

BIG DATA IN PUBLIC & SOCIAL SERVICES

Julia Lan Bui Xuan 882385

Simone Farallo 889719

Michele Salvaterra 891109

Corso di laurea magistrale - Data Science

Prof. Fabio Mercurio

Prof. Andrea Seveso

CONTENUTI

- 01** INTRODUZIONE
- 02** PANORAMICA SUL DOMINIO
- 03** PROBLEMI E LIMITAZIONI
- 04** SVILUPPO DEL PROGETTO
- 05** RISULTATI & VALUTAZIONE
- 06** CONCLUSIONI
- 07** FUTURE IMPLEMENTAZIONI

INTRODUZIONE

L'obiettivo è di sviluppare un prototipo per il settore sanitario centrato sull'ottimizzazione e l'automazione dell'accesso alle informazioni riguardanti le analisi mediche prescritte dai medici e non, al fine di fornire supporto ai cittadini.

Il chatbot utilizza il framework LangChain, alimentato dal modello GPT3-Turbo, per fornire spiegazioni personalizzate e dare priorità a strumenti specifici per questo progetto.

CHATBOT

l'idea prevedeva la creazione di un chatbot per il supporto ai pazienti in ambito sanitario, capace di fornire informazioni sulle malattie, sulle cure, sui farmaci e sulle prestazioni ambulatoriali.

IDEA

È prevista la creazione e l'integrazione con il chatbot della tecnologia di riconoscimento ottico dei caratteri (OCR) per estrarre i codici delle prescrizioni e individuare le cliniche adatte.

AUTOMATIC OCR SEARCH

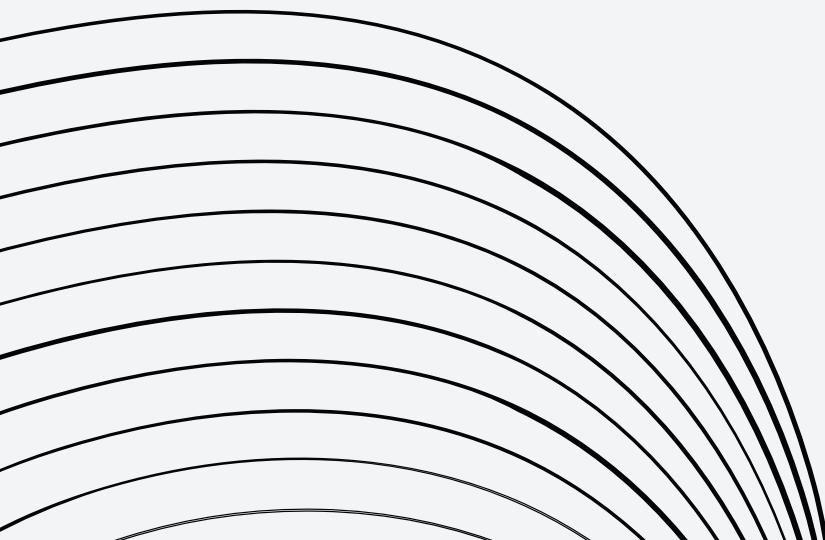
PANORAMICA SUL DOMINIO

Dominio chiuso: l'interazione con il chatbot riguarda solo argomenti rilevanti per la sfera sanitaria.

Sanità medica generale.

Prestazioni ambulatoriali : servizi terapeutici e riabilitativi, nonché procedure di diagnostica strumentale e di laboratorio, gestiti da medici specialisti che operano all'interno di ambulatori e laboratori riconosciute dal Servizio Sanitario Italiano (SSN).

Strutture ambulatoriali regione Lombardia.



PROBLEMI



Privacy e protezione dei dati, tenere presente le leggi sulla protezione dei dati personali e richiedere il consenso esplicito dell'utente.

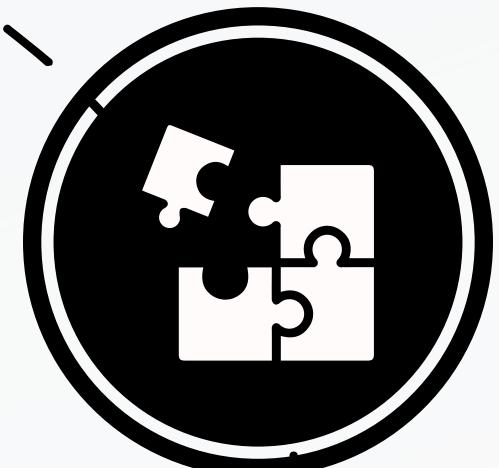
Fornire una dichiarazione di non responsabilità e ricordare agli utenti di confermare le risposte con un medico qualificato.



Trovare il modo di avere informazioni complete e uniformi nel settore sanitario, cercare fonti autorevoli a livello nazionale per supportare il chatbot.



