

Towards Semantically Conscious, Conversation-based Chatbot Services for Migrants

Georgios Meditskos¹[0000-0003-4242-5245], Stergios Tegos³[0009-0000-3857-1552],
Christos Bouas¹, Alexandros Tassios¹, Konstantinos
Manousaridis¹[0009-0003-0557-2542], Maria Papoutsoglou¹[0000-0003-0658-5065],
Thanassis Mavropoulos²[0000-0002-7326-5910], and Stefanos
Vrochidis²[0000-0002-2505-9178]

¹ School of Informatics, Aristotle University of Thessaloniki,
54124, Thessaloniki, Greece

{gmeditsk,stegos,mmchristos,tassiosa,mpapo}@csd.auth.gr,
manousaridis@gmail.com

² Information Technologies Institute, Centre for Research and Technology Hellas,
57001, Thessaloniki, Greece

{mavrathan,stefanos}@iti.gr

Abstract. Many EU countries continue to face significant societal challenges related to the acceptance and integration of Third Country Nationals (TCNs). On one hand, hosting countries need to develop efficient and transparent processes to ensure quick registration, health assistance, integration, and support of TCNs. On the other hand, TCNs often face difficulties finding information about the hosting countries, e.g. about public services, reception centers, migration policies and asylum procedures. This paper presents ongoing work towards the development of SALLY, an intelligent chatbot to assist the information seeking process for migrants, supporting their reception and social integration in Greece. The framework aims to combine a multitude of state-of-the-art technologies on Large Language Models, Dialogue Management, Knowledge Graphs, Sentiment Analysis, and Information Retrieval, to achieve personalized conversational awareness and assist migrants in acquiring information relevant to their needs. The support of the Greek language, which is a low-resource language, and the interaction with the users through smart dialogues, beyond simple question-answering, constitute two key objectives of the framework.

Keywords: Chatbot · Migrants · Large Language Models · Conversational Awareness · Knowledge Graphs.

1 Introduction

In the context of migration, Third Country Nationals (TCNs) often face a number of challenges relevant to their acceptance and integration in the hosting

country. These challenges can vary depending on the TCNs’ educational and cultural background, language barriers, legal status and the specific regulations and procedures the hosting countries implement to support the individuals. Addressing these challenges requires effective and inclusive solutions, able to provide access to essential services and create a welcoming and supportive environment.

In this quest, accessing information about regulations, procedures and services, e.g. healthcare, education, employment, and social services, may serve as a valuable resources and can significantly impact the ease with which migrants are assisted in various aspects of integration. Several countries already provide digital services to migrants, such as online platforms to facilitate access to information, submit applications, language learning, employment opportunities and many others. Nevertheless, the seeking of online guidance and information still remains a time-consuming process.

Artificial Intelligence (AI) solutions can have a significant role in addressing TCNs needs through the more effective and intuitive use of resources. Chatbots, in particular, have evolved significantly and are widely used across various situations to provide comprehensive support and assistance, enhancing the overall experience during time-consuming information seeking, paperwork, documentation, and various procedural tasks [3, 5, 8]. SALLY is motivated by the benefits of using chatbots to assist migrants and develops a conversational agent to further improve user-experience while accessing information through the intelligent combination of state-of-the-art technologies.

More specifically, SALLY is a pioneering initiative aimed at providing crucial information and support to migrants in Greece. This sophisticated virtual assistant is designed to empower migrants on their journey of integration into Greek society, offering guidance on legal procedures, cultural understanding, language learning, and practical aspects of daily life. It draws its knowledge from the official website of the Greek Ministry of Migration and Asylum³, ensuring the reliability and accuracy of the information it provides. At its core, SALLY is a compassionate and informative communication tool, programmed to maintain a respectful and supportive tone throughout its interactions. It acknowledges its limitations and operates transparently, informing users of its specific functions. It’s bilingual capabilities in English and Greek further enhance its accessibility and reach within the migrant community.

The purpose of this paper is to present ongoing work on the development of the chatbot in the context of the “SALLY” project, a Greek national project aiming at developing an intelligent personal assistant for TCNs beyond a simple question-answering system, supporting both the English and Greek languages. The rest of the paper is structured as follows: Section 2 presents related work on chatbots that focus on supporting migrants. Section 3 elaborates on the main concepts and technologies of SALLY, describing also the conceptual architecture. Section 4 describes the workflow and illustrates the basic capabilities of SALLY through examples. Finally, Section 5 concludes our work and discusses future research directions.

