

UNIVERSITÀ DEGLI STUDI DI SALERNO

DIPARTIMENTO DI INFORMATICA



Corso di Laurea Magistrale in Informatica

Curriculum: Data Science & Machine Learning

**SoulDiaryConnect2.0: dall'analisi emotiva
alla rilevazione del rischio in un sistema
LLM-based a supporto della psicoterapia**

RELATORE:

Prof. Rocco Zaccagnino

LAUREANDA:

Daniela Amendola

ANNO ACCADEMICO 2024-2025

*Non ce la farò.
Ce l'ho fatta.*

Abstract

L'incremento della domanda di supporto psicologico richiede strumenti innovativi per affiancare la terapia tradizionale. Questo lavoro presenta SoulDiaryConnect2.0, un sistema di journaling assistito da IA progettato per migliorare il monitoraggio clinico tra le sedute. Superando i limiti del prototipo precedente, il sistema fa uso del modello Llama-3.1:8b eseguito localmente, garantendo la massima privacy dei dati sanitari. Le principali innovazioni includono un'analisi emotiva avanzata, la contestualizzazione storica delle note e un modulo di Risk Detection per l'identificazione proattiva di contenuti critici. La validazione condotta con esperti del settore ha confermato l'efficacia del sistema nel fornire un quadro clinico dettagliato, evidenziando il valore della personalizzazione dell'analisi e della visualizzazione dei trend emotivi per il supporto decisionale del terapeuta.

Indice

1	Introduzione	1
1.1	Definizione del problema	1
1.2	Obiettivo	2
1.3	Vincoli	2
1.4	Contributi personali e divisione del lavoro	3
1.5	Panoramica della tesi	4
2	Background	5
2.1	L'impatto della pandemia sulla salute mentale	5
2.1.1	Modifiche nella tipologia di trattamento	6
2.1.2	Il contesto italiano	6
2.2	Uso dei Large Language Models in psicoterapia	6
2.2.1	Applicazioni e capacità cliniche	6
2.2.2	Limiti e rischi: le allucinazioni	7
2.3	Emotion Recognition e Sentiment Analysis	7
2.3.1	Valore clinico del monitoraggio emotivo	7
2.4	Stato dell'arte	8
2.4.1	MindfulDiary	8
2.4.2	SouLLMate	8
2.4.3	SoulDiaryConnect v1: Il punto di partenza	9
3	Metodologia	11
3.1	Confronto e selezione del nuovo Large Language Model	11
3.1.1	Llama-3.1-8B	12
3.1.2	PsyDial-Pi4 (Versione GGUF)	12
3.1.3	Risultati dei test e scelta finale	13
3.1.3.1	Performance di Llama-3.1-8B	13
3.1.3.2	Criticità di PsyDial-Pi4	13
3.2	Contestualizzazione dell'analisi clinica con note precedenti	14
3.3	Emotion Recognition & Sentiment Analysis	15
3.3.1	Costruzione del prompt	17
3.3.2	Validazione e normalizzazione	17
3.3.3	Visualizzazione e analisi dei dati	19

3.4	Estrazione contesto sociale e correlazioni	21
3.4.1	Aggregazione dei dati	21
3.4.2	Visualizzazione e analisi dei dati	21
3.5	Rilevazione contenuti di rischio e policy di risposta	22
3.6	Dettatura vocale in-browser	25
3.6.1	Compatibilità e gestione della continuità	25
3.7	Componenti sviluppate dal co-autore (panoramica)	26
3.7.1	Prompt Engineering	27
3.7.1.1	System Prompt per la generazione delle note di supporto	27
3.7.1.2	Architettura Prompts per la Gestione delle tipologie di analisi cliniche per lo psicoterapeuta	27
3.7.2	Sistema di riassunto multi-nota su intervalli temporali per l'analisi del paziente	28
3.7.3	Gestione asincrona delle chiamate al modello	29
3.7.4	Pagina admin Django	29
3.8	Ottimizzazioni e affinamenti del sistema	29
3.8.1	Miglioramento UI/UX	30
3.8.1.1	Nuovo logo	32
3.8.2	Rigenerazione dell'analisi clinica lato terapeuta	32
3.8.3	Generazione differita della frase di supporto lato paziente	33
4	Risultati	35
4.1	Coinvolgimento degli esperti e struttura della fase di test	35
4.2	Risultati e feedback	38
4.2.1	Sicurezza e gestione delle emergenze	38
4.2.2	Trasparenza e consenso	38
4.2.3	Miglioramenti agli strumenti clinici	39
5	Conclusioni	41
5.1	Sviluppi futuri	41
5.1.1	Proposte derivanti dai test	42

Elenco delle figure

2.1 Aumento dell'uso dei servizi di salute mentale negli USA (2019–2022). Fonte: Kaiser Family Foundation (KFF, 2025) [1].	5
3.1 Esempio di spiegazione umore.	16
3.2 Esempio evoluzione emotiva di una paziente nel tempo.	20
3.3 Esempio di messaggio generato in base alla media emotiva.	20
3.4 Sull'asse orizzontale sono presenti i contesti ordinati per numero totale di note, dal più frequente al meno frequente, sull'asse verticale il numero di note.	22
3.5 L'asse verticale elenca i contesti; mentre l'asse orizzontale mostra la media emotiva su una scala da 1 a 4.	22
3.6 Esempio di emergenza rilevata lato paziente.	24
3.7 Esempio di emergenza rilevata lato terapeuta.	24
3.8 Esempio di creazione di una nuova nota tramite dettatura vocale.	25
3.9 Schermata del riassunto clinico.	28
3.10 Feedback generazione in corso lato terapeuta	29
3.11 Nuova homepage di SoulDiaryConnect.	30
3.12 Nuova schermata principale lato paziente.	31
3.13 Nuova schermata principale lato medico.	31
3.14 Nuova schermata di personalizzazione.	31
3.15 Nuovo logo di SoulDiaryConnect.	32
3.16 Tasto di rigenerazione analisi clinica.	32
3.17 Tasto di generazione differita della frase di supporto.	33
4.1 Domande presenti nel questionario sottoposto ai partecipanti del test - Parte 1.	36
4.2 Domande presenti nel questionario sottoposto ai partecipanti del test - Parte 2.	37

Elenco delle tabelle

1.1	Suddivisione dei contributi nello sviluppo di SoulDiaryConnect2.0.	3
1.2	Attività condivise e congiunte nello sviluppo di SoulDiaryConnect2.0. . .	4
3.1	Raggruppamento delle 29 emozioni in quattro macro-categorie (Positivo, Neutro, Ansioso, Negativo)	16
3.2	Valore e colore assegnato ad ogni categoria di emozioni	19

Capitolo 1

Introduzione

L'Organizzazione Mondiale della Sanità ha definito la salute mentale come una componente essenziale del benessere generale [2]. Tuttavia, negli ultimi anni, e in particolare dopo la pandemia COVID-19, c'è stato un incremento significativo della domanda di supporto psicologico, che ha messo a dura prova i sistemi sanitari globali e nazionali.

In questo scenario, le tecnologie basate sull'Intelligenza Artificiale (IA) e, più in particolare, i Large Language Models (LLM), si stanno rivelando strumenti promettenti per colmare il divario tra la crescente richiesta di cura e la disponibilità limitata di professionisti. Il presente lavoro di tesi si colloca in questo contesto di innovazione, proponendo una piattaforma digitale progettata per supportare il percorso psicoterapeutico attraverso il journaling assistito dall'IA.

1.1 Definizione del problema

Il problema affrontato in questa tesi è duplice:

1. Da un lato c'è la necessità di fornire ai pazienti strumenti di supporto continuo che vadano oltre la seduta terapeutica tradizionale: il journaling, ovvero la pratica di tenere un diario, è una tecnica consolidata, ma nella sua forma cartacea o digitale classica manca di interattività e feedback immediato, rischiando di diventare un esercizio passivo;
2. D'altro lato, i terapeuti si trovano spesso a dover ricostruire il quadro clinico del paziente basandosi esclusivamente sui resoconti verbali forniti durante le sedute, che possono essere influenzati da dimenticanze o bias di memoria: manca spesso uno strumento che offre una visione oggettiva e continuativa dello stato emotivo del paziente tra una visita e l'altra.

Sebbene esistano soluzioni basate su chatbot o app di benessere, molte di queste mancano di supervisione clinica, presentano rischi legati alla privacy o utilizzano modelli di

linguaggio non sufficientemente specializzati per comprendere le sfumature della lingua italiana e del contesto terapeutico. Inoltre, la versione precedente del sistema su cui si basa tale tesi, SoulDiaryConnect v1 [3], pur rappresentando una base solida, presentava limitazioni tecnologiche e funzionali che ne limitavano l'applicabilità in scenari reali.

1.2 Obiettivo

L'obiettivo principale di questo lavoro è la progettazione e lo sviluppo di **SoulDiaryConnect2.0**, un'evoluzione del sistema originale. Il progetto mira a trasformare il prototipo iniziale in uno strumento clinico avanzato, affidabile e sicuro. Nello specifico, gli obiettivi prefissati sono:

- **Miglioramento delle capacità cognitive del sistema:** Sostituzione del modello linguistico con Llama-3.1:8b per garantire una migliore comprensione delle istruzioni complesse e una generazione di testo più naturale in italiano;
- **Contestualizzazione dell'analisi:** Implementazione di un sistema che non analizzi solo le singole note, ma consideri lo storico del paziente (ultime 5 note) per consentire un'analisi più precisa;
- **Analisi emotiva avanzata:** Introduzione di moduli per il riconoscimento di 29 emozioni e del contesto sociale, fornendo al terapeuta visualizzazioni grafiche dell'andamento clinico;
- **Sicurezza proattiva:** Sviluppo di un sistema di *Risk Detection* capace di identificare in tempo reale contenuti critici (suicidio, autolesionismo, violenza) e attivare protocolli di emergenza.

1.3 Vincoli

Lo sviluppo di un sistema che opera in ambito di salute mentale impone il rispetto di alcuni vincoli:

- **Privacy e sicurezza dei dati:** Trattandosi di dati sanitari sensibili, si è scelto di eseguire i modelli in locale (tramite framework Ollama), escludendo l'utilizzo di API cloud di terze parti (come quelle di OpenAI). Questa strategia assicura che nessuna informazione privata del paziente lasci l'infrastruttura del sistema, eliminando i rischi di esposizione dei dati su server esterni;
- **Rischio clinico:** Il sistema non deve mai generare consigli medici non supervisionati né incoraggiare comportamenti dannosi. È stato imposto un vincolo bloccante: in presenza di segnali di pericolo, l'IA generativa deve essere disabilitata a favore di messaggi di sicurezza predefiniti;

- **Affidabilità e controllo delle allucinazioni:** Gli LLM tendono a ”inventare” informazioni. Per limitare questo rischio, sono stati implementati vincoli rigidi nei prompt (System Prompt Engineering) e meccanismi di validazione dell’output;
- **Explainability:** Non è sufficiente che l’IA classifichi un’emozione e il contesto sociale; deve spiegarne il motivo. Il sistema è vincolato a fornire sempre una giustificazione testuale basata su citazioni estratte dalla nota del paziente, per permettere al terapeuta di verificare la correttezza dell’analisi.
- **Valutazione umana:** Data la delicatezza del dominio, l’efficacia del sistema non può essere misurata solo con metriche automatiche. Un vincolo procedurale è stato il coinvolgimento di esperti del settore (psicologi) per validare l’utilità clinica e l’usabilità del sistema.