Considerare le limitazioni del modello OpenAI che è probabilistico e generativo, per cui può dare risposte non del tutto accurate o inventate.



LIMITAZIONI



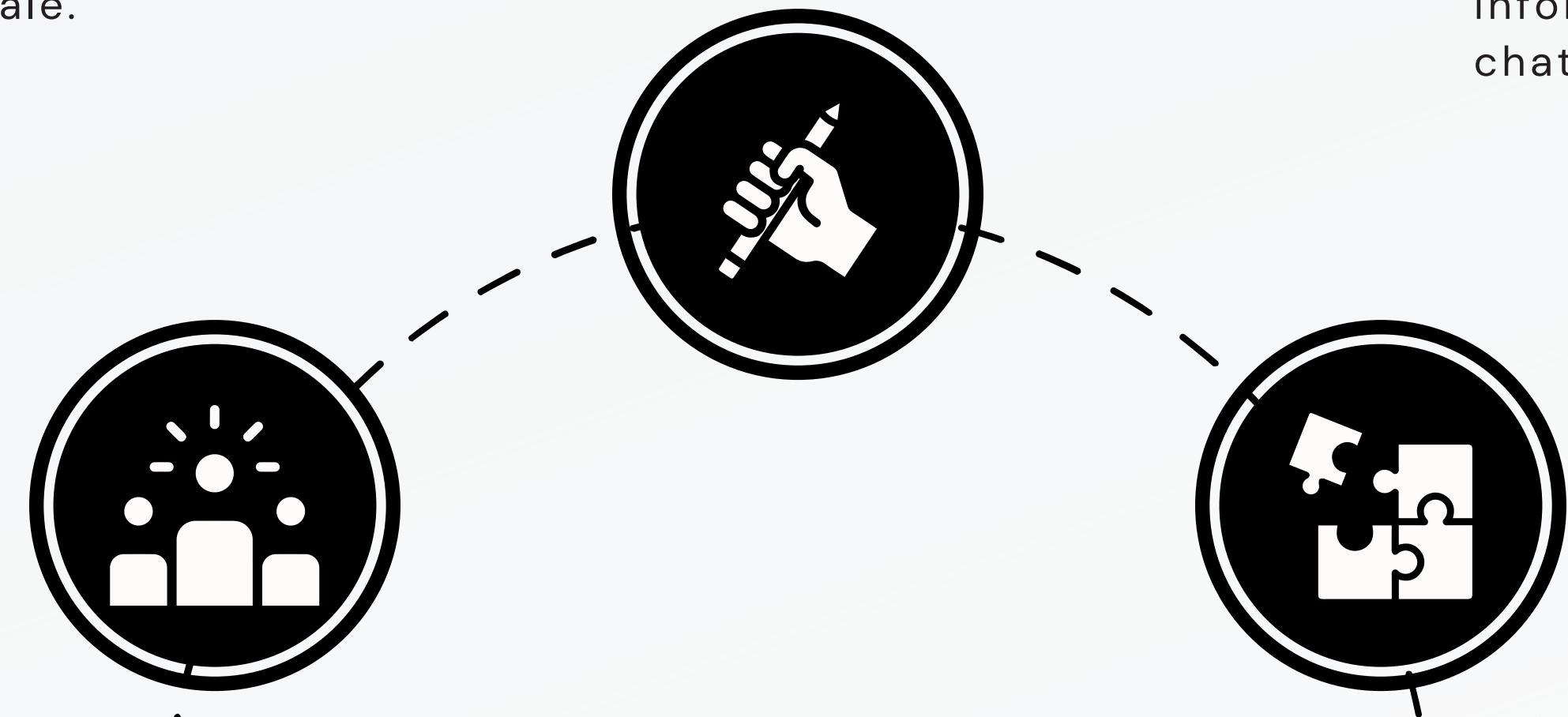
La precisione delle risposte del chatbot può essere imprecisa, nonostante l'addestramento su vasti dati e inoltre il chatbot non sostituisce un parere medico professionale.



L'OCR può presentare sfide legate alla qualità delle immagini, alla calligrafia o ai layout complessi, influenzando la sua precisione



Limitazioni dati disponibili: la frammentazione e la limitata disponibilità dei dati possono influire sulla completezza e accuratezza delle informazioni fornite dal chatbot.