³ <https://migration.gov.gr/>

2 Related Work

The advancements in AI technology have fueled the popularity and utility of chatbots across industries [2]. Especially in the context of migration, chatbots have the potential to play an increasingly important role in helping migrants access information and overcome the challenges of integration more effectively, capitalising on different technologies according to the aimed objectives.

Digital Companion [5] is a rule-based chatbot developed to aid migrants and refugees in Europe by facilitating access to essential services. Main concept of this chatbot is to foster better communication between public administrations and migrants. It implements pictograms and mediated oral communication systems, allowing users to leave video messages for service providers via the platform.

The MyWelcome Agent [8] supports multilingual conversational assistance for migrants, aiding in reception and integration processes, employing knowledge models for personalized assistance. Through Behavior Trees and Semantic Service Computing, the agent coordinates tasks and services efficiently. Multilingual language technology services enable communication in languages like Arabic, Catalan, English and Spanish, supported by advanced language understanding and generation technologies.

The user experience can further be enhanced taking into account emotional aspects. This emotional bond has been investigated in [3], where a personable chatbot named Eike was developed to provide information about living in a German city. Users found the psychological and emotional interaction with chatbots very important, valuing features like adhering to social norms and providing reliable information promptly. MyMigrationBot [4] is a chatbot aimed at migrants which attempts to diagnose the psychosocial capital of the user and make it easier for them to find suitable employment in the country they migrated to. It incorporates a personality test which attempts to score the user in the Big Five personality dimensions.

There are a lot of time-consuming and often difficult issues a migrant has to take care of when first settling in another country and NADINE-bot [6] intends to simplify this process. It supports casual conversation as well as domain-specific FAQ conversation related to asylum seeking, education, healthcare, job seeking and other topics incorporating knowledge from a variety of sources across Europe. NADINE-bot is based on BERT and can handle conversation in different languages by translating the user's query into English or into their target language.

OntBot [1] is an ontology-based chatbot which uses a mapping technique to transform the knowledge stored in ontologies into a relational database and then utilize that knowledge in its chats. The OWL to database mapping is done with the OWL2RDB tool. OntBot works using rules, with inference in three parts: a) a scope specifier that specifies the scope of the question asked, b) a rule matcher which tries to find a matching rule given the input tokens and c) a query processor which executes the query against the database. OntBot has the advantage of being application-agnostic, seamlessly supporting different application domains.

3 Overview of SALLY

SALLY capitalises on the recent advances in Large Language Models (LLMs) and builds an ecosystem of several components to empower the SALLY personal assistant with smart dialogues. Through dynamic user profiling, SALLY is able to adapt its behaviour to TCNs’ characteristics (profile information and conversational context) for more user-friendly and human-like interactions. Domain knowledge is represented using OWL 2 ontologies⁴. User profiles and conversational history, as well as the semantics of guided dialogues to elicit information from the users are captured in the form of RDF Knowledge Graphs, fostering interoperability and interlinking. A sentiment analysis component detects the behaviour of users in order to infuse SALLY with emotion intelligence and increase the user experience. Finally, administrative dashboards and analytics aim to inform the solution owners, e.g. local migration authorities, about specific patterns of interaction in order to further improve the provided services.

SALLY’s technical foundation integrates the aforementioned cutting-edge technologies to deliver a seamless and intelligent user experience. The key concepts that underpin the architecture are presented in the following.

Knowledge Base. SALLY must provide trustworthy information to the end users relevant to the migration domain. To this end, a custom-built crawler meticulously extracts and structures information from the website of the Greek Ministry of Migration and Asylum. The crawling component is periodically running and extracting the hyperlinks of the domain. Once the crawling process has been completed, the discovered URLs are forwarded to the scraping component for extracting the content of the pages. This data is meticulously organized into five distinct knowledge hubs, i.e. contacts, facilities, benefits, procedures and services, ensuring efficient retrieval and processing. Furthermore, vector embeddings are generated for all text within these knowledge hubs, enabling advanced semantic understanding and search capabilities. This semantic layer allows SALLY to grasp the nuances and relationships between words and concepts, leading to more accurate information retrieval and contextually relevant responses.