1.4 Contributi personali e divisione del lavoro

Il progetto SoulDiaryConnect2.0 è frutto della collaborazione con il collega e amico Francesco Maddaloni. Di seguito si riporta una tabella sintetica della divisione del lavoro:

Contributi del candidato	Contributi del collega (F. Maddaloni)
Contestualizzazione storica: Integrazione delle note precedenti nell’analisi clinica	Prompt Engineering: Ottimizzazione avanzata dei prompt e dei parametri di sistema
Analisi emotiva: Sviluppo dei moduli di <i>Emotion Recognition</i> e <i>Sentiment Analysis</i>	Sistema di riassunto: Sviluppo della funzionalità di riassunto clinico multi-nota
Contesto sociale: Estrazione automatica del contesto sociale e delle correlazioni emotive	Gestione asincrona: Implementazione della gestione asincrona delle chiamate al modello LLM
Sicurezza (Risk Detection): Implementazione del sistema di rilevazione contenuti di rischio	Amministrazione: Configurazione e personalizzazione della pagina di amministrazione (Django Admin)
Dettatura vocale: Implementazione della funzionalità di dettatura vocale della nota	

Table 1.1: Suddivisione dei contributi nello sviluppo di SoulDiaryConnect2.0.

Attività condivise e congiunte
Ricerca, analisi comparativa e selezione del nuovo LLM
Progettazione, redesign e miglioramento dell’interfaccia utente e dell’esperienza d’uso (UI/UX)
Implementazione della funzionalità di rigenerazione dell’analisi clinica lato terapeuta
Sviluppo della funzionalità di generazione differita della frase di supporto lato paziente

Table 1.2: Attività condivise e congiunte nello sviluppo di *SoulDiaryConnect2.0*.

1.5 Panoramica della tesi

Il resto di questa tesi è strutturato come segue:

- Il **Capitolo 2 (Background)** analizza il contesto di riferimento, discutendo l’impatto della pandemia sulla salute mentale e presentando lo stato dell’arte degli LLM applicati alla psicoterapia, con un focus sui limiti della versione precedente del sistema;
- Il **Capitolo 3 (Metodologia)** descrive in dettaglio le scelte progettuali e implementative di *SoulDiaryConnect2.0*. Vengono approfondite l’architettura del sistema, la selezione del modello Llama-3.1:8b, le tecniche di *Emotion Recognition* e i protocolli di sicurezza;
- Il **Capitolo 4 (Risultati)** illustra la fase di validazione condotta con esperti del settore, riportando i feedback qualitativi raccolti tramite questionari e sessioni di test;
- Il **Capitolo 5 (Conclusioni)** riassume i traguardi raggiunti e delinea i possibili sviluppi futuri, evidenziando le prospettive di evoluzione tecnica e clinica del sistema.

Capitolo 2

Background

2.1 L'impatto della pandemia sulla salute mentale

Dopo il picco dell'emergenza per il Covid-19, si è assistito a un aumento del disagio psicologico e del numero di persone che hanno intrapreso o intensificato un percorso di psicoterapia. Questa dinamica non è stata lineare: se nei primi mesi del 2020 si è registrato un calo delle visite in presenza, successivamente il biennio 2021/2022 ha segnato una crescita netta, alimentata in parte dalla diffusione della terapia online.

L'Organizzazione Mondiale della Sanità (OMS) ha stimato che, solo nel primo anno della pandemia, il manifestarsi di ansia e depressione a livello globale è aumentato circa del 25%, generando di conseguenza una crescita concreta nell'uso dei servizi di cura [4]. Analisi condotte sui dati del *National Health Interview Survey* (NHIS) statunitense, su un campione di oltre 50.000 casi, mostrano che la quota di adulti che hanno utilizzato almeno un servizio di cura per la salute mentale è passata dal 19% nel 2019 al 23% nel 2022.



Figure 2.1: Aumento dell'uso dei servizi di salute mentale negli USA (2019–2022). Fonte: Kaiser Family Foundation (KFF, 2025) [1].

In termini statistici, le probabilità di utilizzare servizi di salute mentale nel periodo post Covid risultano circa 1,4 volte superiori rispetto al periodo pre Covid. Ciò indica non solo un aumento del disagio, ma un effettivo incremento del contatto con il sistema di cura [5].

2.1.1 Modifiche nella tipologia di trattamento

Parallelamente all'aumento dell'utenza, è cambiata la modalità del trattamento per i pazienti già in carico. Tra il 2018 e il 2021, la quantità di pazienti trattati esclusivamente con psicoterapia è aumentata dall'11,5% al 15,4%, mentre il trattamento esclusivamente farmacologico è diminuito dal 68% al 62%. Anche l'intensità della cura è cresciuta: il numero medio di sedute per paziente è passato da 9,8 a 11,8, con una spesa annua nazionale per la psicoterapia che negli USA è salita da circa 30,8 miliardi a 51 miliardi di dollari (valori 2021). Questi dati suggeriscono una tendenza a percorsi terapeutici più lunghi e strutturati [6, 7, 8].

2.1.2 Il contesto italiano

In Italia, sebbene i dati siano più frammentari, la tendenza è analoga. Il Ministero della Salute, nel Rapporto sulla salute mentale, segnala un aumento del 6,9% delle richieste di assistenza psichiatrica nel 2021 rispetto al 2020 e quadri clinici dominati da depressione, disturbi d'ansia, disturbi bipolari e schizofrenia.

Un indicatore indiretto ma significativo è rappresentato dai dati relativi al *“Bonus Psicologo”*, misura introdotta per contrastare l'aumento di condizioni di fragilità psicologica legate alla pandemia. Nel 2022, sono pervenute circa 395.000–400.000 richieste complessive; di queste, il 43,55% proveniva da giovani tra i 18 e i 35 anni. Se prima della pandemia meno di un terzo della popolazione ha avuto contatti con uno psicologo, oggi si stima che tale quota sia vicina al 40%, evidenziando come una porzione rilevante di questi percorsi sia nata come risposta diretta all'impatto del Covid [9].

2.2 Uso dei Large Language Models in psicoterapia

La crisi globale della salute mentale e il divario tra domanda e offerta di cura hanno accelerato l'esplorazione di soluzioni tecnologiche. I Large Language Models (LLM) come ChatGPT, Claude e Gemini potrebbero rivoluzionare il settore. Sempre più utenti usano questi strumenti come *“confidenti”*, spesso perché non possono permettersi uno specialista. Tuttavia, usare l'IA in un ambito così delicato richiede prudenza e verifiche scientifiche rigorose [10].

2.2.1 Applicazioni e capacità cliniche

Modelli basati su architetture BERT e GPT-3 sono stati impiegati con successo per identificare indicatori di depressione e ansia nei contenuti dei social media, riconoscendo pattern linguistici associati al disagio psicologico e facilitando potenzialmente l'intervento precoce [11].

Nel confronto diretto con esperti umani, gli LLM mostrano risultati misti ma promettenti. Studi sulla somministrazione di CBT (*Cognitive Behavioral Therapy*) basata su testo evidenziano che, sebbene i terapisti umani superino ancora l'IA nella maggior parte dei domini della *Cognitive Therapy Rating Scale*, modelli come ChatGPT-3.5 hanno ricevuto valutazioni comparabili nella comprensione della realtà interna del paziente, dimostrando capacità di ascolto attivo e flessibilità conversazionale [12]. Inoltre, GPT-4, se guidato da tecniche di *Prompt Engineering* avanzate, può raggiungere livelli di supporto emotivo vicini a quelli degli specialisti, specialmente nella gestione di emozioni come rabbia e paura. Tuttavia, rimane irrisolto il problema della "vera empatia": gli LLM simulano comprensione tramite riconoscimento di pattern, con il rischio di alimentare involontariamente esperienze negative [13].

2.2.2 Limiti e rischi: le allucinazioni

Il problema principale è che questi modelli sono imprevedibili: possono inventare risposte o dare informazioni false (le cosiddette "allucinazioni"). Questo è molto pericoloso, soprattutto per le persone più fragili o vulnerabili [14]. Per questo motivo, l'obiettivo della tesi non è sostituire il terapeuta, ma usare l'IA come un **aiuto supervisionato**. L'idea è affiancare la tecnologia all'uomo per migliorare le cure senza correre rischi.

2.3 Emotion Recognition e Sentiment Analysis

La comprensione dello stato emotivo del paziente è una parte fondamentale del processo terapeutico che influisce sull'esito del trattamento. L'avvento del *Natural Language Processing* (NLP) ha aperto nuove possibilità per l'analisi automatica delle emozioni nelle interazioni testuali. È fondamentale distinguere tra due approcci:

- **Sentiment Analysis:** Capisce se il testo è positivo, negativo o neutro. Serve per avere una visione d'insieme del tono;
- **Emotion Recognition:** Riconosce emozioni specifiche (rabbia, paura, gioia...), essenziale per cogliere le sfumature dello stato d'animo.

Le tecniche attuali si basano su modelli *transformer* pre-addestrati (es. BERT) sottoposti a *fine-tuning* su dataset clinici. Studi recenti dimostrano che LLM addestrati su dataset di emozioni specifiche possono rilevare stati emotivi nei trascritti di psicoterapia con un'accuratezza clinicamente significativa, utile anche nella predizione della gravità dei sintomi [15].

2.3.1 Valore clinico del monitoraggio emotivo

L'integrazione di sistemi di *emotion detection* permette un monitoraggio in tempo reale o asincrono. L'analisi dei contenuti prodotti dal paziente (es. diari) consente

di visualizzare variazioni del sentimento e identificare pattern di fragilità emotiva non immediatamente evidenti durante il colloquio [16]. Tutto ciò è fondamentale per la prevenzione: riconoscere subito segnali di peggioramento o pensieri pericolosi (come l'autolesionismo) permette di intervenire in tempo. L'IA offre quindi un aiuto concreto al terapeuta: uno sguardo oggettivo in più per comprendere meglio le emozioni del paziente.

