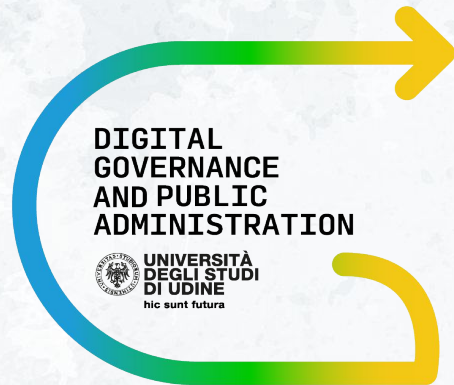


Towards a Conversational-Based Agent for Healthcare

Michael Soprano, Kevin Roitero, Vincenzo Della Mea, and Stefano Mizzaro

CARE: Conversational Agents and Responsible Ethics in Healthcare

Workshop @ *European Federation of Medical Informatics - Special Topic Conference 2023 (EFMI STC 2023)* - Turin, October 27, 2023



Introduction

Conversational agents are software-based systems designed to interact with humans using natural language

The recent surge of **advanced language models** (GPT family) has brought further attention to the usage of such agents

The unexpected backfire in the popularity of chatbot-like agents could be beneficial to less explored fields, such as **public administration** (P.A.)

Introduction

One of the services provided by the P.A. is **allowing citizens to access their Electronic Health Record (EHR)**

The Italian Government published, in 2014, the guidelines for the presentation of **regional project plans for the creation of the EHR** (*Fascicolo Sanitario Elettronico*)

In February 2023, the most recent technical specifications for **interoperability between regional systems** have been published (v2.4.1)

Conversational Agents

They have been well known for some time, having a rather long history, starting from the ELIZA system (1966) to ChatGPT (2023)

In the past, they were based on classical and symbolic AI approaches, while **today they rely on machine learning** (Deep Learning, Natural Language Processing, **Large Language Models**, ...)



Aims

Explore **conversational agent-based eGovernment applications** for the Italian public administration, with a particular focus on **healthcare**

Reducing the digital divide in segments of the population like the elderly and people with disabilities

Usage of **Large Language Models (LLMs)**, **development of prototypes** and **evaluation** of their **effectiveness**

In Sum...

- **eGovernment** applications
- **Healthcare** domain
- **Conversational agents**
- **EHR-related** features
- **Reducing digital divide**
- Italy → **several regional systems**
 - We focus on a single system
 - The case of **Friuli-Venezia Giulia** region

Where To Start?

The EHR access modality provided by the Friuli-Venezia Giulia region is **SeSaMo**, a **web-based portal**

The portal's **EHR-related features** are **complemented by** a set of so-called “quick services”, such as searching for a doctor

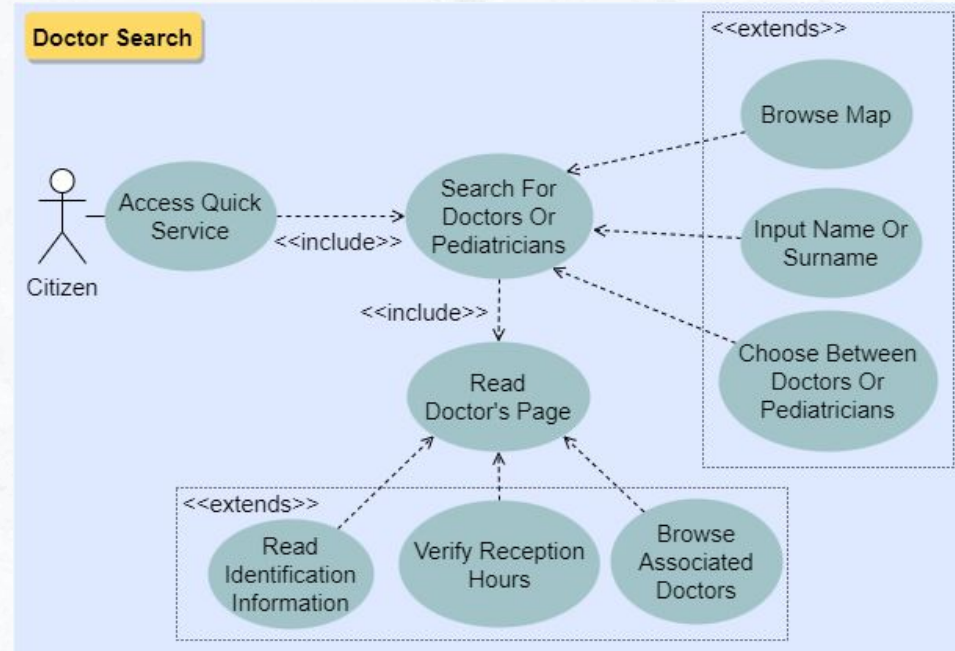
Quick services can be seen as **simple and specific use cases** of the portal