SVILUPPO DEL PROGETTO

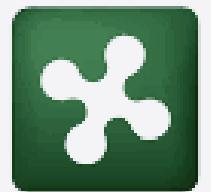


PREPROCESSING & EDA



QUATTRO DATASET:

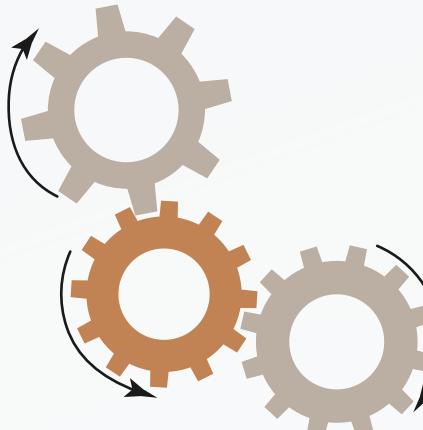
Prestazioni ambulatoriali
Strutture accreditate
Transcodifica noemnclature
Nomenclatura e prezziario nazionale



Regione
Lombardia

OPERAZIONI FATTE:

Rimozione e controllo valori nulli
Controllo dei duplicati
Formattazione codici
Integrazione datasets
Controllo qualità dati



LANGCHAIN

CHATBOT



Modello: GPT-3.5 Turbo.

Agente: zero-shot react description.

Temperatura: 0

Interazioni: limite massimo di 5 iterazioni per la ricerca di risposte.

Memoria: memoria della cronologia della chat attiva solo per la sessione corrente.

Tools: definiti da nome, funzione e descrizione.

LANGCHAIN TOOLS

WIKIPEDIA

Motore di ricerca. Da usare per rispondere a domande relative alla sanità in modo prioritario.

GOOGLE SEARCH

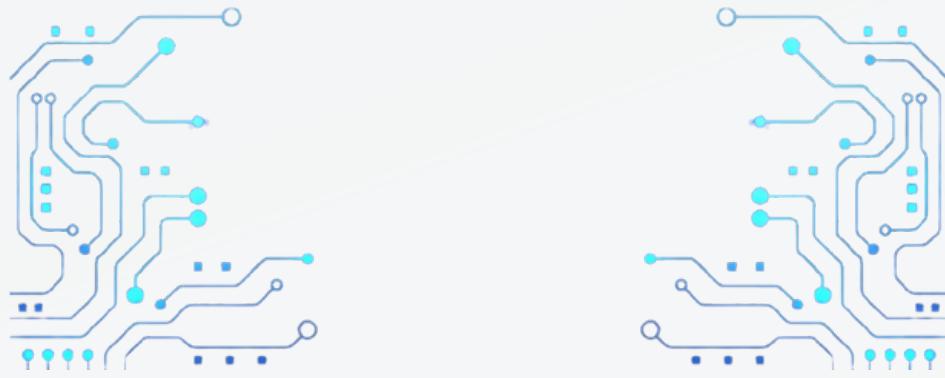
Motore di ricerca. Da usare per rispondere a domande relative alla sanità. Usare specialmente per eventi recenti.

NO ANSWER

Funzione personalizzata: "**Non mi è consentito rispondere a questo tipo di domande perché non sono legate all'ambito sanitario.**
Descrizione = **Usare per rispondere a domande non relative alla sanità.**

PROMPT ENGINEERING

ZERO SHOT PROMPTING



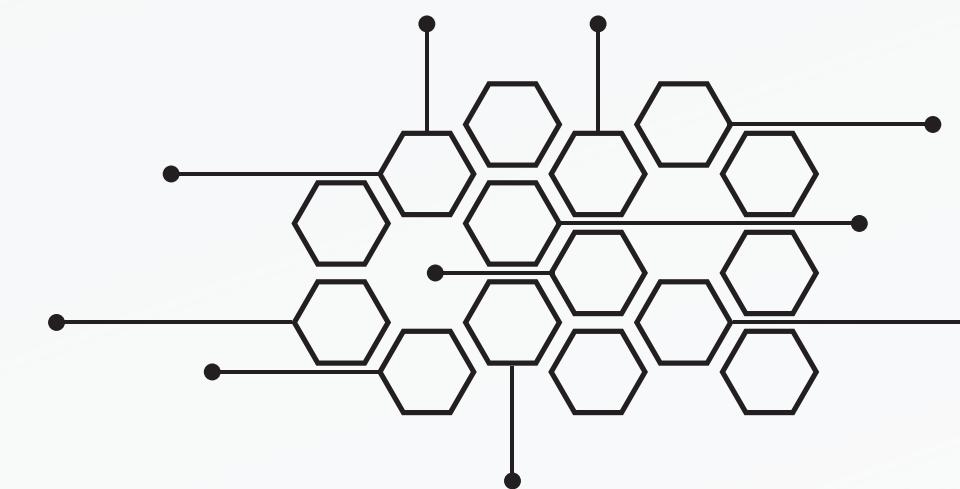
Il modello può generare una risposta coerente senza aver mai visto esempi specifici relativi alla richiesta specifica.

Istruzione: Usa massimo 2 - 3 frasi per rispondere a domande relative alla sanità.