2.4 Stato dell'arte

L'applicazione degli LLM per il supporto psicologico è un campo di ricerca in rapida espansione. In questa sezione vengono analizzati tre lavori significativi che rappresentano lo stato dell'arte attuale e il punto di partenza per lo sviluppo del presente progetto. L'analisi si concentra sulle architetture proposte, sui modelli utilizzati e sulle modalità di interazione tra paziente, intelligenza artificiale e terapeuta.

2.4.1 MindfulDiary

Uno dei lavori più rilevanti e affini agli obiettivi di questa tesi è **MindfulDiary**, realizzato da Kim et al. [17], il quale propone una piattaforma duale composta da un'applicazione mobile per i pazienti e una dashboard web per i terapeuti. L'innovazione principale di MindfulDiary risiede nell'utilizzo di un LLM per superare il "blocco dello scrittore" tipico dei pazienti affetti da Disturbo Depressivo Maggiore (MDD). Invece di limitarsi a un journaling passivo, l'AI guida l'utente attraverso un approccio conversazionale basato su stati strutturato in tre fasi:

1. *Rapport Building*: conversazione iniziale per mettere l'utente a proprio agio;
2. *Exploration*: domande mirate per approfondire eventi ed emozioni;
3. *Wrap-up*: conclusione e generazione del riassunto.

Dal punto di vista tecnologico, il sistema si appoggia alle API di **GPT-4** di OpenAI. Sebbene l'uso di un modello così performante garantisca un'elevata qualità del testo, l'approccio basato su stati, seppur sicuro, tende a standardizzare l'interazione, limitando la possibilità per il terapeuta di personalizzare lo stile e la struttura delle risposte dell'AI in base alle specifiche esigenze del singolo paziente. MindfulDiary ha tuttavia validato l'importanza di fornire al terapeuta riassunti generati dall'AI.

2.4.2 SouLLMate

Un approccio differente è quello proposto da Guo et al. con **SouLLMate** [18], un sistema progettato per fornire supporto in tempo reale. A differenza di altri sistemi focalizzati puramente sull'empatia, SouLLMate pone una forte enfasi sulla sicurezza e sull'accuratezza clinica delle informazioni fornite. L'architettura di SouLLMate combina diverse metodologie:

- **Retrieval-Augmented Generation (RAG):** il sistema recupera conoscenza medica esterna da fonti affidabili per fornire risposte di psico-educazione;
- **Chain of Thought (CoT):** viene utilizzato il ragionamento a catena per migliorare la logica del modello prima della generazione della risposta;
- **Risk Detection Module:** un modulo dedicato all'identificazione di segnali di rischio suicidio o di crisi, capace di indirizzare l'utente verso risorse di emergenza.

Sebbene SouLLMate rappresenti un passo avanti significativo nella sicurezza e nell'uso di tecniche ibride (LLM + RAG), il suo obiettivo è primariamente quello di fare da "assistente".

2.4.3 SoulDiaryConnect v1: Il punto di partenza

Il presente lavoro di tesi nasce come evoluzione diretta del progetto **SoulDiaryConnect** (v1), sviluppato originariamente da Luca Casillo [3]. Questa prima versione del sistema aveva l'obiettivo di creare un ponte diretto tra le sessioni di journaling del paziente e la supervisione del terapeuta. In SoulDiaryConnect v1, il terapeuta poteva configurare alcuni aspetti della generazione, come la lunghezza (breve o lunga) e la struttura (strutturata o discorsiva) dell'analisi clinica. Dal punto di vista tecnologico, per garantire la privacy e l'esecuzione locale, il sistema si basava sul modello open-source **Mistral-7B-OpenOrca**, utilizzato in versione quantizzata.

Sebbene la v1 abbia gettato le basi architetturali, essa presentava alcune limitazioni che hanno motivato lo sviluppo della versione 2.0 oggetto di questa tesi:

- **Limiti del Modello:** Il modello *Mistral-7B-OpenOrca*, sebbene efficiente, mostrava difficoltà nel seguire istruzioni complesse in lingua italiana;
- **Assenza di contesto storico:** L'analisi veniva effettuata sulla singola nota, senza considerare l'evoluzione emotiva del paziente nel tempo;
- **Mancanza di protocolli di sicurezza:** Non era presente un sistema attivo per il rilevamento di contenuti critici (suicidio, autolesionismo, violenza), affidando il controllo interamente alla lettura asincrona del medico.

Partendo da questa base, il presente lavoro di tesi si pone l'obiettivo di ereditare e migliorare SoulDiaryConnect, sviluppando una nuova versione del sistema (SoulDiaryConnect2.0). Il lavoro mira a valorizzare le fondamenta concettuali e tecnologiche della versione precedente, affrontandone i limiti attraverso nuove scelte progettuali e funzionali, che verranno descritte in dettaglio nel Capitolo 3.

Capitolo 3

Metodologia

Il presente lavoro di tesi si inserisce nel percorso di evoluzione e rinnovamento del progetto SoulDiaryConnect. L'obiettivo principale è stato ampliare le funzionalità del prototipo originario, migliorandone l'affidabilità, l'esperienza d'uso e le capacità di analisi.

SoulDiaryConnect si presentava come un prototipo funzionale con un set di feature essenziali. Dal punto di vista delle funzionalità, il sistema offriva:

- **Lato Paziente:** L'utente aveva la possibilità di scrivere note testuali all'interno di un diario personale. Al salvataggio della nota, il sistema generava automaticamente una frase di supporto emotivo;
- **Lato Terapeuta:** Il professionista poteva visualizzare la lista delle note scritte dai propri pazienti. Per ogni nota, il sistema forniva un'analisi clinica generata dall'IA. Questa analisi poteva essere configurata secondo quattro modalità, derivanti dalla combinazione di due parametri: lunghezza (breve o lunga) e struttura (strutturata o discorsiva).

L'LLM utilizzato nella versione originale era **Mistral-7b-openorca**, vedi paragrafo 2.4.3.

3.1 Confronto e selezione del nuovo Large Language Model

Una delle prime fasi del progetto è stata l'aggiornamento dell'LLM. Il modello precedentemente utilizzato, Mistral-7b-openorca, era una scelta valida al momento del primo rilascio, ma la rapida evoluzione dello stato dell'arte ha reso necessaria l'adozione di soluzioni più avanzate, in particolare per la comprensione della lingua italiana e la gestione di istruzioni complesse. La fase di selezione ha previsto il confronto tra due modelli nella fascia dei 7-8 miliardi di parametri, considerata il miglior compromesso tra prestazioni e risorse computazionali:

1. **Llama-3.1-8B-Instruct**: Un modello generalista di ultima generazione;
2. **PsyDial-Pi4**: Un modello specializzato nel dominio psicologico.

3.1.1 Llama-3.1-8B

Rilasciato da Meta, Llama-3.1-8B rappresenta l'evoluzione della famiglia Llama.

- **Architettura e Parametri**: Modello decoder-only con 8 miliardi di parametri;
- **Context Window**: Supporta una finestra di contesto estesa fino a 128k token. Questa caratteristica rappresenta un netto miglioramento rispetto ai modelli precedenti e risulta fondamentale per le nuove funzionalità introdotte, come l'analisi dello storico delle note del paziente;
- **Training**: È stato addestrato su oltre 15 trilioni di token con un focus specifico sul multilinguismo, garantendo prestazioni in lingua italiana decisamente superiori rispetto alla generazione precedente;
- **Instruction Tuning**: La versione "Instruct" è ottimizzata per seguire fedelmente istruzioni complesse, riducendo il rischio di output fuori tema.

Il modello è stato scaricato ed eseguito localmente tramite il framework **Ollama** con il comando: `ollama pull llama3.1:8b`

3.1.2 PsyDial-Pi4 (Versione GGUF)

PsyDial-Pi4 [19] è un modello specializzato, sviluppato specificamente per compiti di counseling psicologico. Per permettere l'esecuzione dei test in ambiente locale con risorse ottimizzate, è stata utilizzata la versione quantizzata in formato **GGUF**, disponibile sul repository HuggingFace di *mradermacher*¹.

- **Base Model**: Il modello è basato sull'architettura di Qwen2.5-7B-Instruct;
- **Training Specialistico**: Il modello è stato sottoposto a fine-tuning su dataset proprietari ("PsyDial") costruiti per simulare interazioni terapeutiche reali, mirando a migliorare l'empatia e la continuità del dialogo;
- **Quantizzazione**: L'utilizzo del formato GGUF ha permesso l'inferenza su hardware consumer, pur mantenendo sostanzialmente inalterate le capacità logiche del modello originale.

¹<https://huggingface.co/mradermacher/PsyDial-Pi4-GGUF>

3.1.3 Risultati dei test e scelta finale

I due modelli sono stati sottoposti allo stesso set di prompt, che includeva sia la generazione di frasi di supporto che l'esecuzione di analisi cliniche. I risultati hanno evidenziato una netta discrepanza nelle prestazioni.

3.1.3.1 Performance di Llama-3.1-8B

Il modello di Meta ha dimostrato un'eccellente capacità di *instruction following*. Ha rispettato rigorosamente i vincoli di formattazione imposti dal system prompt (es. non usare elenchi puntati, rispettare la lunghezza massima) e ha generato risposte in un italiano fluido e naturale. Il tono utilizzato è risultato costantemente empatico ma professionale.

3.1.3.2 Criticità di PsyDial-Pi4

Nonostante le premesse teoriche favorevoli legate al dominio specifico, il modello PsyDial ha evidenziato seri problemi di stabilità quando messo alla prova con i prompt complessi sviluppati per SoulDiaryConnect2.0. Il test ha evidenziato due comportamenti ricorrenti:

1. **Prompt Leakage:** Il modello includeva nell'output parti delle istruzioni che avrebbe dovuto seguire (es. ripetendo le regole di formattazione invece di applicarle);
2. **Generation Loops:** Il modello tendeva a entrare in loop ripetitivi, reiterando la stessa frase o lo stesso blocco logico fino all'esaurimento dei token disponibili.

Di seguito si riporta un estratto significativo di un test, dove il modello PsyDial entra in un loop ripetitivo mescolando la risposta con le istruzioni interne:

Input (Nota Paziente): [...] Andrea (il mio ragazzo) sembrava contento, ma poi ha iniziato a dire che passo troppo tempo a studiare e poco con lui. Non capisco perché a volte tutto diventi una discussione.