Searching For a Doctor

A user **can search for a doctor** in Friuli-Venezia Giulia by writing a full name or by exploring those available in a given municipality

Then, the user can **read the doctor's details**, verify reception hours, etc.



User Interface

Searching for **doctors and pediatricians** in the municipality of Udine

Filtra elenco per:

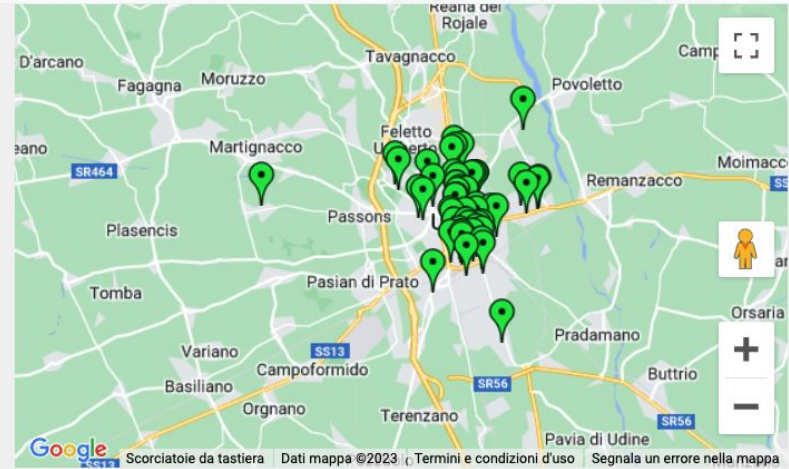
Cognome e/o nome

Digita almeno 3 caratteri

Comune dell' ambulatorio

UDINE

Medici **Pediatri**






Trovati 69 medici in base ai filtri impostati

 **ANTONUTTI VIVIANA (ASU FC – Azienda Sanitaria Universitaria Friuli Centrale)**

 **BARACCHINI PAOLA (ASU FC – Azienda Sanitaria Universitaria Friuli Centrale)**

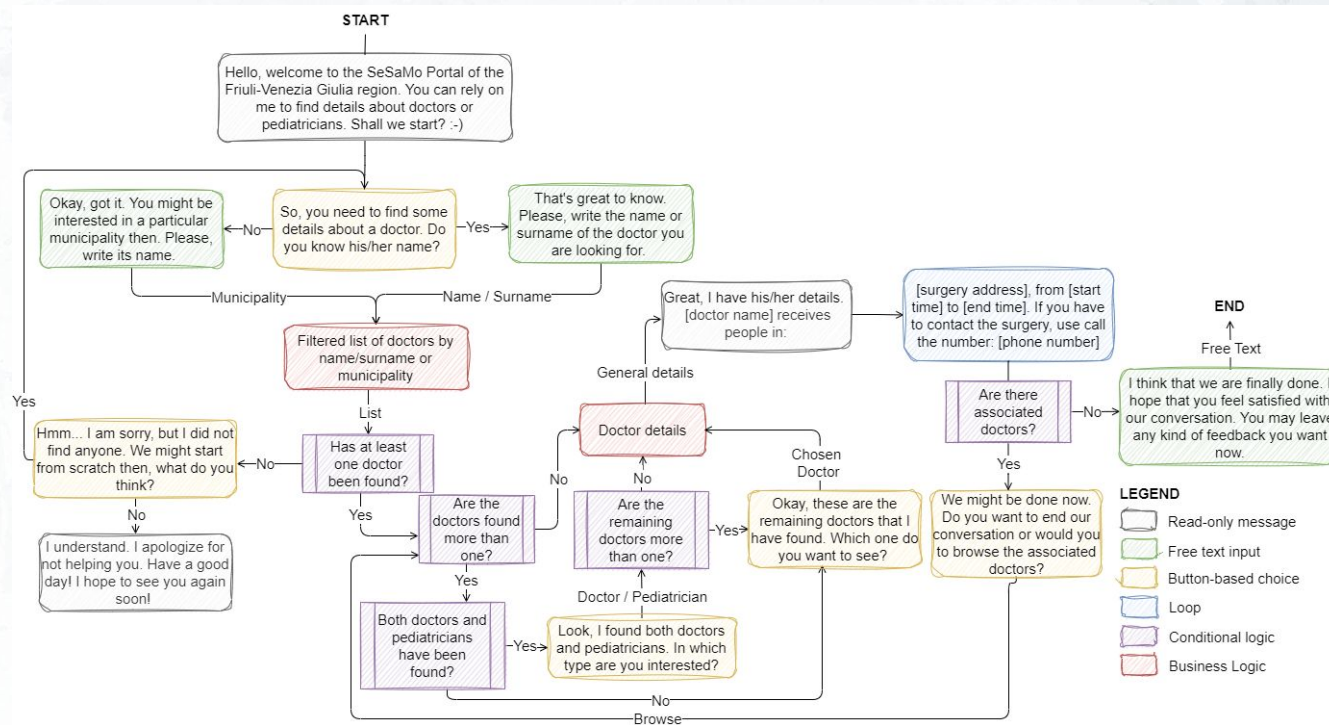
Case Study

We consider the doctor search quick service as the case study to build a prototype:

1. Definition of a **conversation model** 
2. Usage of LLMs-based approaches 
3. **Design** and **implementation** of the prototype 
4. **Evaluation** principles and metrics

The Conversation Model

Graph used to represent the **conversation between the citizen and the conversational agent** while using the quick service







Design Approach

There exist guidelines concerning how to build a “successful” conversational agent:

1. **Identification** of the **key scenarios**
2. Define the **goals that a user can satisfy** (*intents*)
3. **Identification of each intent** (*query*)
4. **Disambiguation** of multiple intents
5. **Carrying out** the conversation

Intent Definition

- Our prototype involves a **single key scenario**
