A Multilingual Chatbot for Migrants: Concept and Implementation

Maria Papoutsoglou[1,*], Alexandros Tassios[1], Stergios Tegos[1], Christos Bouas[1], Konstantinos Manousaridis[1], Maria Kaltsa[2,3], Thanassis Mavropoulos[3], Stefanos Vrochidis[3] and Georgios Meditskos[1]

[1]School of Informatics, Faculty of Sciences, Aristotle University of Thessaloniki, Greece
[2]Department of Theoretical & Applied Linguistics, School of English, Faculty of Philosophy, Aristotle University of Thessaloniki, Greece
[3]Information Technologies Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece

Abstract

Migration is one of the main societal challenges faced by many countries in the European Union (EU). Thus, efficient and transparent processes to achieve the swift registration, health support, and social integration of Third Country Nationals (TCNs) is imperative. Although governments usually provide essential information through various official websites, these resources often lack multilingual accessibility. In some cases, critical content is available in the hosting country's native language. This language barrier can prevent migrants from accessing important information about public services, migration policies, and asylum procedures. This paper presents the development framework for the SALLY chatbot, designed to address some of these challenges by leveraging advancements in generative AI. The proposed solution integrates state-of-the-art technologies in Large Language Models, Dialogue Management, Knowledge Graphs, and Information Retrieval to facilitate personalized, multilingual interactions.

Keywords

Multilingual, Chatbot, Natural Language Processing, Large Language Models, migrants

1. Introduction

Third Country Nationals (TCNs) often face significant challenges integrating into host countries due to differences in education, culture, language, and legal status [1]. To support their inclusion, accessible information and a welcoming environment are essential. Clear guidance on healthcare, education, employment, and social services remains a key need for TCNs. While digital platforms exist, they are often difficult to navigate. AI—particularly chatbot technology [2], —offers a promising solution by providing personalized, real-time support[3, 4].

The SALLY project (Semantically Conscious Conversation-Based Chatbot Services for Migrants) aims to develop an intelligent chatbot that assists migrants in Greece. Combining Large Language Models [5], Dialogue Management [6], Knowledge Graphs [7], Sentiment Analysis [8], and Information Retrieval [9]. SALLY delivers context-aware conversations tailored to users' needs. A central goal is strong support for Greek, a low-resource language, enabling more natural, meaningful interactions. This paper presents the chatbot's framework and explores its benefits and potential challenges.

4th International Conference on "Multilingual digital terminology today. Design, representation formats and management systems" (MDTT) 2025, June 19-20, 2025, Thessaloniki, Greece.

*Corresponding author.
mpapo@csd.auth.gr (M. Papoutsoglou); tassiosa@csd.auth.gr (A. Tassios); stergios@enchatted.com (S. Tegos); chrisbouas@gmail.com (C. Bouas); manousaridis@gmail.com (K. Manousaridis); mkaltsa@iti.gr (M. Kaltsa); mavrathan@iti.gr (T. Mavropoulos); stefanos@iti.gr (S. Vrochidis); gmeditsk@csd.auth.gr (G. Meditskos)
0000-0003-0658-5065 (M. Papoutsoglou); 0009-0009-7922-339X (A. Tassios); 0009-0000-3857-1552 (S. Tegos); 0009-0002-8809-3303 (C. Bouas); 0009-0003-0557-2542 (K. Manousaridis); 0000-0002-2422-7889 (M. Kaltsa); 0000-0002-7326-5910 (T. Mavropoulos); 0000-0002-2505-9178 (S. Vrochidis); 0000-0003-4242-5245 (G. Meditskos)

2. Related Work

Chatbot technology that supports individuals in different domains is a widely researched topic. For example, CataractBot [10] is a multilingual chatbot designed in collaboration with a tertiary eye hospital in India to provide expert-verified, reliable information about cataract surgery using a curated knowledge base. Using GPT-4 and a Retrieval Augmented Generation (RAG) approach, it integrates language technologies, a vector database, and WhatsApp services to offer accurate, concise responses in multiple formats and Indic languages, addressing challenges of information overload and limited communication time with healthcare professionals.

The author in [11] present a chatbot aims to enhance e-commerce customer service by delivering personalized, sentiment-aware responses tailored to users' interests and emotional states. Using advanced NLP techniques, including intent classification, sentiment analysis, and information extraction, alongside a BERT-based language model, it provides accurate, contextually relevant answers while maintaining coherent multi-turn dialogue and engaging user interactions.