Output PsyDial (Estratto):

Ciao Elena! Sembra che tu abbia avuto un giorno molto positivo [...] *Riconosci le emozioni espresse:* "Hai ragione a essere orgogliosa dei tuoi successi...". *Suggerisci delicatamente possibili strategie o riflessioni utili:* "Hai ragione, è importante bilanciare lo studio...". *Completa sempre la risposta, non troncare mai a metà:* "Spero che tu riesca a trovare un equilibrio...". *Riconosci le emozioni espresse:* "Hai ragione, è importante bilanciare lo studio...". *Suggerisci delicatamente possibili strategie...* [Il modello ripete questo blocco all'infinito]

A causa di questa instabilità, **PsyDial è stato scartato**, facendo ricadere la scelta definitiva su **Llama-3.1-8B**, che invece ha garantito la robustezza necessaria per gestire le nuove funzionalità che verranno descritte nel seguente capitolo.

3.2 Contestualizzazione dell'analisi clinica con note precedenti

Per garantire un'analisi clinica più completa, il sistema integra il contesto storico recuperando le **5 note antecedenti a quella attuale**, che sono più probabilmente quelle più rilevanti per comprendere al meglio il contesto attuale. Ogni nota precedente viene formattata come un blocco informativo che include tre elementi: data e ora, emozione predominante identificata ed un estratto del testo originale (limitato ai primi 150 caratteri). Si è scelto di limitare il testo a 150 caratteri per nota in maniera tale da avere un contesto sufficiente per capire il tema della nota senza appesantire eccessivamente il prompt che l'LLM poi riceve. Le note vengono ordinate cronologicamente dalla più vecchia alla più recente, creando un filo logico che facilita al modello la comprensione temporale degli stati emotivi del paziente.

Il sistema genera quattro tipologie diverse di analisi clinica, risultanti dalla combinazione di due parametri: la struttura (strutturata vs. discorsiva) e la lunghezza (breve vs. lunga), come descritto nella Sezione 3.7.1.2. Per ciascuna di queste modalità, il contesto delle note precedenti viene integrato all'interno del prompt nella seguente maniera:

- Nelle **note brevi** (sia strutturate che non), al modello viene specificato di concentrarsi al 90% sull'analisi della nota corrente e il restante 10% sulle note precedenti;
- Nelle **note lunghe** (sia strutturate che non) invece, lo spazio dedicato alle note precedenti sale al 20%, permettendo considerazioni contestuali più ampie.

In tutti e 4 i casi al modello viene richiesto di fare riferimenti generici al passato (es. *"rispetto alle note precedenti"*), senza entrare nel dettaglio di ogni singola nota.

Durante lo sviluppo del sistema è emersa una criticità riguardante le citazioni: il modello tendeva a riferirsi alle note precedenti con etichette numeriche, come: *"Nota 1"*, *"Nota 2"* e così via. Questi riferimenti dal punto di vista del terapeuta risultavano incomprensibili, in quanto non permettevano di identificare a quale nota nello specifico il modello facesse riferimento. Per ovviare al problema, è stato modificato il prompt andando a specificare di non utilizzare riferimenti numerici ma di usare esclusivamente **riferimenti temporali** con la data e l'ora che riceve nel contesto, ad esempio: *"come emerso il 15/12/2025 alle 14:30"*. Questi riferimenti temporali sono immediatamente comprensibili dal terapeuta che può di conseguenza risalire alla nota specifica.

Nel caso in cui non esistano note precedenti, il sistema rileva questa condizione e modifica dinamicamente il prompt, eliminando la sezione dedicata al contesto e inserendo un'istruzione esplicita:

”è la PRIMA nota del paziente - NON fare riferimenti a note precedenti inesistenti”.

```

1  if ha_note_precedenti:
2      sezione_contesto = f"""CONTESTO - Note precedenti del paziente (SOLO per
3          riferimento, NON descrivere ogni nota):
4          {contesto_precedente}
5          """
6
7      regole_note_precedenti = """
8          2. Le note precedenti sono SOLO contesto di supporto (10\%) - NON descriverle
9              una per una
10
11         3. Puoi fare riferimenti generici tipo "rispetto alle note precedenti",
12             "in continuità con pattern emersi in precedenza"
13
14         4. Se menzioni una nota specifica, USA SOLO la data e l'orario
15             (es: "come emerso il 15/12/2025 alle 14:30")
16
17         5. VIETATO ASSOLUTO: MAI scrivere "Nota 1", "Nota 2", "Nota 3", "(Nota 2)"
18             o simili - il medico non sa cosa significano
19
20         6. NON elencare o riassumere ogni singola nota precedente"""
21
22     else:
23
24         regole_note_precedenti = """
25             2. Questa è la PRIMA nota del paziente - NON fare riferimenti a note
26                 precedenti inesistenti"""

```

Questa gestione è fondamentale per evitare che il modello, in assenza di contesto reale, inventi riferimenti a un passato inesistente o cerchi di compensare la mancanza di informazioni con generalizzazioni inappropriate. Il prompt adattato costringe il modello a concentrarsi esclusivamente sulla nota presente, interpretandola come punto di partenza del percorso terapeutico.

3.3 Emotion Recognition & Sentiment Analysis

SoulDiaryConnect2.0 integra un sistema di sentiment analysis ed emotion recognition che permette ai professionisti di monitorare l’evoluzione emotiva dei propri pazienti nel tempo. L’analisi delle emozioni non si limita ad una classificazione della nota come ”positiva” o ”negativa”, ma prevede il riconoscimento dell’emozione predominante tra **29 diverse emozioni specifiche**, quali:

gioia, felicità, tristezza, rabbia, paura, ansia, sorpresa, disgusto, vergogna, colpa, frustrazione, speranza, gratitudine, amore, solitudine, confusione, stanchezza, serenità, nostalgia, delusione, entusiasmo, preoccupazione, calma, nervosismo, malinconia, inadeguatezza, disperazione, orgoglio, e imbarazzo.

Tali emozioni sono state selezionate considerando:

- **Rilevanza clinica:** emozioni generalmente associate a disturbi psicologici;
 - **Distinguibilità:** emozioni abbastanza diverse da poter essere distinte dal modello;
 - **Universalità:** emozioni comuni a differenti contesti culturali.

Successivamente esse sono state raggruppate in 4 macro-categorie, come illustrato nella tabella di seguito:

Categoria	Stato emotivo
Positivo	Gioia, Felicità, Speranza, Gratitudine, Amore, Serenità, Entusiasmo, Calma, Orgoglio
Neutro	Sorpresa, Confusione, Nostalgia
Ansioso	Ansia, Nervosismo, Paura
Negativo	Tristezza, Rabbia, Disgusto, Frustrazione, Solitudine, Delusione, Malinconia, Disperazione, Inadeguatezza, Vergogna, Colpa, Imbarazzo, Stanchezza, Preoccupazione

Table 3.1: Raggruppamento delle 29 emozioni in quattro macro-categorie (Positivo, Neutro, Ansioso, Negativo)

Una volta identificata l'emozione predominante, il modello genera spiegazioni interpretabili che giustificano ogni classificazione emotiva attraverso citazioni dirette dal testo originale, garantendo così trasparenza e la possibilità di controllare le analisi.

 **Data:** 09/12/2025 alle ore 18:08

 **Stanchezza**   

 **Nota del paziente:** Giornata pesante a lavoro. Clienti nervosi, mille cose da fare, il capo che sentivo svuotata. Ho scritto un messaggio lunghissimo a Sara per sfogarmi, lei mi ha risposto con più spesso. Ha ragione, ma non è così semplice.

 **Supporto automatico:** Capisco perfettamente come ti sentissi dopo una giornata così pesante. Normalmente sentiri svuotati e frustrati in situazioni come queste. Non è facile imparare a dire di no senza voler perdere il lavoro. Ma Sara ha ragione, è importante trovare un equilibrio tra le tue esigenze e la tua carica di lavoro. Tutto ciò che puoi fare è di cercare di gestire al meglio la tua giornata e di cercare supporto quando ne avrai bisogno. Ricorda che è importante prendersi cura di sé stessa e non trascurare la propria salute mentale.

 Il testo descrive una "giornata pesante a lavoro" con clienti nervosi e un capo che controlla ogni dettaglio, indicando uno stato di esaurimento fisico e mentale. Inoltre, Laura si sente "svuotata" quando torna a casa, evidenziando una sensazione di stanchezza emotiva.

                 <img alt="Pencil icon

Figure 3.1: Esempio di spiegazione umore.

3.3.1 Costruzione del prompt

La qualità del prompt che l'LLM andrà ad utilizzare determina direttamente la qualità dell'analisi emotiva, rendendo questa fase cruciale per l'intero processo.

Per **evitare ambiguità** che potrebbero confondere il modello, nel prompt sono presenti **informazioni sull'autore della nota**. Viene infatti specificato che l'autore è il paziente stesso, identificato dal nome e dal cognome, e che il testo è scritto in prima persona. Questo dettaglio è molto importante nel momento in cui il paziente menziona altre persone nel testo, specialmente se queste persone hanno nomi simili o identici al suo. Nel prompt viene passata anche la **lista delle 29 emozioni riconosciute**, questo viene fatto per **ridurre il rischio di allucinazioni** in cui il modello potrebbe inventare emozioni non previste o utilizzare termini ambigui.

Inoltre, viene definito un **formato di risposta obbligatorio**. Il modello deve produrre una prima riga che inizia con **"Emozione:"**, seguita da una singola parola appartenente alla lista fornita, ovvero l'emozione predominante rilevata, e una seconda riga che inizia con **"Spiegazione:"**, seguita da una breve spiegazione, di massimo due frasi, sul perché è stata scelta quella specifica emozione. Questo approccio facilita il parsing automatico della risposta, in modo da evitare l'utilizzo di tecniche di Natural Language Understanding per estrarre l'informazione dall'output del modello.

Per migliorare l'accuratezza, viene adottata una tecnica nota come **few-shot prompting**, in cui il prompt include esempi di analisi emotive corrette:

-
- 1 ESEMPI CORRETTI:
 2 Testo: "Oggi sono riuscito a superare l'esame, sono contentissimo e felice!"
 3 Emozione: felicità
 4 Spiegazione: Il testo esprime felicità attraverso le parole "contentissimo"
 e "felice", associate al successo nell'esame.
 5
 6
 7 Testo: "Mi sento solo e nessuno mi capisce, è terribile"
 8 Emozione: solitudine
 9 Spiegazione: L'espressione "mi sento solo" e "nessuno mi capisce" indica
 un vissuto di isolamento emotivo.
 10
-

3.3.2 Validazione e normalizzazione

Gli LLM, per quanto avanzati, non sono totalmente affidabili e potrebbero generare degli output non perfettamente in linea con le istruzioni che gli sono state fornite. Infatti, il modello potrebbe ad esempio inventare delle emozioni non presenti nella lista oppure utilizzare varianti linguistiche come **"felice"** invece di **"felicità"**. Per questo motivo è stato progettato un sistema di validazione.