- The aim of the quick service
 - Searching for a doctor
- **Four intents** can be outlined

Intent	Description	Implementation
IN1	Searching for a doctor by full name or municipality	
IN2	Searching for a pediatrician by full name or municipality	
IN3	Exploring the details of the doctor/pediatrician chosen	
IN4	Exploring the list of associated doctors	

Intent Identification (+ Disambiguation)

LLMs can be leveraged **to recognize and understand the user's intent** (in Italian)

The “developer” provides a **list of reference intents that should be recognized**

- Sentences that identify services provided by SeSaMo
- “Information on the municipality doctor”

The **user tells the agent** what s/he needs to achieve

- “I need my GP's schedule”

The **LLM provides a ranking of the most likely intents**, based on semantic similarity

Example

- **User input** → “I need my GP’s schedule”
- **Intent to match** → “Information on the municipality doctor”

User Input

Ho bisogno degli orari del mio medico di base

Intents (one per line)

informazioni sul medico del comune
elenco dei medici associati del comune
ricerca la farmacia per nome
ricerca le strutture per anziani del comune
verifica tempi di attesa e prenotazione per una prestazione
visualizza le informazioni sul farmaco
annullare una prenotazione per una visita
pagare un ticket sanitario
prenotare il vaccino per il COVID-19
vedere dei referti medici sanitari
verifica tempi di attesa in pronto soccorso

Example

Four different LLMs; two of them provide a “wrong” ranking...