In [12] a chatbot aids students in understanding learning-path recommendations by providing clear, accurate explanations and connecting them with human mentors when needed is proposed. The system brings together large language models (LLMs) and a knowledge graph (KG) to give more accurate answers, focusing specifically on learning-path questions. It also considers the user's context, checks their intent, and includes backup options—like involving a mentor—when needed to keep the conversation helpful and on track.

In another case, a chatbot [13] created for the Museum of Paleontology and Geology in Athens helps remote visitors explore exhibits by suggesting items and sharing multimedia content in both English and Greek. Built with Rasa's DIETClassifier and supported by a knowledge graph, it uses customized natural language processing, including named entity recognition trained on synthetic data, to respond to detailed paleontology-related questions and make the experience more engaging.

Shifting to migrant support, the NADINE-bot [14] was developed to help asylum seekers and vulnerable migrants by answering administrative questions about EU countries in their native languages. It combines translation tools, a Universal Sentence Encoder, and a two-step retrieval method to find relevant answers. A dialogue manager keeps the interaction smooth, even allowing for small talk to make the experience more user-friendly. Moreover, MyMigrationBot [15] is a cloud-based social chatbot deployed on Facebook to support migrants by providing personalized feedback on personality traits and job-competency fit, aiding labor market integration. Powered by Twilio Autopilot and hosted on AWS, it uses machine learning for dialogue management while enabling admin monitoring through Facebook's admin panel to ensure unbiased and effective interactions.

In contrast to these works, the proposed chatbot focuses on the assistance of migrants in governmental related tasks and utilizes a plethora of state of the art technologies such as Large Language Models, Dialogue Management, Knowledge Graphs, and Information Retrieval.

3. Proposed Framework

The architecture of the SALLY chatbot has been designed to address the complex requirements of information delivery and interaction management for migrants in Greece. It combines multiple interconnected components that enable both natural language understanding and the delivery of precise and contextually appropriate information. Figure 1 provides an overview of the system's architecture.

At the heart of the system is a Knowledge Base that organizes practical information related to migration procedures, services, contact points, facilities, and available benefits. To keep this data current, a dedicated web crawler regularly pulls updates from the official website of the Greek Ministry of Migration and Asylum. The collected content is grouped into five main hubs—procedures, benefits, services, facilities, and contacts—making it easier to retrieve the right information during conversations.

SALLY needs to be able to handle real-world conversations that may be informal, unclear, or multilingual. To make sense of user questions and generate helpful replies, SALLY uses Large Language Models