LLM Agnosticism. LLMs play a crucial role for the development of SALLY, enhancing its capabilities by providing natural language understanding and response generation. LLMs are employed for understanding user input, extracting meaning, recognising entities, and analysing sentiment. In addition, LLMs are used for response generation based on the conversational context producing coherent, contextually relevant replies. An important aspect of SALLY is that it follows an LLM-agnostic design. This flexibility allows for seamless integration of various LLMs like Mistral-8x7B or GPT-4-Turbo, enabling the adaptation of the system based on evolving requirements and technological advancements.

⁴ <https://www.w3.org/TR/owl2-overview/>

Retrieval-Augmented Generation (RAG). Standard LLMs can perform various common tasks, such as analysing sentiment and recognising named entities. These tasks typically do not necessitate additional contextual knowledge. However, for more intricate, knowledge-intensive applications, such as a domain-specific chatbot, it is required to construct a language model-based system that leverages external knowledge sources. This approach enhances factual consistency, increases the reliability of generated responses, and mitigates the issue of “hallucinations”. SALLY’s chatbot functionality is founded upon a Retrieval-Augmented Generation (RAG) model. This approach empowers SALLY to access relevant information from its knowledge base and leverage the power of LLMs to generate informative and contextually appropriate responses. To incorporate external knowledge effectively, SALLY employs a sophisticated mechanism that dynamically integrates information from its knowledge base into the conversation flow. This is achieved by mapping user queries to the relevant knowledge hubs and retrieving the most pertinent information in real time. By doing so, SALLY ensures that the dialogue is not only contextually relevant but also enriched with up-to-date and accurate domain-specific knowledge, thereby enhancing the overall effectiveness and personalisation of the interaction.

Dialogue Management. SALLY includes several noteworthy features that enhance its overall functionality. The Dialogue Manager incorporates a long-term memory function, allowing users to seamlessly resume their conversations at a later time by providing their email address. Additionally, a collection of multi-layered LLMs operates in parallel, responsible for tasks such as language fluency detection, user sentiment analysis, and confusion level assessment. The output of these layered LLMs contributes to the creation of a user profile graph that dynamically tracks the user’s state throughout their interaction with the system. This graph provides valuable insights to the Dialogue Manager, enabling it to fine-tune the flow of the human-bot conversation for a more personalised and effective experience. As such, SALLY a) extracts relevant information from user input, such as topics, entities and sentiment, b) provides tailored responses to users, taking into account users’ preference, history and context, c) handles errors and misunderstandings, recognising ambiguous context and responding accordingly asking for clarifications, and d) understands the intentions of users, proactively initiating dialogues and suggesting conversation topics.

Semantic Spaces. SALLY develops standardised semantic representations by promoting the concept of Semantic Conversational Spaces for advanced knowledge representation towards interoperability and knowledge reuse. The Semantic Spaces support the standardised capturing of chatbot’s dialogue management policies, domain knowledge, such as Personal Identifiable Information (PII) through the use of the Extended Personal Data categories (Fig. 1) for the DPV Data Privacy Vocabulary⁵, and conversational context (e.g. detected topics, text analysis results, dialogue history), aiming at building semantically rich and inter-linked knowledge graphs with generic and domain-pertinent situational context.

⁵ <https://w3c.github.io/cg-reports/dpvcg/CG-FINAL-dpv-owl-pd-20221205/>

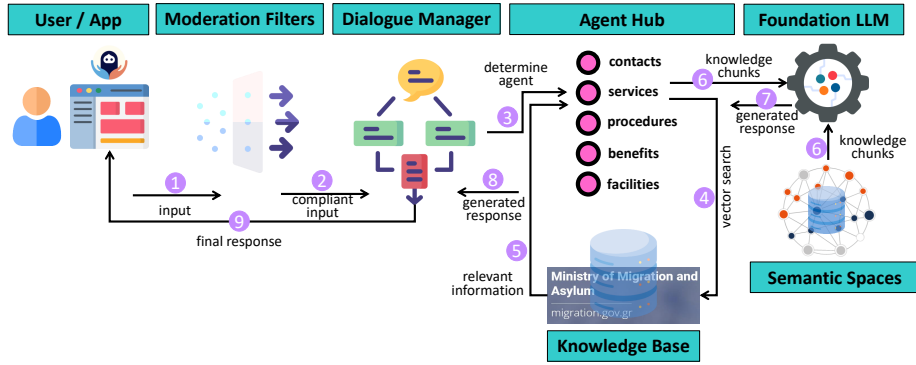


Fig. 2. Conceptual architecture of SALLY.