vedere dei referti medici sanitari	0.38	verifica tempi di attesa in pronto soccorso	0.45
informazioni sul medico del comune	0.32	vedere dei referti medici sanitari	0.45
verifica tempi di attesa in pronto soccorso	0.31	informazioni sul medico del comune	0.45
ricerca la farmacia per nome	0.27	elenco dei medici associati del comune	0.35
elenco dei medici associati del comune	0.25	verifica tempi di attesa e prenotazione per una prestazione	0.33
verifica tempi di attesa e prenotazione per una prestazione	0.21	ricerca la farmacia per nome	0.3
pagare un ticket sanitario	0.2	pagare un ticket sanitario	0.27
visualizza le informazioni sul farmaco	0.18	visualizza le informazioni sul farmaco	0.27
prenotare il vaccino per il COVID-19	0.16	annullare una prenotazione per una visita	0.16
ricerca le strutture per anziani del comune	0.16	prenotare il vaccino per il COVID-19	0.15
annullare una prenotazione per una visita	0.1	ricerca le strutture per anziani del comune	0.12

Example

... while the remaining two **“guess” the “correct” intent**, the second “in a better way”. We are working on maximizing precision...

informazioni sul medico del comune	0.64	informazioni sul medico del comune	0.68
elenco dei medici associati del comune	0.56	elenco dei medici associati del comune	0.62
vedere dei referti medici sanitari	0.54	vedere dei referti medici sanitari	0.6
verifica tempi di attesa in pronto soccorso	0.52	verifica tempi di attesa in pronto soccorso	0.59
pagare un ticket sanitario	0.45	verifica tempi di attesa e prenotazione per una prestazione	0.49
ricerca la farmacia per nome	0.4	pagare un ticket sanitario	0.43
visualizza le informazioni sul farmaco	0.35	ricerca la farmacia per nome	0.42
verifica tempi di attesa e prenotazione per una prestazione	0.34	prenotare il vaccino per il COVID-19	0.37
prenotare il vaccino per il COVID-19	0.18	visualizza le informazioni sul farmaco	0.35
ricerca le strutture per anziani del comune	0.17	ricerca le strutture per anziani del comune	0.28
annullare una prenotazione per una visita	0.12	annullare una prenotazione per una visita	0.24

How To

Recognizing the user intent consists of three subproblems:

1. **Generate paraphrases for a set of reference intents**, from the SeSaMo portal
 - Finetuning on a dataset of paraphrases in Italian...
2. **Match each paraphrase generated with the correct intent**
 - Generate embeddings for the reference + paraphrased intent and the user intent with a sentence transformer (e.g., SBERT)
 - Find the most similar embedding via k-nearest neighbor search
3. **Refine the sentence model capabilities**
 - contrastive learning, vector-db for model usage in real-time, ...

We are investigating step 3...

Carrying Out The Conversation

The user **accesses the agent and receives a rich card** that shows the **available services**

S/he can either **click the chosen service** or **write a message**. The LLM will **receive the message** and **understand the user intent...**

Then, a **predefined waterfall dialog** to **collect the info** needed to find doctors

The screenshot displays a chatbot interface for 'sesamo'. It starts with a welcome message: 'Ciao User, benvenuto! Sono SeSaBot, il chatbot del portale SeSaMo del Friuli-Venezia Giulia. Ho il compito di aiutarti a fruire i servizi offerti dal portale. Ti va di cominciare? :-)' followed by a 'Adesso' button. A second message says: 'Sono SeSaBot, il chatbot del portale SeSaMo del Friuli-Venezia Giulia. Ho il compito di aiutarti a fruire i servizi offerti dal portale. Ti va di cominciare? :-)' with another 'Adesso' button. Below this is a 'rich card' for 'sesamo' with the tagline 'servizi salute in mobilità'. It lists 'Servizi Disponibili' and provides a search button labeled 'Ricerca medico e/o pediatra'. The interface then transitions to a search form with a 'SI' button. The form asks: 'Per favore, scrivimi nome e/o cognome del medico che stai cercando.' with an 'Adesso' button. Below the form, it states: 'Ho individuato 76 medici, te ne mostro al massimo 3.' Two doctor profiles are shown: 'Chiara Zivic' (Azienda Sanitaria Universitaria Giuliano I... 72389) and 'Chiara Ianna' (Azienda Sanitaria Friuli Occidentale 71142). Each profile has a 'Seleziona' button. The interface concludes with another 'Adesso' button.