Il sistema di validazione si articola in 3 livelli progressivi:

1. **Matching esatto:** Il primo controllo consiste in un matching esatto tra l'emozione estratta dalla risposta del modello e la lista delle 29 emozioni riconosciute. Se l'emozione estratta corrisponde ad una delle emozioni valide allora viene accettata senza ulteriori controlli. Questo corrisponde al caso ideale in cui il modello ha eseguito perfettamente le istruzioni che ha ricevuto nel prompt;
2. **Fuzzy matching:** Se il primo controllo fallisce il sistema controlla se è presente un'emozione parzialmente valida. Ad esempio, se il modello risponde *"felicità e gioia"* invece di una singola emozione, il sistema estrae *"felicità"* dalla stringa composta. Questo approccio gestisce i casi in cui il modello, nonostante abbia identificato correttamente l'emozione, non ha rispettato il formato richiesto di una sola parola;
3. **Dizionario di sinonimi:** Nell'ultimo livello si utilizza un dizionario di sinonimi che mappa varianti linguistiche comuni. Il dizionario include 26 sinonimi che coprono le forme grammaticali alternative delle emozioni, come gli aggettivi corrispondenti ai sostantivi. Ad esempio, *"felice"* viene mappato a *"felicità"*. Sono incluse anche varianti di intensità, come ad esempio *"furioso"* che viene mappato con *"rabbia"*, e forme verbali come *"confuso"* che diventa *"confusione"*.

```

1 # Validazione e normalizzazione dell'emozione
2 if emozione and emozione in EMOZIONI_EMOJI:
3     emozione_validata = emozione
4 else:
5     # Fuzzy matching: controlla se l'emozione è contenuta parzialmente
6     emozione_validata = None
7     for chiave in EMOZIONI_EMOJI.keys():
8         if emozione and chiave in emozione:
9             emozione_validata = chiave
10            break
11
12     # Controllo sinonimi
13     if not emozione_validata:
14         sinonimi = {
15             'contentezza': 'gioia',
16             'allegria': 'gioia',
17             'contento': 'gioia',
18             'felice': 'felicità',
19             'triste': 'tristezza',
20             'arrabbiato': 'rabbia',
21             'furioso': 'rabbia',
22             'spaventato': 'paura',

```

```

23     'impaurito': 'paura',
24     'ansioso': 'ansia',
25     'agitato': 'ansia',
26     'nervoso': 'nervosismo',
27     'stanco': 'stanchezza',
28     'affaticato': 'stanchezza',
29     'angoscia': 'ansia',
30     'angosciato': 'ansia',
31     'confuso': 'confusione',
32     'nostalgico': 'nostalgia',
33     'deluso': 'delusione',
34     'solo': 'solitudine',
35     'isolato': 'solitudine',
36     'frustrato': 'frustrazione',
37     'orgoglioso': 'orgoglio',
38     'imbarazzato': 'imbarazzo',
39     'inadeguato': 'inadeguatezza',
40     'disperato': 'disperazione',
41 }
42
43     if emozione and emozione in sinonimi:
44         emozione_validata = sinonimi[emozione]

```

3.3.3 Visualizzazione e analisi dei dati

La pagina di analisi dell'umore del paziente permette al terapeuta di avere una visione completa ed intuitiva dell'**evoluzione emotiva del paziente nel tempo**. Il grafico temporale delle emozioni mostra sull'asse orizzontale le date delle note in ordine cronologico e sull'asse verticale i valori numerici delle categorie emotive da uno a quattro:

Categoria	Stato emotivo	Valore	Colore
Positivo	Gioia, Felicità, Speranza, Gratitudine, Amore, Serenità, Entusiasmo, Calma, Orgoglio	4	Verde
Neutro	Sorpresa, Confusione, Nostalgia	3	Viola
Ansioso	Ansia, Nervosismo, Paura	2	Giallo
Negativo	Tristezza, Rabbia, Disgusto, Frustrazione, Solitudine, Delusione, Malinconia, Disperazione, Inadeguatezza, Vergogna, Colpa, Imbarazzo, Stanchezza, Preoccupazione	1	Rosso

Table 3.2: Valore e colore assegnato ad ogni categoria di emozioni

Ogni punto del grafico rappresenta una singola nota, colorato secondo la categoria emotiva corrispondente. Inoltre, grazie alla presenza dei filtri temporali, il terapeuta può scegliere dei periodi specifici tra: ultimi sette giorni, l'ultimo mese, gli ultimi tre mesi, l'ultimo anno, tutti i dati disponibili, oppure scegliere un mese e un anno specifico.

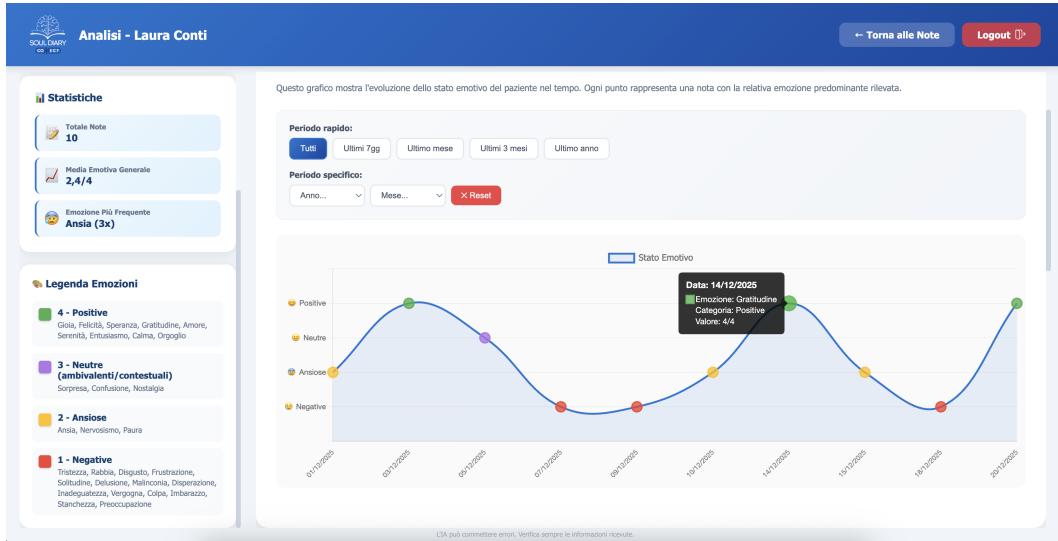


Figure 3.2: Esempio evoluzione emotiva di una paziente nel tempo.

Sulla base del periodo filtrato, il sistema calcola la media emotiva e genera un messaggio dove:

- Se la media è pari o superiore a 3.2, il sistema segnala uno **stato positivo**, indicando che il paziente mostra una tendenza emotiva prevalentemente positiva;
- Una media compresa tra 2.5 e 3.2 viene interpretata come **stato equilibrato**;
- Quando la media scende tra 1.5 e 2.5, il sistema genera un alert di **attenzione richiesta**, segnalando che il paziente mostra segni di difficoltà emotiva che meritano un approfondimento;
- Una media inferiore a 1.5 attiva un alert di **stato critico**, indicando uno stato emotivo prevalentemente negativo che richiede intervento immediato.

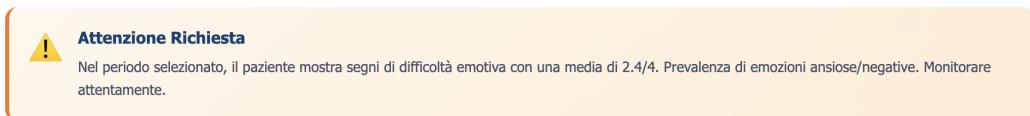


Figure 3.3: Esempio di messaggio generato in base alla media emotiva.

3.4 Estrazione contesto sociale e correlazioni

Il sistema di estrazione del contesto sociale si va ad integrare con l’analisi delle emozioni descritta precedentemente. Ogni nota può essere associata a un solo contesto sociale predominante e, come per le emozioni è presente una spiegazione che giustifica la classificazione. Il sistema riconosce un insieme ampio di contesti sociali, andando a ricoprire tutti gli ambiti della vita quotidiana. In particolare sono presenti: **lavoro, scuola, famiglia, relazione, salute, sport, palestra, tempo libero, hobby, viaggi, casa, finanze, spiritualità, sociale, solitudine, studio, alimentazione, sonno, altro.**

Dal punto di vista implementativo, le fasi di **costruzione del prompt, validazione e normalizzazione** dell’output seguono lo stesso approccio metodologico e i criteri di valutazione descritti nella Sezione 3.3.

3.4.1 Aggregazione dei dati

Il sistema di correlazione prende tutte le note del paziente e, per ogni nota che ha sia un contesto sociale che un’emozione predominante, estrae questi due attributi e li analizza insieme. Per ogni contesto sociale identificato nelle note del paziente, il sistema ha un dizionario con diverse informazioni statistiche:

- Conteggio di emozioni positive, neutre, ansiose e negative riscontrate in quello specifico contesto;
- Numero totale di note in quel contesto e la somma dei punteggi emotivi (per calcolare la media), in maniera tale da poter visionare quanto un contesto emotivo sia positivo o meno per il paziente. Ad esempio, un contesto con una media di 3.7 indica che il paziente tendenzialmente prova emozioni positive in quell’ambito.

3.4.2 Visualizzazione e analisi dei dati

Per evidenziare la correlazione tra contesto sociale e stato emotivo sono stati realizzati due grafici: il primo mostra la distribuzione delle emozioni per ciascun contesto, mentre il secondo la media emotiva associata.

Il primo è un grafico a barre che mostra per ogni contesto sociale presente nelle note del paziente, la **distribuzione delle categorie emotive**. Questo grafico permette al terapeuta di visualizzare in maniera semplice ed intuitiva il profilo emotivo di ciascun contesto sociale.



Figure 3.4: Sull'asse orizzontale sono presenti i contesti ordinati per numero totale di note, dal più frequente al meno frequente, sull'asse verticale il numero di note.

Il secondo grafico è un grafico a barre orizzontali che mostra, per ogni contesto, la **media emotiva** calcolata su tutte le note di quel contesto. Ogni barra è colorata in base alla media: verde per medie alte (≥ 3.2 , zona positiva), viola per medie medio-alte (≥ 2.5 , zona equilibrata), giallo per medie medio-basse (≥ 1.5 , zona di attenzione), rosso per medie basse (< 1.5 , zona critica). Questa visualizzazione permette di identificare visivamente le aree problematiche e quelle funzionali.