Contesto: Sei un agente AI per fornire informazioni generali all'utente legati all'ambito sanitario. Ricorda che non puoi effettuare prenotazioni o azioni simili.

Domanda: {domanda}

TEMPLATE



CONVERSAZIONE CHATBOT

Che cos'è l'emocromo e perché devo fare le analisi di sangue dell'emocromo?"

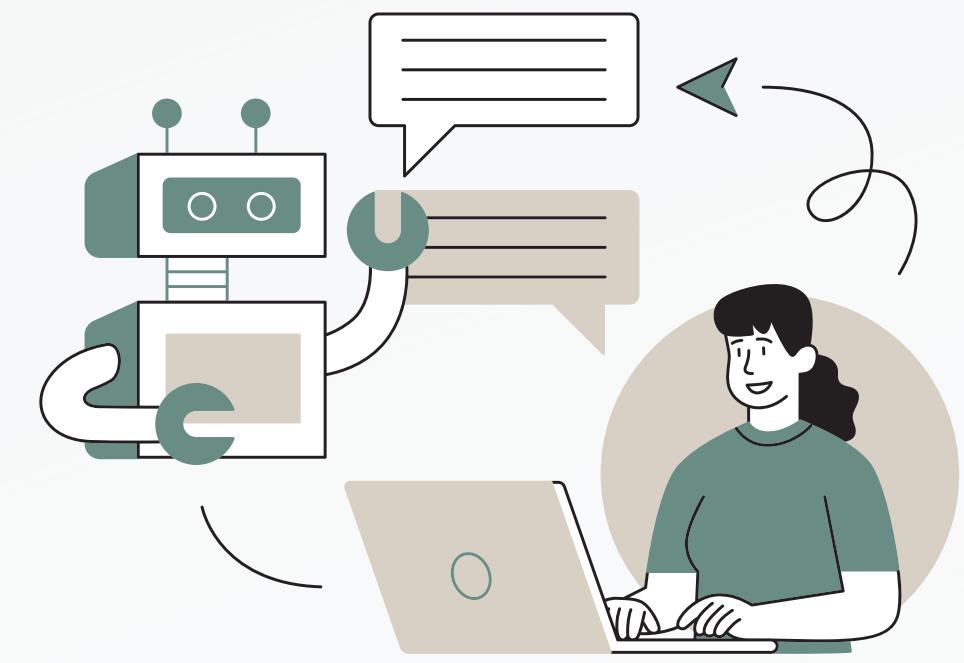
L'emocromo è un esame di laboratorio sul sangue che determina la quantità degli elementi corpuscolati (leucociti o globuli bianchi, eritrociti o globuli rossi, e trombociti o piastrine), i livelli dell'ematocrito (HCT), e dell'emoglobina (Hb), nonché altri parametri del sangue. È importante farlo perché può aiutare a diagnosticare malattie come la leucemia mielomonocitica giovanile

Quali sono i valori normali di emocromo per una donna sana?

I valori normali di emocromo per una donna sana sono: emoglobina (Hb) 12-16 g/dl, eritrociti/ μ l (RBC) 4,2-5,4 x 10⁶, volume corpuscolare medio (MCV) 81-99 fl, emoglobina corpuscolare media (MCH) 26-32 pg.

Consigliami la pizzeria migliore di Milano.

Non posso rispondere a questa domanda perché non è correlata alla sanità



EXTRA TOOLS

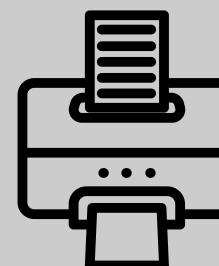
OCR (Optical Character Recognition) su pdf e png.

Fuzzy Matcher: *distanza di levenshtein.*

Automatic OCR Search

Strumento che prende in input una ricetta e restituisce in output la lista delle strutture disponibili.

Le strutture possono essere filtrate per Città, Tipologia,etc



Search medical facilities

Strumento che permette di cercare le strutture inserendo il nome della prestazione



AUTOMATIC OCR SEARCH

Input:

SERVIZIO SANITARIO NAZIONALE		RICETTA ELETTRONICA-PROMEMORIA PER L'ASSISTITO	
 REGIONE PUGLIA			
COGNOME E NOME/INIZIALI DELL'ASSISTITO: [REDACTED]			
INDIRIZZO [REDACTED] CAP: 70056 CITTÀ: MOLFETTA		PROV: BA [REDACTED]	
ESENZIONE:NON ESENTE		SIGLA PROVINCIA:BA CODICE ASL:114	
TIPOLOGIA PRESCRIZIONE(S,H):		DISPOSIZIONI REGIONALI: ALTRO: PRIORITA' PRESCRIZIONE (U,B,D,P):programmabile	
PRESCRIZIONE		QTA	NOTA
90482 (71801) - ANTICORPI ANTI CITOPLASMA DEI NEUTROFILI (ANCA)		1	---
90524 (72901) - ANTICORPI ANTI NUCLEO (ANA)		1	---
90521 (72751) - ANTICORPI ANTI MITOCONDRI (AMA)		1	---
90335 (68251) - MIOGLOBINA [S]		1	---
90155 (63751) - CREATINCHINASI ISOENZIMA MB (CK-MB)		1	---
90154 (63701) - CREATINCHINASI (CPK O CK)		1	---
QUESITO DIAGNOSTICO: <i>accertamenti</i>			
N.CONFEZIONI/PRESTAZIONI:6 TIPO RICETTA:Assist.SSN DATA:26/04/2023 CODICE FISCALE DEL MEDICO:MRLNLG59M06F2840			
CODICE AUTENTICAZIONE:260420231123125190006417000052 COGNOME E NOME DEL MEDICO:MUROLO NICOLA GIACOMO MARIO			
Rilasciato ai sensi dell'art.11, comma 16 del DL 31 mag 2010, n.78 e dell'art.1, comma 4 del DM 2 nov 2011			

1. Ricetta medica

Output:

Codice prestazione ambulatoriale	Codice branca regionale	Branca regionale	Tipo di prestazione	Nome struttura	Comune struttura	Provincia struttura	Indirizzo	Descrizione ASL	Codice struttura	Tipo struttura	Telefono	Fax	Struttura privata
ANTICORPI ANTI CITOPLASMA DEI NEUTROFILI (ANCA)	90482	IMMUNOEMATOLOGIA-TRASFUSIONALE	LABORATORIO	LABORATORIO BIOANALITICO DELTA SRL	BERGAMO	BG	VIA CLARA MAFFEI 14	ATS DI BERGAMO	1252.0	Servizio di Medicina di Laboratorio	035221565	035235848	Si
ANTICORPI ANTI CITOPLASMA DEI NEUTROFILI (ANCA)	90482	IMMUNOEMATOLOGIA-TRASFUSIONALE	LABORATORIO	CENTRO BIOMEDICO BERGAMASCO SRL	BERGAMO	BG	ROTONDA DEI MILLE 3/A	ATS DI BERGAMO	1359.0	Servizio di Medicina di Laboratorio	035222332	035236267	Si
ANTICORPI ANTI NUCLEO (ANA)	90524	IMMUNOEMATOLOGIA-TRASFUSIONALE	LABORATORIO	LABORATORIO BIOANALITICO DELTA SRL	BERGAMO	BG	VIA CLARA MAFFEI 14	ATS DI BERGAMO	1252.0	Servizio di Medicina di Laboratorio	035221565	035235848	Si
ANTICORPI ANTI NUCLEO (ANA)	90524	IMMUNOEMATOLOGIA-TRASFUSIONALE	LABORATORIO	CENTRO BIOMEDICO BERGAMASCO SRL	BERGAMO	BG	ROTONDA DEI MILLE 3/A	ATS DI BERGAMO	1359.0	Servizio di Medicina di Laboratorio	035222332	035236267	Si