Prototype Implementation

- The **prototype** of the agent can be **deployed online easily**, on any website...
 - `https://conversational-agent-application-service.azurewebsites.net/`
 - ... and on different channels, like Telegram
 - `https://telegram.me/WorkPackage5Bot`
- The chosen technology allows for **further integration within other channels**
 - Alexa, Slack, Teams, ...

Web Interface and Telegram Bot

SeSaMo Portal



The screenshot shows the SeSaMo Portal web interface. At the top, the title "SeSaMo Portal" is displayed. Below it is a large blue-bordered box containing the "sesamo" logo in white lowercase letters. Underneath the logo, the text "servizi salute in mobilità" is written in a smaller blue font. Below this, the heading "Servizi del portale SeSaMo" is followed by a paragraph: "Il portale offre diversi servizi. Qui sotto ti presento l'elenco di tutti quelli che sono in grado di fornire allo stato attuale delle cose." A button labeled "Ricerca medico e/o pediatra" is visible. At the bottom, there is a text input field with the placeholder "Digita il tuo messaggio" and a send button. A timestamp "48 minuti fa" is shown at the bottom left.



The screenshot shows the SeSaMo Telegram Bot interface. The chat header displays the bot's name "SeSaBoT" and "bot" next to a green profile picture with a white 'S'. The chat content includes a message from the bot: "Vorrei cominciare presentandoti una rassegna dei servizi che sono in grado di fornirti." followed by a timestamp "10:54". Below this is a card identical to the one in the web interface, showing the "sesamo" logo and "servizi salute in mobilità" text. Another message from the bot follows: "Il portale offre diversi servizi. Qui sotto ti presento l'elenco di tutti quelli che sono in grado di fornire allo stato attuale delle cose." with a timestamp "10:54". A third message says: "Puoi scegliere il servizio che intendi utilizzare toccando uno dei pulsanti" with a timestamp "10:54". At the bottom, a button labeled "Ricerca medico e/o pediatra" is visible. The chat input field at the very bottom contains the placeholder "Scrivi un messaggio...".

Evaluation

Irvine et al. **propose four metrics** as a proxy to measure the engagement level of conversational agents...

- **Average length** (of the conversation): mean number of user queries per conversation session
- **User satisfaction**: fraction of ratings by users equal to a given value
- **Retry rate**: fraction of system responses that the user requested to regenerate at least once
- **Retention rate**: fraction of users that engage with the agents on the x-th day after their first conversation

... but we need **metrics to estimate their effectiveness in reducing digital divide...**

Conclusions

- **eGovernment application** based on **conversational agents**
- **Healthcare** domain, **EHR**-related
- Focus on **Italian language**
- **Conversation model** for a simple use case
 - not trivial, when conversational...
- **LLM-based approach** for user intent detection
- Implementation of a **working prototype** in progress
 - Chatbot-like software, easy to deploy

Open Challenges

- Should the **conversational agent** be **based solely on LLMs**? Could the **conversation follow predefined paths**, sometimes?

According to Zachary Tellman:

“Wherever possible, business logic should be described by code rather than training data. This keeps our system's behavior principled, predictable, and easy to change.”

Do you agree with him?

Open Challenges

- Should we **deploy the agent** on **each channel** or provide a **single user interface**?
- Which **metrics** should we use to **measure an agent's effectiveness in reducing the digital divide**?
- How can we ensure that the **conversational agent adheres to ethical principles**, such as transparency and privacy?
 - User authentication is required for accessing EHRs...
- Can a **voice-based interaction** be more fruitful?

Acknowledgments

This work is being carried out within the
Interdepartmental Project

*Digital Governance and Public
Administration*

Strategic Plan of the **University of Udine**
(Italy), 2022-2025

Thanks!

Any question?