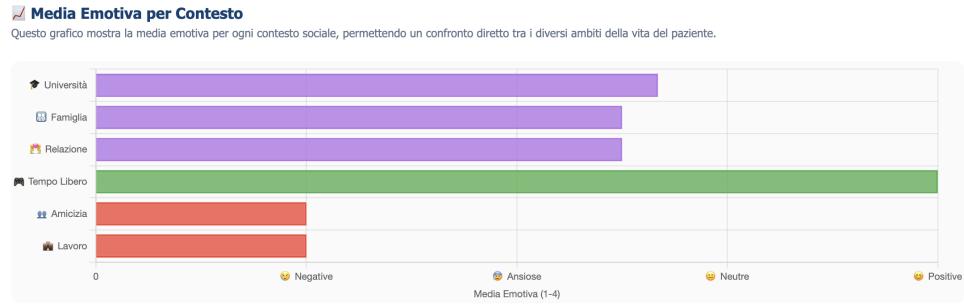


Figure 3.5: L'asse verticale elenca i contesti; mentre l'asse orizzontale mostra la media emotiva su una scala da 1 a 4.

3.5 Rilevazione contenuti di rischio e policy di risposta

SoulDiaryConnect2.0 integra un sistema di rilevazione automatica dei contenuti di rischio che analizza ogni nota del paziente alla ricerca di segnali di pericolo imminente. Nel momento in cui un paziente invia una nuova nota al diario, prima che vengano

eseguite tutte le varie analisi, il sistema analizza il contenuto della nota alla ricerca di indicatori di rischio. Se viene rilevato un contenuto di rischio, il sistema entra in una **modalità di risposta di emergenza** che ha delle caratteristiche e un flusso diverso da quello standard. In particolare, la generazione automatica delle frasi di supporto da parte dell'LLM viene disabilitata, così come la funzionalità di analisi clinica lato terapeuta. Questo viene fatto per evitare che il modello possa produrre dei messaggi inappropriati o potenzialmente dannosi in una situazione già critica.

Il sistema è in grado di riconoscere tre categorie di emergenza:

- **Suicidio:** Rappresenta la situazione più critica per la vita del paziente. Questa categoria include non solo riferimenti specifici come ad esempio *"voglio uccidermi"*, ma anche segnali più indiretti come *"tutti starebbero meglio senza di me"*;
- **Violenza/Stalking:** Comprende tutti i casi in cui il paziente è vittima attuale o potenziale di aggressioni fisiche, psicologiche o sessuali da parte di altri individui;
- **Autolesionismo:** Identifica comportamenti in cui il paziente si procura intenzionalmente danni fisici senza intento suicida.

Queste tre categorie non sono sufficienti a coprire tutte le possibili situazioni di rischio in salute mentale, ma coprono le categorie più comuni e quelle per le quali esistono dei servizi di emergenza specifici a cui il paziente può accedere direttamente.

Quando il sistema rileva un'emergenza, il paziente riceve un **messaggio di sicurezza** e l'interfaccia del terapeuta viene modificata per evidenziare immediatamente la situazione critica. Nello specifico, per ogni categoria di emergenza, il sistema dispone di un messaggio prestabilito che viene mostrato al paziente subito dopo il salvataggio della nota. Ogni messaggio inizia con una validazione emotiva esplicita:

- Suicidio: *"Capisco che stai attraversando un momento di grande sofferenza. Quello che provi è reale e importante"*;
- Violenza: *"Mi preoccupo per la tua sicurezza. Quello che stai vivendo non è giusto"*;
- Autolesionismo: *"Capisco che stai soffrendo molto e che forse senti il bisogno di sfogare il dolore"*.

La validazione emotiva è seguita poi dall'invito a contattare subito il proprio terapeuta oppure dei servizi di emergenza, con numeri di telefono specifici e informazioni sugli orari. Nel dettaglio:

- Per il rischio suicidio, il sistema fornisce due opzioni: il **Telefono Azzurro** (19696, attivo 24/7) e il **Telefono Amico Italia** (02 2327 2327, attivo 9:00-24:00). Entrambi i servizi offrono ascolto e possono fornire supporto immediato;

- Per la violenza e lo stalking, il sistema fornisce il numero 1522, **numero nazionale antiviolenza e stalking**. Questo servizio, gratuito e attivo 24/7, è specificamente progettato per supportare le vittime di violenza di genere e stalking;
- Per l'autolesionismo, il sistema fornisce gli stessi numeri del rischio suicidio (Telefono Azzurro e Telefono Amico).

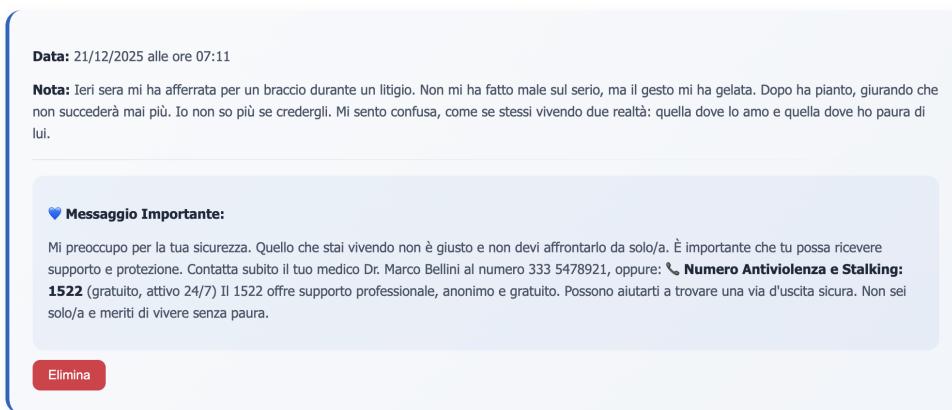


Figure 3.6: Esempio di emergenza rilevata lato paziente.

Il terapeuta visualizza le note critiche con una grafica differente rispetto alle note che non sono state classificate come rischiose. Infatti, le note rilevate come rischiose vengono mostrate con una colorazione sul rosso, per far risaltare subito all'occhio del terapeuta la situazione di emergenza.



Figure 3.7: Esempio di emergenza rilevata lato terapeuta.

3.6 Dettatura vocale in-browser

La funzionalità di dettatura vocale è stata pensata e progettata per facilitare la creazione di nuove note senza la necessità di digitare manualmente il testo, facilitando così la spontaneità. La nuova funzionalità è stata integrata all'interno della pagina home del paziente, nella sezione dedicata alla creazione di nuove note.

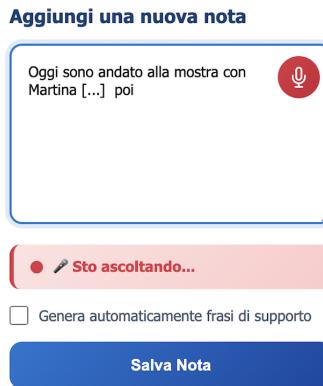


Figure 3.8: Esempio di creazione di una nuova nota tramite dettatura vocale.

L'implementazione si basa sulla **Web Speech API** che fornisce funzionalità di riconoscimento vocale direttamente nel browser. Questa scelta tecnologica permette di integrare capacità di **speech-to-text** senza la necessità di installare librerie esterne o di configurare infrastrutture server dedicate al processamento audio.

La Web Speech API si compone di due interfacce principali:

- **SpeechRecognition:** si occupa del riconoscimento vocale convertendo parlato in testo (*speech-to-text*);
- **SpeechSynthesis:** genera voce sintetica da testo scritto (*text-to-speech*).

Nel contesto di SoulDiaryConnect2.0, essendo l'obiettivo quello di facilitare l'input testuale da parte dei pazienti, è stata utilizzata esclusivamente l'interfaccia di SpeechRecognition. L'utilizzo dei motori di riconoscimento nativi del browser garantisce velocità di trascrizione elevate e latenza minima, aspetti cruciali per fornire feedback in tempo reale all'utente.

3.6.1 Compatibilità e gestione della continuità

La prima sfida riguarda la compatibilità browser: la Web Speech API è pienamente supportata solo nei **browser basati su Chromium** (Chrome, Edge, Opera), mentre

Firefox non ha implementato il supporto e Safari offre solo un supporto parziale.

Un secondo aspetto critico, strettamente legato all'architettura di processamento, è la **gestione della continuità della registrazione**. Poiché la Web Speech API trasmette l'audio a server remoti (tipicamente di Google) per l'elaborazione, l'interazione è soggetta a timeout imposti dal browser dopo periodi di silenzio o a interruzioni dovute a latenze di rete. Per garantire un'esperienza utente fluida in un contesto terapeutico, dove le pause di riflessione sono frequenti e necessarie, è stato implementato un **meccanismo di riavvio automatico** all'interno dell'handler `onend`. Il sistema verifica lo stato di una variabile di controllo (`isRecording`): se l'evento di interruzione viene scatenato mentre il flag è ancora impostato su `true`, il sistema deduce che l'arresto non è stato richiesto dall'utente e il riconoscimento viene riavviato automaticamente:

```

1  recognition.onend = function() {
2      if (isRecording) {
3          // Riavvia se deve continuare
4          try {
5              recognition.start();
6          } catch (e) {
7              console.log('Recognition già attiva');
8          }
9      } else {
10          // Logica di chiusura standard
11          updateRecordingUI(false);
12          // Pulizia del testo dai marcatori provvisori
13          const textarea = document.getElementById('notaTextarea');
14          textarea.value = textarea.value.replace(/\[\.\.\.\]\.*$/ , '').trim();
15      }
16  };

```

Questo meccanismo è fondamentale per l'usabilità del sistema; infatti, senza il riavvio automatico, ogni pausa prolungata costringerebbe il paziente a riattivare manualmente il microfono, interrompendo il flusso narrativo ed emotivo.

3.7 Componenti sviluppate dal co-autore (panoramica)

Come anticipato precedentemente, questo progetto di tesi è stato sviluppato in collaborazione con il collega e amico Francesco Maddaloni. Nello specifico, il collega si è occupato degli aspetti gestionali e di IA Generativa relativa al sistema di generazione frasi di supporto, analisi cliniche e riassunto dello stato clinico del paziente. Tali funzionalità verranno descritte nei paragrafi successivi.

3.7.1 Prompt Engineering

Una parte fondamentale del lavoro dal co-autore si basa sulla ristrutturazione ed ottimizzazione dei prompt utilizzati per la generazione dei testi destinati sia al terapeuta che al paziente. In tutti i prompt è stato utilizzato l'approccio **few-shot prompting** fornendo al modello un esempio di input e di output.