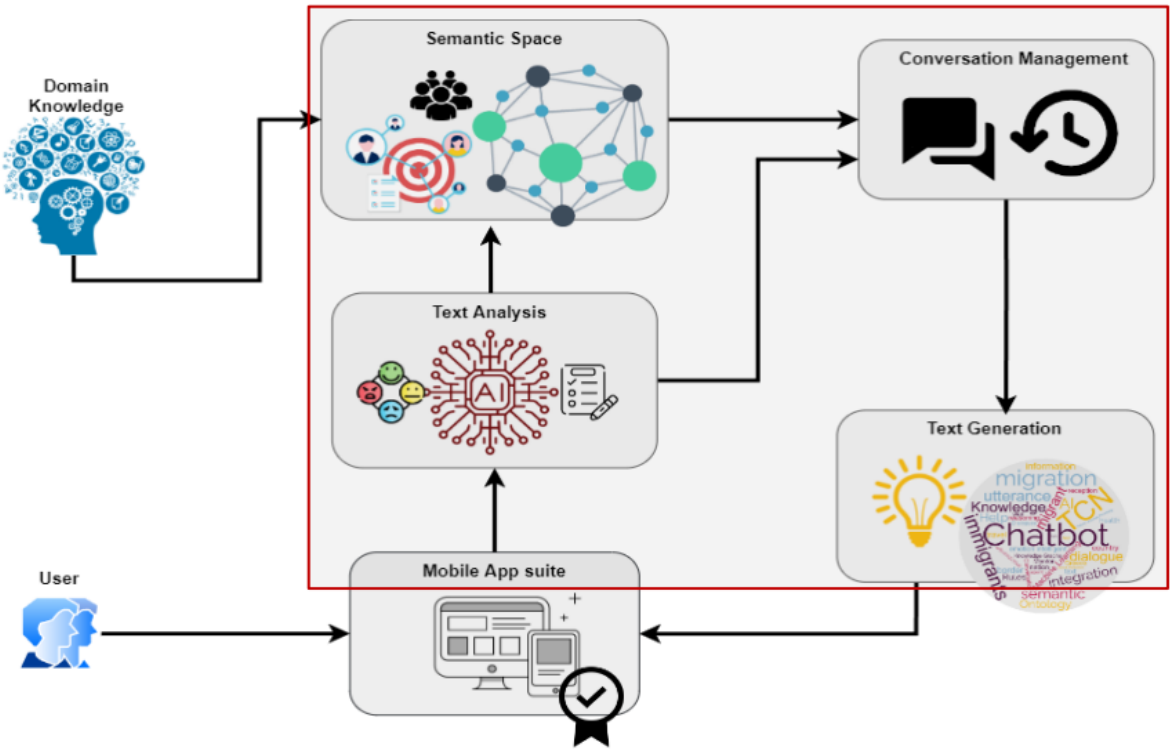

Figure 1: Architecture of the SALLY chatbot showing its main components and data flow.

(LLMs) such as Mistral, LLaMA 3, and GPT-4-Turbo. These models help interpret user input, generate relevant responses, and carry out tasks like detecting sentiment and identifying important entities.

To make sure that responses aren't just fluent, but that they're also accurate and up to date, the system utilizes the Retrieval-Augmented Generation (RAG) process. The RAG component is key to SALLY as it blends the broad language abilities of LLMs with factual information pulled directly from the knowledge base.

To keep conversations on track, a Dialogue Manager oversees the flow of interaction. It keeps track of context, user profiles, and past messages, so conversations stay coherent—even if they're interrupted or continued after a break. It can also detect when a user seems confused or struggles with language, and adjust its responses to offer clearer guidance.

SALLY also makes use of Semantic Conversational Spaces, which connect user data, dialogue history, and domain-specific knowledge through knowledge graphs. This setup helps the system understand relationships between different ideas and respond in a more informed, relevant way. External sources like DBpedia, ConceptNet, and BabelNet are used to enrich these spaces and support deeper understanding of the topics being discussed.

Finally, Moderation Filters are built into the system to help protect users from harmful or inappropriate content. These filters check for issues such as harassment, hate speech, violence, or privacy risks. SALLY combines advanced language tools, structured knowledge, and built-in safety measures to provide a helpful, respectful, and personalized experience for migrants in Greece seeking support.

4. Discussion

The development of the SALLY chatbot highlights how combining knowledge-based systems with large language models can help meet information needs in complex and sensitive areas like migration. Thanks to LLMs, the system can understand natural, everyday language—even when it's informal or influenced by different linguistic backgrounds, as is often the case with migrant users. At the same time, drawing

on a well-structured knowledge base helps ensure that the information provided stays accurate and in line with official procedures. During development, several challenges came up. One of the biggest was keeping the generated responses consistent with verified information from the knowledge base. LLMs are good at sounding fluent, but sometimes they stray from the facts. Using a retrieval-augmented generation (RAG) setup helped reduce this issue by anchoring answers in real data, but more work is needed to fully control what the model produces.

Another challenge involved handling user profiles and keeping track of conversation history—especially when users return after long breaks or interact in unpredictable ways. The Dialogue Manager and Semantic Conversational Spaces helped keep things coherent across sessions, but there's still room for improvement, particularly in personalizing the experience to each user. Privacy and safety were also top priorities, considering how vulnerable the target users may be. We used moderation filters to prevent inappropriate or harmful content, but ongoing updates and monitoring are essential, especially as new issues or user behavior patterns show up. We also had to consider how SALLY would perform in real-world situations. Things like poor internet access, limited devices, or language-related obstacles can impact the user experience. While early testing has been positive under standard conditions, further trials with actual users will be needed in order to understand real-life constraints and make necessary adjustments.

5. Conclusions

The SALLY chatbot marks a meaningful step toward using generative AI to support migrant integration. By bringing together tools like Large Language Models, Dialogue Management, Knowledge Graphs, and Retrieval-Augmented Generation, it creates a conversational experience that feels personal, helpful, and tailored to the needs of migrants in Greece. Its support for both English and Greek allows for smooth communication, helping users find the information they need about services, procedures, and available support. At the same time, built-in safety measures promote respectful, constructive interactions—something especially important for vulnerable users.