ANTICORPI ANTI MITOCONDRI (AMA)	90521	IMMUNOEMATOLOGIA-TRASFUSIONALE	LABORATORIO	LABORATORIO BIOANALITICO DELTA SRL	BERGAMO	BG	VIA CLARA MAFFEI 14	ATS DI BERGAMO	1252.0	Servizio di Medicina di Laboratorio	035221565	035235848	Si
ANTICORPI ANTI MITOCONDRI (AMA)	90521	IMMUNOEMATOLOGIA-TRASFUSIONALE	LABORATORIO	CENTRO BIOMEDICO BERGAMASCO SRL	BERGAMO	BG	ROTONDA DEI MILLE 3/A	ATS DI BERGAMO	1359.0	Servizio di Medicina di Laboratorio	035222332	035236267	Si
MIOGLOBINA [S/U]	90335	CHIMICA CLINICA	LABORATORIO	LABORATORIO BIOANALITICO DELTA SRL	BERGAMO	BG	VIA CLARA MAFFEI 14	ATS DI BERGAMO	1252.0	Servizio di Medicina di Laboratorio	035221565	035235848	Si
MIOGLOBINA [S/U]	90335	CHIMICA CLINICA	LABORATORIO	CENTRO BIOMEDICO BERGAMASCO SRL	BERGAMO	BG	ROTONDA DEI MILLE 3/A	ATS DI BERGAMO	1359.0	Servizio di Medicina di Laboratorio	035222332	035236267	Si
MIOGLOBINA [S/U]	90335	CHIMICA CLINICA	LABORATORIO	HUMANITAS GAVAZZENI	BERGAMO	BG	VIA M. GAVAZZENI 21	ATS DI BERGAMO	30143.0	Struttura di ricovero e cura	0354204111	0354204919	Si
CREATINCHINASI ISOENZIMA MB (CK-MB)	90155	CHIMICA CLINICA	LABORATORIO	CENTRO BIOMEDICO BERGAMASCO SRL	BERGAMO	BG	ROTONDA DEI MILLE 3/A	ATS DI BERGAMO	1359.0	Servizio di Medicina di Laboratorio	035222332	035236267	Si
CREATINCHINASI ISOENZIMA MB (CK-MB)	90155	CHIMICA CLINICA	LABORATORIO	HUMANITAS GAVAZZENI	BERGAMO	BG	VIA M. GAVAZZENI 21	ATS DI BERGAMO	30143.0	Struttura di ricovero e cura	0354204111	0354204919	Si
CREATINCHINASI ISOENZIMA MB (CK-MB)	90155	CHIMICA CLINICA	LABORATORIO	LABORATORIO BIOANALITICO DELTA SRL	BERGAMO	BG	VIA CLARA MAFFEI 14	ATS DI BERGAMO	1252.0	Servizio di Medicina di Laboratorio	035221565	035235848	Si
CREATINCHINASI (CPK o CK)	90154	CHIMICA CLINICA	LABORATORIO	HUMANITAS GAVAZZENI	BERGAMO	BG	VIA M. GAVAZZENI 21	ATS DI BERGAMO	30143.0	Struttura di ricovero e cura	0354204111	0354204919	Si
CREATINCHINASI (CPK o CK)	90154	CHIMICA CLINICA	LABORATORIO	LABORATORIO BIOANALITICO DELTA SRL	BERGAMO	BG	VIA CLARA MAFFEI 14	ATS DI BERGAMO	1252.0	Servizio di Medicina di Laboratorio	035221565	035235848	Si
CREATINCHINASI (CPK o CK)	90154	CHIMICA CLINICA	LABORATORIO	CENTRO BIOMEDICO BERGAMASCO SRL	BERGAMO	BG	ROTONDA DEI MILLE 3/A	ATS DI BERGAMO	1359.0	Servizio di Medicina di Laboratorio	035222332	035236267	Si