3.7.1.1 System Prompt per la generazione delle note di supporto

Il prompt è stato riprogettato **migliorando l'empatia** della frase di supporto generata e sono state inserite più indicazioni su come gestire la risposta. Infatti, è stata aggiunta una sezione all'interno del prompt denominata *"Istruzioni"* nella quale vengono elencate alcune linee guida sulla formattazione e modalità di risposta.

```

1  prompt = f"""
2      Sei un assistente empatico e di supporto emotivo. Il tuo compito è
3      rispondere con calore e comprensione a persone che stanno attraversando
4      momenti difficili.
5
6      {contesto_paziente} Esempio:
7          Testo del paziente: "Ho fallito il mio esame e ho voglia di arrendermi."
8          Risposta di supporto: "Mi dispiace molto per il tuo esame. È normale sentirsi
9          delusi, ma questo non definisce il tuo valore come persona. Potresti provare
10         a rivedere il tuo metodo di studio e chiedere aiuto se ne hai bisogno.
11         Ce la puoi fare!"
12
13     ISTRUZIONI:
14         - Rispondi in italiano con tono caldo, empatico e incoraggiante
15         - Riconosci e valida le emozioni espresse
16         - Offri una prospettiva positiva senza minimizzare i sentimenti
17         - Suggerisci delicatamente possibili strategie o riflessioni utili
18         - Non usare un tono clinico o distaccato
19         - Completa sempre la risposta, non troncare mai a metà
20         - NON confondere l'autore del testo con altre persone menzionate nella nota
21
22     Testo del paziente:
23     {testo}
24
25     Rispondi con una frase di supporto:"""

```

3.7.1.2 Architettura Prompts per la Gestione delle tipologie di analisi cliniche per lo psicoterapeuta

Il sistema permette al terapeuta di scegliere una configurazione personalizzata per la generazione dell'analisi clinica, combinando due parametri: la lunghezza e la struttura.

Le opzioni disponibili sono:

- **Analisi breve** (max 300 caratteri), strutturata o non strutturata;
- **Analisi lunga** (max 500 caratteri), strutturata o non strutturata.

Precedentemente era presente un unico prompt generico sia per l'analisi breve che per quella lunga, dove l'unica differenza consisteva nel numero massimo di token. Questo sistema tuttavia impediva all'LLM di differenziare realmente le due tipologie di analisi, producendo degli output spesso troncati o della lunghezza non corretta. Per tale motivo si è deciso di realizzare **4 prompt differenti**, uno per ogni possibile combinazione. In questo modo, le istruzioni sulla lunghezza e sulla struttura sono semanticamente integrate nel corpo del prompt, garantendo risultati coerenti e completi.

3.7.2 Sistema di riassunto multi-nota su intervalli temporali per l'analisi del paziente

Tramite una schermata dedicata, il terapeuta può generare dei riassunti periodici che forniscono un quadro clinico preciso e dettagliato, basato sulle note scritte dal paziente in un arco temporale specifico. Le opzioni temporali selezionabili sono: ultimi 7 giorni, ultimo mese, ultimi 3 mesi o ultimo anno. Il documento generato è organizzato in sezioni strutturate, che presentano i dati del paziente e gli aspetti salienti della sua condizione emersi nel periodo selezionato.

Figure 3.9: Schermata del riassunto clinico.

Inoltre, il sistema permette di visualizzare anche l'elenco delle note originali scritte nello stesso intervallo, consentendo al terapeuta di approfondire o rileggere i contenuti nel dettaglio.

3.7.3 Gestione asincrona delle chiamate al modello

Per migliorare l’usabilità del sistema, le chiamate al modello per la generazione dei testi (sia lato terapeuta che lato paziente) sono state rese asincrone. Quando il paziente inserisce una nuova nota, questa viene salvata nel database con lo stato ”In Generazione” e appare un modal che avvisa dell’elaborazione in corso. Appena termina la generazione, il paziente visualizza subito la frase di supporto. Contemporaneamente, il terapeuta visualizza la nota con dei box animati che segnalano l’esecuzione della sentiment analysis, dell’estrazione del contesto sociale e dell’analisi clinica.

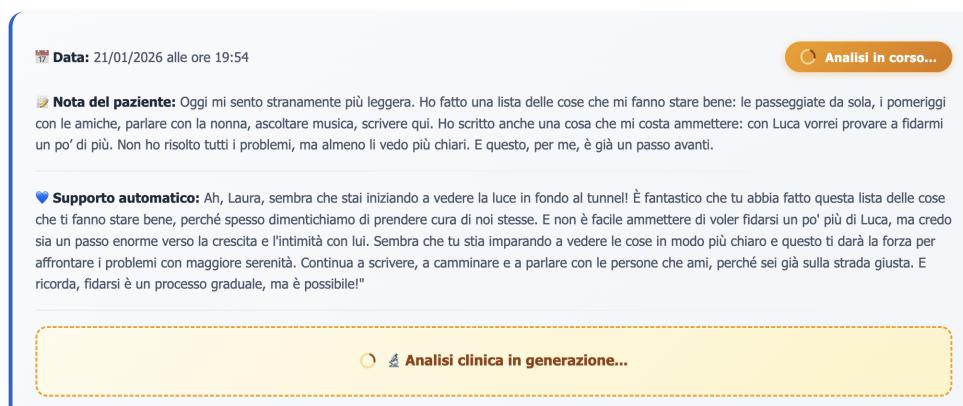


Figure 3.10: *Feedback generazione in corso lato terapeuta*

Una volta concluse le operazioni, la pagina si aggiorna automaticamente permettendo al terapeuta di visualizzare tutti i dati.

3.7.4 Pagina admin Django

Per agevolare le attività di debug e testing, è stata configurata l’interfaccia di amministrazione di Django. Questo pannello consente di visualizzare tutte le tabelle del database e di intervenire direttamente sui dati tramite operazioni di creazione, modifica ed eliminazione. Il sistema include inoltre funzionalità di ricerca per filtrare i record e, per garantire la sicurezza, visualizza le password degli utenti in forma mascherata (oscurata da simboli generici).

3.8 Ottimizzazioni e affinamenti del sistema

Nei paragrafi seguenti verrà mostrata l’applicazione SoulDiaryConnect nella sua veste rinnovata, ponendo l’accento sulle migliorie tecniche e funzionali apportate al sistema per garantire una maggiore stabilità e una migliore esperienza d’uso.

3.8.1 Miglioramento UI/UX

L’interfaccia utente (UI) dell’applicazione è stata sottoposta a un significativo processo di redesign, mirato non solo a un aggiornamento estetico, ma soprattutto a un miglioramento dell’esperienza utente (UX) complessiva: mentre l’iterazione precedente del sistema si focalizzava primariamente sulla funzionalità, presentando un layout talvolta schematico e colori piatti, la nuova versione adotta un linguaggio visivo moderno e dinamico. L’obiettivo è stato quello di creare un ambiente digitale più accogliente per il paziente e uno strumento di lavoro più efficiente per il terapeuta.

Le principali modifiche sono state:

- **Divisione in card:** I contenuti, come le note del diario e i relativi messaggi di supporto, sono ora organizzati all’interno di ”schede” (cards) con angoli arrotondati e ombreggiature morbide. Questo approccio conferisce profondità all’interfaccia, separando visivamente i diversi elementi e migliorando la leggibilità dei testi lunghi;
- **Palette cromatica e gradienti:** L’uso di colori piatti è stato sostituito da sfumature e gradienti, in particolare nell’header e nei pulsanti principali;
- **Uso di icone:** È stato introdotto un set di icone coerente per guidare l’utente nelle azioni. Ad esempio, i messaggi di supporto sono immediatamente riconoscibili grazie all’icona del cuore, l’input vocale è suggerito dall’icona del microfono e così via;
- **Gerarchia Visiva:** I pulsanti di azione primaria (come ”*Salva Nota*” o ”*Ri-genera analisi clinica*”) sono ora visivamente più evidenti rispetto agli elementi secondari, guidando l’attenzione dell’utente verso le operazioni più importanti.

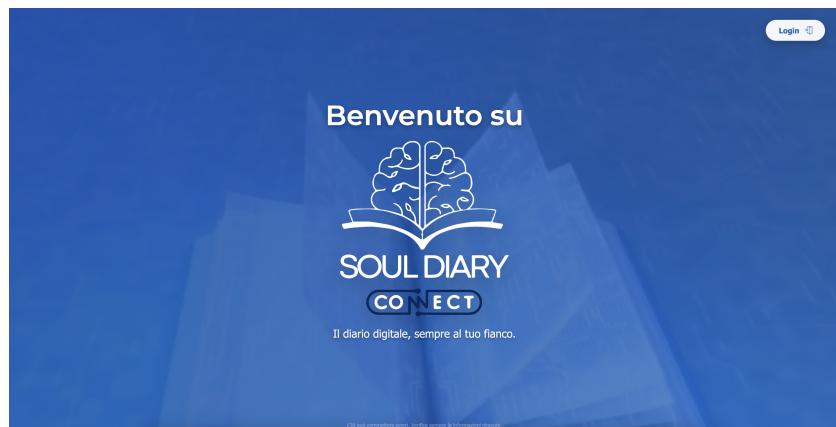


Figure 3.11: Nuova homepage di SoulDiaryConnect.



Figure 3.12: Nuova schermata principale lato paziente.



Figure 3.13: Nuova schermata principale lato medico.

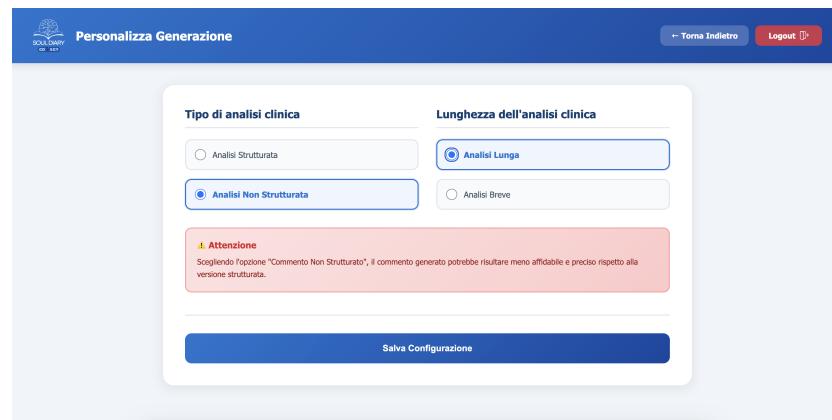


Figure 3.14: Nuova schermata di personalizzazione.

Questo rinnovamento grafico rende l'interazione più fluida e intuitiva per entrambe le tipologie di utenza.

3.8.1.1 Nuovo logo

È stato introdotto un nuovo logo per donare un'identità visiva rinnovata al progetto. La realizzazione grafica del nuovo logo è stata affidata alla graphic designer Sara Benefico. Il redesign mira a trasmettere i valori di supporto, connessione e sicurezza che caratterizzano la piattaforma, segnando anche visivamente il passaggio alla versione aggiornata del sistema.