compliance with SALLY’s ethical guidelines and preserve user privacy. If the message satisfies the moderation criteria, it is routed to the system’s Dialogue Manager. This central control unit plays a pivotal role in determining the most suitable agent among the five available, based on the content of the user’s query. The Dialogue Manager activates the selected agent, which employs a Chain-of-Thought (CoT) [9] prompting technique to access its designated knowledge hub. CoT prompting encourages the LLM to break down complex queries into smaller, more manageable steps, improving the system’s reasoning and problem-solving capabilities. A vector search is then performed to identify the most relevant information within the knowledge base. Relevant content chunks are ranked for prioritization and loaded into the agent’s working memory. The agent then utilizes a foundation LLM (e.g., Mistral or GPT-4), incorporating the available and retrieved knowledge chunks to generate a comprehensive response. The Dialogue Manager receives the LLM-generated response, applies any necessary post-processing refinements, and presents the final output to the user through the conversational interface.

4.1 Example Interactions

We present two targeted interaction scenarios between a user (U) and the system (S) that illustrate basic capabilities of SALLY. In Fig. 3 the dialogue manager, based on user’s input, determines the sub-agent that is most relevant to the request (the procedure agent in this case), which will be the one that drives the rest of the conversation. The Moderation Filters classifies the name and the country of origin that are provided by the user as PII’s (U1) and are ignored. SALLY supports stateful interactions through user accounts, capitalising on account profile information, history, etc. During the conversation, the agent further supports users by suggesting relevant pages from the Knowledge Base (S2). In each turn, the agent detects the sentiment of the user and responds accordingly. For example, in U6 the system detects that the user is concerned or confused,

complementing the response with a reassuring message. When needed, e.g. when the initial question is too generic, the agent proactively suggests topics of discussion or asks follow-up questions in order to elicit additional information. This is the case in the dialogue in Fig. 3 b), where SALLY suggests a more specific topic of discussion (S1), based on the initial (generic) question of the user.

[U1] Hello SALLY, my name is Saleh and I am from Syria.
 [S1] Hi, how can I help you?
 [U2] I would need information about the Asylum Procedure in Greece.
 [S2] Of course. Access to the international protection procedure is unrestricted and free of charge... You can find more information in the following page: <https://migration.gov.gr/en/gas/diadikasia-asylov/i-aitisi-gia-asylo/>. Are you looking for more specific information?
 [U3] Yes, I would like to know where I should submit the application.
 [S3] The International Protection application is submitted in person. If you have not been registered, you can book an appointment...
 [U4] What details should I provide for the registration?
 [S4] Full registration of your application includes your identity details, your country of origin,
 [U5] Should I submit the application in Greek?
 [S5] When submitting your application, there will be an interpreter available to assist you in a language that you understand...
 [U6] I am a little bit concerned, as I am not sure about how my personal information is going to be used.
 [S6] You do not need to worry. All information you provide within your application is confidential....
 [U7] Thank you very much for the information.
 [S7] You are welcome.

(a)

[U1] Good morning SALLY. I am looking for information about Residence Permits. Can you help me?
 [S1] Good morning. You can find information about Residence Permits in the following link <https://migration.gov.gr/en/gas/aitovntes-kai-dikaioychoi/adeies-diamonis/>. I can also give you information about the Residence Permit categories for Third Country Citizens.
 [U2] Yes, sure.
 [S2] There are seven Residence Permits categories:
 o Residence Permit for Work and Professional Purposes
 o Residence Permit for Humanitarian, Exceptional and Other Reasons
 o Residence Permit for Studies, Voluntary Work, Research and Vocational Training
 o Residence Permit for Victims of Trafficking in Human Beings and Illegal Immigration
 o Residence Permit for Family Reunification
 o Long-Stay Residence Permit
 o Special Certificate of Legal Residence
 You can find more information in the following link <https://migration.gov.gr/migration-policy/metanasteusi-stin-ellada/katigories-adeion-diamonis-politon-triton-choron-dikaiologitika%E2%80%8B/>.
 [U3] Thank you very much. Bye.
 [S3] Bye.