2. Strutture private ambulatoriali nella città di Bergamo

INTERFACCIA

Chatbot Automatic OCR Search Search medical facilities

Chatbot

Ciao! Sono il tuo assistente sanitario virtuale. Sono qui per fornirti informazioni generali in ambito sanitario. Tuttavia, è importante sottolineare che non posso sostituire il parere di un medico esperto. Farò del mio meglio per chiarire i tuoi dubbi e offrirti informazioni affidabili.

input
Scrivi qui la tua domanda

Clear Submit

output

Flag

3. Chatbot

Chatbot Automatic OCR Search Search medical facilities

Automatic OCR search tool

Per accedere all'elenco degli ambulatori disponibili in Lombardia, è necessario inserire la propria ricetta medica, specificare la città (in maiuscolo) e indicare se si tratta di una struttura privata (Sì) o meno (No).

Inserisci file pdf o png
Drop File Here
- OR -
Click to Upload

output
1 2 3

Flag

Comune

Struttura privata

Clear Submit

4. Automatic OCR search tool

Chatbot Automatic OCR Search Search medical facilities

Search medical facilities

Inserisci il nome delle prestazione ambulatoriale per ottenere tutte le strutture disponibili per la regione Lombardia.

words
Scrivi il nome dell'analisi

output
1 2 3

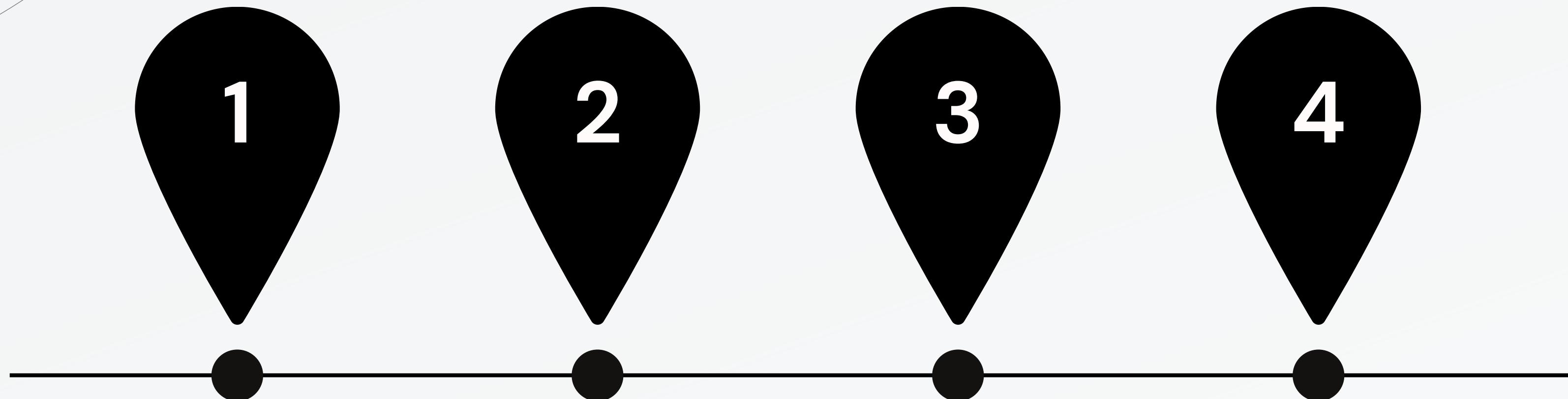
Clear Submit

Flag

5. Search medical facilities



VALUTAZIONE EURISTICA



CHIAREZZA E COMPRENSIBILITÀ

Violazione di secondo
livello, essendo che si
tratta di un chatbot in
campo sanitario le
spiegazioni fornite hanno
un linguaggio tecnico

COERENZA E UNIFORMITÀ

Euristica rispettata

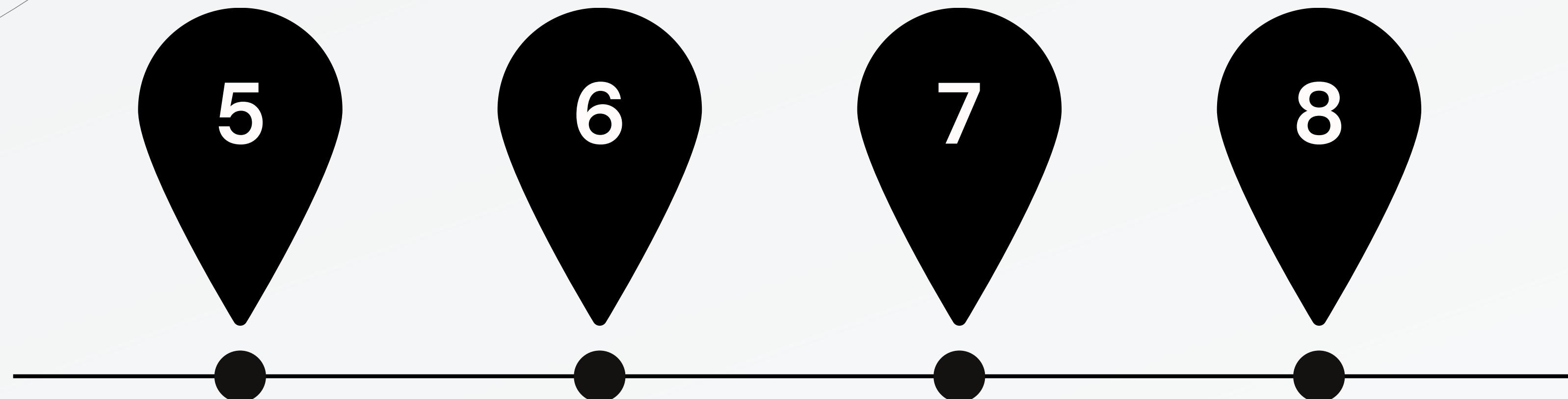
FEEDBACK E CONFERMA

Euristica rispettata, il
sistema nell'elaborare la
risposta presenta una
barra di caricamento che
fa intendere che il
sistema sta elaborando
una risposta.

FACILITÀ DI NAVIGAZIONE

Euristica rispettata

VALUTAZIONE EURISTICA



REATTIVITÀ E TEMPESTIVITÀ

Violazione di secondo livello, le tempistiche di risposta del chatbot variano dai 20 ai 40 secondi, il che vuol dire che viola questa euristica

ADATTABILITÀ E FLESSIBILITÀ

Euristica rispettata

ERRORI E RECUPERO

Violazione di secondo livello, infatti, quando viene caricato un documento nel formato errato non viene emesso alcun avviso sull'errore e non vengono fornite suggerimenti all'utente per risolverlo.

SODDISFAZIONE DELL'UTENTE

Violazione di secondo livello, il chatbot ha una memoria il che vuol dire che può conversare però essendo progettato in un ambito sanitario le risposte che da sono di carattere tecnico.