Beyond its specific purpose, SALLY also shows how generative AI can help build more inclusive digital services. As more public and private organizations turn to AI-powered tools, systems like SALLY can serve as examples of how to design accessible, multilingual platforms that truly meet people where they are. By making it easier to access accurate information and get real-time help, SALLY not only assists individual users—it also supports better service delivery and stronger connections between migrants and their host communities.

6. Acknowledgments

This research project is carried out within the framework of H.F.R.I call "Basic research Financing (Horizontal support of all Sciences)" under the National Recovery and Resilience Plan "Greece 2.0" funded by the European Union - NextGenerationEU (H.F.R.I. Project Number: 15010).

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT-4 for grammar and spelling checks. The authors have subsequently reviewed and edited the content and take full responsibility for the publication's final version.

References

[1] N. Doerr, Bridging language barriers, bonding against immigrants: A visual case study of transnational network publics created by far-right activists in europe, Discourse & Society 28 (2017) 3–23.

[2] E. Adamopoulou, L. Moussiades, An overview of chatbot technology, in: IFIP international conference on artificial intelligence applications and innovations, Springer, 2020, pp. 373–383.

[3] Z. Chen, Y. Lu, M. P. Nieminen, A. Lucero, Creating a chatbot for and with migrants: chatbot personality drives co-design activities, in: Proceedings of the 2020 acm designing interactive systems conference, 2020, pp. 219–230.

[4] Y.-C. Tseng, W. Jarupreechachan, T.-H. Lee, Understanding the benefits and design of chatbots to meet the healthcare needs of migrant workers, Proceedings of the ACM on Human-Computer Interaction 7 (2023) 1–34.

[5] Y. Chang, X. Wang, J. Wang, Y. Wu, L. Yang, K. Zhu, H. Chen, X. Yi, C. Wang, Y. Wang, et al., A survey on evaluation of large language models, ACM Transactions on Intelligent Systems and Technology 15 (2024) 1–45.

[6] H. Brabra, M. Báez, B. Benatallah, W. Gaaloul, S. Bouguelia, S. Zamanirad, Dialogue management in conversational systems: a review of approaches, challenges, and opportunities, IEEE Transactions on Cognitive and Developmental Systems 14 (2021) 783–798.

[7] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. D. Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, et al., Knowledge graphs, ACM Computing Surveys (Csur) 54 (2021) 1–37.

[8] M. Wankhade, A. C. S. Rao, C. Kulkarni, A survey on sentiment analysis methods, applications, and challenges, Artificial Intelligence Review 55 (2022) 5731–5780.

[9] A. Singhal, et al., Modern information retrieval: A brief overview, IEEE Data Eng. Bull. 24 (2001) 35–43.

[10] P. Ramjee, B. Sachdeva, S. Golechha, S. Kulkarni, G. Fulari, K. Murali, M. Jain, Cataractbot: An llm-powered expert-in-the-loop chatbot for cataract patients, arXiv preprint arXiv:2402.04620 (2024).

[11] A. El-Ansari, A. Beni-Hssane, Sentiment analysis for personalized chatbots in e-commerce applications, Wireless Personal Communications 129 (2023) 1623–1644.

[12] H. Abu-Rasheed, M. H. Abdulsalam, C. Weber, M. Fathi, Supporting student decisions on learning recommendations: An llm-based chatbot with knowledge graph contextualization for conversational explainability and mentoring, arXiv preprint arXiv:2401.08517 (2024).

[13] K. Tsitseklis, G. Stavropoulou, A. Zafeiropoulos, A. Thanou, S. Papavassiliou, Recbot: Virtual museum navigation through a chatbot assistant and personalized recommendations, in: Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization, 2023, pp. 388–396.

[14] A. Lelis, N. Vretos, P. Daras, Nadine-bot: An open domain migrant integration administrative agent, in: 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), IEEE, 2020, pp. 1–6.

[15] A. JASTRZĘBOWSKA-TYCZKOWSKA, K. CHLASTA, P. Sochaczewski, I. GRABOWSKA, Mymigrationbot: A cloud-based facebook social chatbot for migrant populations, Annals of Computer Science and Information Systems 31 (2022).