(b)

Fig. 3. Example interactions.

5 Conclusions

This paper presented the key concepts and technologies that underpin the ongoing development of SALLY, an intelligent chatbot aiming at empowering and supporting migrants by providing them with useful information, resources, and support services. Compared to existing approaches, SALLY's main focus is to act as a compassionate and informative communication tool, infusing interactions

with intelligent dialogue features, such as clarifications and proactive questions, beyond a simple QA system.

As already described, this is a preliminary work and the framework is still under development. We are currently exploring the performance of both open-source and proprietary models for understanding and generating the Greek language, aiming to enhance SALLY’s linguistic capabilities and ensure its effectiveness across diverse linguistic contexts. For example, taking into account that the majority of open-source models (except for Aya [7]) are not fluent in Greek, we are actively investigating the effectiveness and efficiency of using a dynamic form of translation in real time. This approach could potentially bridge the linguistic gap between open-source and proprietary LLMs in non-English languages, making it a critical area of research for enhancing the accessibility and usability of SALLY for Greek-speaking users.

In addition, several use cases are under development following a co-design approach with domain experts and relevant stakeholders to ensure that the solution meet the evolving needs of migrant populations in ethical and responsible ways. Evaluation trials are planned to assess the agent in terms of reliability and satisfaction.

Acknowledgements 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).

We would also like to thank Enchatted P.C. (<https://enchatted.com/>) that has generously shared its technical expertise and insights on chatbots.

References

1. Al-Zubaide, H., Issa, A.A.: Ontbot: Ontology based chatbot. In: International Symposium on Innovations in Information and Communications Technology. pp. 7–12 (2011). <https://doi.org/10.1109/ISIICT.2011.6149594>
2. Alsharhan, A., Al-Emran, M., Shaalan, K.: Chatbot adoption: A multiperspective systematic review and future research agenda. *IEEE Transactions on Engineering Management* pp. 1–13 (2023). <https://doi.org/10.1109/TEM.2023.3298360>
3. Chen, Z., Lu, Y., Nieminen, M.P., Lucero, A.: Creating a chatbot for and with migrants: Chatbot personality drives co-design activities. In: Proceedings of the 2020 ACM Designing Interactive Systems Conference. p. 219–230. DIS ’20, Association for Computing Machinery, New York, NY, USA (2020)
4. Chlasta, K., Sochaczewski, P., Grabowska, I., Jastrzebowska-Tyczkowska, A.: Mymigrationbot: A cloud-based facebook social chatbot for migrant populations. In: Conference on Computer Science and Information Systems (2022)
5. Garito, M.A., Caforio, A., Pollini, A.: Equity, Diversity and Inclusion in Designing AI-Based Solutions for Migrants’ Integration: REBUILD Project, pp. 177–195. Springer International Publishing, Cham (2022)

6. Lelis, A., Vretos, N., Daras, P.: Nadine-bot: An open domain migrant integration administrative agent. In: 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). pp. 1–6 (2020). <https://doi.org/10.1109/ICMEW46912.2020.9106024>
7. Üstün, A., Aryabumi, V., Yong, Z.X., Ko, W.Y., D’souza, D., Onilude, G., Bhandari, N., Singh, S., Ooi, H.L., Kayid, A., et al.: Aya model: An instruction finetuned open-access multilingual language model. arXiv preprint arXiv:2402.07827 (2024)
8. Wanner, L., Klusch, M., Mavropoulos, A., Jamin, E., Marín Puchades, V., Casamayor, G., Černocký, J., Davey, S., Domínguez, M., Egorova, E., Grivolla, J., Jaramillo Rojas, G.E., Karakostas, A., Ntioudis, D., Pecina, P., Sobko, O., Vrochidis, S., Wertmann, L.: Towards a versatile intelligent conversational agent as personal assistant for migrants. In: Dignum, F., Corchado, J.M., De La Prieta, F. (eds.) *Advances in Practical Applications of Agents, Multi-Agent Systems, and Social Good. The PAAMS Collection*. pp. 316–327. Springer International Publishing, Cham (2021)
9. Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q.V., Zhou, D., et al.: Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems* **35**, 24824–24837 (2022)