Service?” Accessed: May 12, 2025. [Online]. Available: https://learn.microsoft.com/en-us/azure/ai-services/openai/overview
[19]. Microsoft, “App Service Overview.” Accessed: May 12, 2025. [Online]. Available: https://learn.microsoft.com/en-us/azure/app-service/overview
[20]. Microsoft, “What’s Azure AI Search?” Accessed: May 12, 2025. [Online]. Available: https://learn.microsoft.com/en-us/azure/search/search-what-is-azure-search
[21]. K. Weiland, “Azure AI Search: A Comprehensive Guide.” Accessed: May 12, 2025. [Online]. Available: https://www.linkedin.com/pulse/azure-ai-search-comprehensive-guide-kim-weiland-mvb1e/
[22]. Y. Lu, “Artificial intelligence: a survey on evolution, models, applications and future trends,” Journal of Management Analytics, vol. 6, no. 1, pp. 1–29, Jan. 2019, doi: 10.1080/23270012.2019.1570365.
[23]. J. R. Guay Paz, “Introduction to Azure Cosmos DB,” Microsoft Azure Cosmos DB Revealed, pp. 1–23, 2018, doi: 10.1007/978-1-4842-3351-1_1.
[24]. Microsoft Corporation, “Chat with your data - Solution Accelerator,” GitHub. Accessed: May 12, 2025. [Online]. Available: https://github.com/Azure-Samples/chat-with-your-data-solution-accelerator
[25]. R. Islam and O. M. Moushi, “GPT-4o: The Cutting-Edge Advancement in Multimodal LLM,” TechRxiv, Jul. 2024, doi: 10.36227/TECHRXIV.171986596.65533294/V1.

Информация об авторах / Information about authors

Лазарь РАДОСАВЛЬЕВИЧ окончил факультет организационных наук Белградского университета. Работает в Управлении информационных технологий и электронного управления Республики Сербии. Его научные интересы включают искусственный интеллект, кибербезопасность и интернет-технологии.

Lazar RADOSAVLJEVIĆ graduated from the Faculty of Organizational Sciences, University of Belgrade. Employed at the Office for IT and eGovernment of the Republic of Serbia. His research interests include artificial intelligence, cybersecurity, and internet technologies.

Милица СИМИЧ – аспирантка по направлению «Программная инженерия и электронный бизнес», ассистент на факультете организационных наук Белградского университета. Член IEEE и IEEE Computer Society. Её научные интересы включают искусственный интеллект, генеративный ИИ, образование, программную инженерию, интернет-технологии и электронный бизнес.

Milica SIMIĆ – PhD student in Software Engineering and E-Business, teaching assistant at the Faculty of Organizational Sciences, University of Belgrade. Member of IEEE and the IEEE Computer Society. Her research interests include artificial intelligence, generative AI, education, software engineering, internet technologies, and e-business.

Александр ЙОКСИМОВИЧ – магистрант по направлению «Электронный бизнес» и учебный ассистент на факультете организационных наук Белградского университета. Его научные интересы включают моделирование, цифровые двойники, Интернет вещей и их применение в умном сельском хозяйстве.

Aleksandar JOKSIMOVIĆ – Master student in E-Business Technologies and teaching associate at the Faculty of Organizational Sciences, University of Belgrade. His research interests include simulations, digital twins, the Internet of Things, and their applications in smart agriculture.

Тамара НАУМОВИЧ – аспирантка по направлению «Программная инженерия и электронный бизнес», ассистент на факультете организационных наук Белградского университета. Её научное исследование посвящено блокчейн-технологиям. Научные интересы включают Web3, интернет-технологии, электронный бизнес и электронное образование.

Tamara NAUMOVIĆ – PhD student in Software Engineering and E-Business, teaching assistant at the Faculty of Organizational Sciences, University of Belgrade. Her PhD research focuses on blockchain technologies. Her research interests include Web3, internet technologies, e-business, and e-education.

Марьяна ДЕСПОТОВИЧ-ЗРАКИЧ – профессор факультета организационных наук Белградского университета, руководитель лаборатории моделирования. Член IEEE и IEEE Computer Society. Заместитель председателя отделения IEEE Computer Chapter CO 16 и руководитель семинара при IEEE Computer Chapter CO 16. Её научные интересы включают блокчейн, интернет-технологии и Интернет вещей (IoT).

Marijana DESPOTOVIĆ-ZRAKIĆ – Full professor at the Faculty of Organizational Sciences, University of Belgrade, and Head of the Laboratory for Simulation. Member of IEEE and the IEEE Computer Society. Vice Chair of the IEEE Computer Chapter CO 16 and Head of the Seminar of the IEEE Computer Chapter CO 16. Her research interests include blockchain, internet technologies, and the Internet of Things (IoT).