CONCLUSIONI

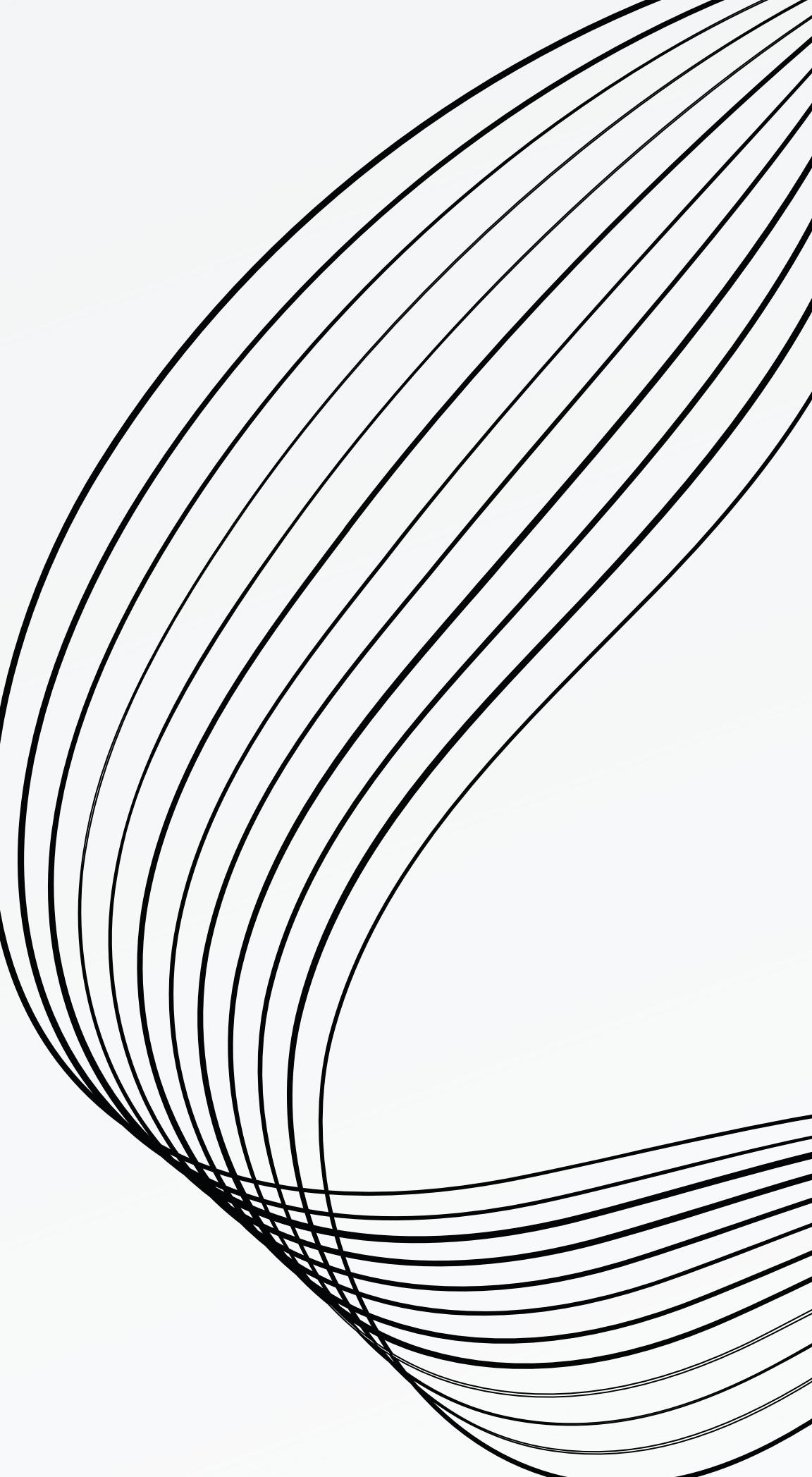
Abbiamo sviluppato un prototipo per supportare il sistema sanitario italiano, tramite OCR e LangChain.

Obiettivo: comprensione delle analisi mediche e ricerca di informazioni sulle cliniche.

Modello GPT-3.5 Turbo: limitazioni nella generazione di risposte accurate.

Dominio sanitario: molto complesso, difficile acquisizione di informazioni affidabili.

Nonostante ciò, il progetto ha gettato le basi per futuri progressi nell'assistenza sanitaria digitale.



FUTURE IMPLEMENTAZIONI



- Integrare funzioni extra e creare strumenti ad hoc.
- Costruire una base di conoscenza valida, accurata e verificata dagli esperti.



- Estendere il dominio a tutto il paese.
- Provare diverse tecniche di prompt per definire il dominio in modo più dettagliato.



- Sperimentare diversi modelli e tecniche, inclusa la creazione di embedding ad hoc
- Sviluppare un chatbot più dinamico, rendendo la conversazione più piacevole e fluida.

**GRAZIE PER
L'ATTENZIONE**