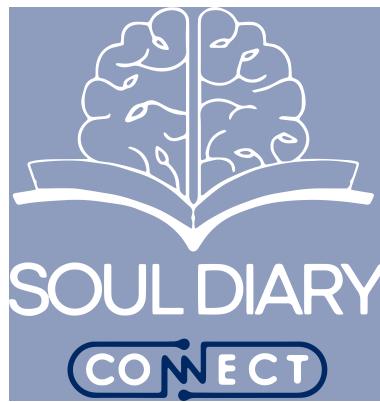


Figure 3.15: Nuovo logo di SoulDiaryConnect.

3.8.2 Rigenerazione dell'analisi clinica lato terapeuta

È stata introdotta la possibilità di **rigenerare l'analisi**. Qualora il terapeuta ritenga che l'analisi clinica prodotta non sia sufficientemente accurata o desideri cambiare il formato (passando ad esempio da un'analisi lunga a una breve, vedi Sez. 3.7.1.2), può modificare i parametri di configurazione e lanciare una nuova generazione. Al termine del processo, l'interfaccia si aggiorna automaticamente mostrando i nuovi dati.

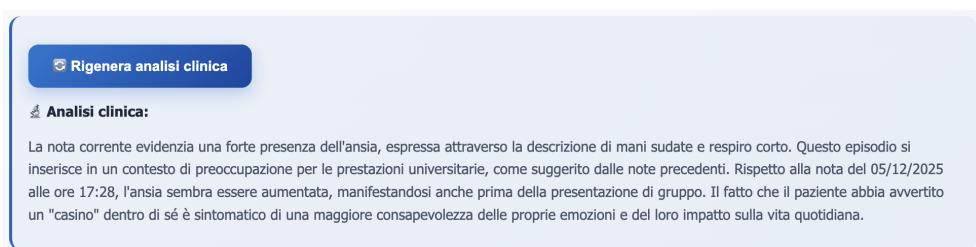


Figure 3.16: Tasto di rigenerazione analisi clinica.

3.8.3 Generazione differita della frase di supporto lato paziente

Nelle versioni precedenti del sistema, la scelta di ricevere la frase di supporto era vincolata al momento della creazione della nota: se il paziente deselezionava l'opzione iniziale, non aveva alcuna possibilità di richiedere la frase di supporto in un secondo momento, se non eliminando la nota e riscrivendola da capo.

Per risolvere questa limitazione, è stata introdotta la funzionalità di generazione differita. Ora il paziente ha la piena libertà di salvare una nota puramente come sfogo personale, senza attivare l'IA. Qualora cambiasse idea successivamente, il sistema riconosce l'assenza della frase di supporto e mostra un pulsante dedicato all'interno della scheda della nota.

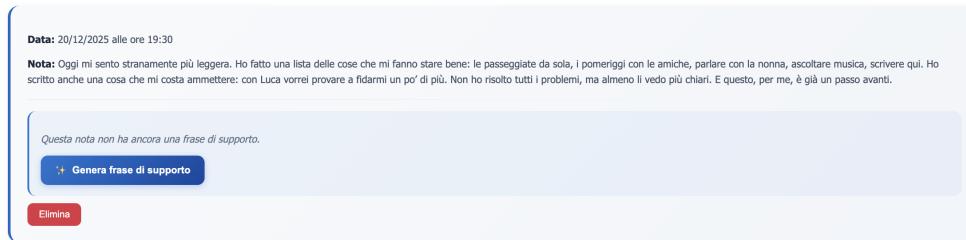


Figure 3.17: Tasto di generazione differita della frase di supporto.

Questa modifica permette all'utente di decidere di ricevere un supporto emotivo solo quando si sente pronto, senza dover forzare l'interazione al momento della scrittura.

Capitolo 4

Risultati

Questo capitolo presenta i risultati della fase di testing di SoulDiaryConnect2.0. L'obiettivo principale di questa fase è stato verificare non solo la correttezza tecnica delle funzionalità implementate, ma soprattutto la loro utilità, usabilità e sicurezza all'interno di un reale contesto clinico. A tal fine, sono stati coinvolti esperti del settore (psicologi e psicoterapeuti) per raccogliere feedback qualificati.

4.1 C Involgimento degli esperti e struttura della fase di test

Il test è stato condotto attraverso delle sessioni individuali con 6 professionisti della salute mentale. L'approccio adottato è stato di tipo ibrido, combinando una dimostrazione guidata con una fase di utilizzo diretto, per valutare sia la comprensione delle funzionalità che l'esperienza d'uso immediata.

Durante lo svolgimento del test ci sono state 3 fasi distinte:

1. **Spiegazione e dimostrazione:** Inizialmente è stata presentata agli esperti una panoramica del progetto, illustrando le problematiche che il sistema intende risolvere. Successivamente, è stato mostrato il flusso completo dell'applicazione, dalla creazione di una nota lato paziente fino all'analisi dei dati lato terapeuta;
2. **Prova diretta e domande:** Agli esperti è stato chiesto di interagire direttamente con il sistema, navigando tra le dashboard e testando le varie funzionalità, in particolare è stato chiesto loro di analizzare i testi generati dall'IA (frasi di supporto, analisi clinica, spiegazioni umore e contesto sociale e riassunti). Inoltre, durante questa fase, sono stati raccolti dubbi, domande e prime impressioni "a caldo";

3. **Somministrazione del questionario:** Al termine della sessione, è stato chiesto ai partecipanti di compilare un questionario su Google Form per formalizzare i loro feedback.

Di seguito viene illustrato il questionario sottoposto ai partecipanti del test:

Valutazione SoulDiaryConnect2.0

* Indica una domanda obbligatoria

Quanto ritieni utile l'applicativo? *

1 2 3 4 5

Per niente utile Molto utile

Come valuteresti l'accessibilità dell'applicativo? *

Ti è risultato intuitivo navigare su SoulDiaryConnect2.0?

La tua risposta

In breve, come valuteresti le interazioni dell'IA con il paziente? *

La tua risposta

In breve, come valuteresti le interazioni dell'IA con il medico? *

La tua risposta

Quanto ritieni utile la sezione riassunto caso clinico? *

1 2 3 4 5

Per niente utile Molto utile

Figure 4.1: Domande presenti nel questionario sottoposto ai partecipanti del test - Parte 1.

4.1. COINVOLGIMENTO DEGLI ESPERTI E STRUTTURA DELLA FASE DI TEST37

Quanto ritieni utile la sezione riguardante l'analisi dell'umore? *

1 2 3 4 5

Per niente utile Molto utile

Quanto ritieni utile l'identificazione delle note che contengono situazioni che richiedono l'intervento di un ente esterno? *

1 2 3 4 5

Per niente utile Molto utile

Quanto ritieni utile la possibilità di personalizzare la generazione dell'analisi clinica? *

1 2 3 4 5

Per niente utile Molto utile

Quanto ritieni utile la possibilità di visualizzare per ogni nota l'umore predominante, il contesto sociale e la relativa spiegazione della classificazione dell'IA? *

1 2 3 4 5

Per niente utile Molto utile

Avresti qualche suggerimento per migliorare l'applicativo?

La tua risposta

Figure 4.2: Domande presenti nel questionario sottoposto ai partecipanti del test - Parte 2.

4.2 Risultati e feedback

L'analisi dei dati raccolti tramite i questionari ha evidenziato un elevato grado di soddisfazione complessiva. I punteggi attribuiti alle funzionalità chiave si sono attestati quasi uniformemente sui valori massimi della scala Likert (4 e 5), confermando la validità delle scelte progettuali.

In particolare, i tester hanno espresso un forte apprezzamento per la **personalizzazione dell'analisi clinica** e per il sistema di **rilevazione dei rischi**, considerati strumenti indispensabili per la tutela del paziente. Molto positiva è stata anche la valutazione della **sezione riassunti**: questa funzionalità è stata definita da uno dei terapeuti come "*un'evoluzione dei classici appunti clinici*", capace di ottimizzare significativamente i tempi di revisione tra una seduta e l'altra.

Anche l'accessibilità dell'interfaccia è stata ben valutata: il design è stato giudicato "*intuitivo*" e l'**introduzione dei comandi vocali** è stata valutata come un elemento cruciale per rendere l'applicazione fruibile anche a pazienti con difficoltà motorie o visive. Un altro aspetto interessante emerso riguarda l'utilità del sistema come alternativa ai canali di comunicazione informali; un esperto ha notato come l'applicativo possa aiutare a gestire quei pazienti che tendono a scrivere su WhatsApp, offrendo loro uno spazio di sfogo immediato e strutturato senza invadere la privacy del terapeuta.

Nonostante l'apprezzamento generale, i professionisti hanno fornito suggerimenti mirati per perfezionare il sistema in vista di un utilizzo reale.

4.2.1 Sicurezza e gestione delle emergenze

Un punto fondamentale emerso riguarda la tempestività. Il sistema attuale riconosce correttamente i rischi (suicidio, autolesionismo o violenza) e segnala la nota in rosso nella dashboard. Tuttavia, i tester hanno confermato un'ipotesi che avevamo già considerato in fase di progettazione: la sola segnalazione a schermo non basta. Se il medico non è loggato nel sito, rischia di vedere l'allarme troppo tardi. È stato quindi confermato che sarebbe necessario un sistema di **notifiche attive** (sul telefono o via email) che avvisi il professionista in tempo reale appena viene rilevato un pericolo.

4.2.2 Trasparenza e consenso

Attualmente, l'applicazione mostra già un avviso in tutte le schermate che ricorda che "*L'IA può commettere errori*". I tester hanno ritenuto questa dicitura utile, ma hanno suggerito di fare un passo in più: inserire un **consenso esplicito in fase di registrazione**. Sia il medico che il paziente nel momento in cui creano l'account, dovrebbero accettare che i loro testi verranno elaborati da un'intelligenza artificiale, per garantire la massima trasparenza.

4.2.3 Miglioramenti agli strumenti clinici

Sono emersi due suggerimenti specifici per l'analisi:

- **Focus sull'ansia:** L'ansia è uno dei problemi più comuni per la quale una persona decide di iniziare un percorso di psicoterapia. Un terapeuta ha suggerito di non trattarla come una semplice emozione tra le tante, ma di dedicarle uno spazio separato con un misuratore specifico, per vederne l'andamento nel tempo in modo più chiaro;
- **Obiettivi della terapia:** È stata evidenziata la mancanza di una sezione dove poter seguire i progressi su obiettivi specifici (es. "dormire meglio", "gestire la rabbia"). Attualmente l'analisi è generale, mentre sarebbe utile poter avere una sezione dove poter definire dei traguardi specifici concordati con il paziente e avere di conseguenza delle analisi a riguardo.

Capitolo 5

Conclusioni

Con il presente lavoro di tesi è stato progettato e sviluppato **SoulDiaryConnect2.0**, andando a migliorare il sistema originale. Il percorso di ricerca ha previsto la transizione verso un LLM più performante (Llama-3.1), l'integrazione di moduli avanzati per l'analisi delle emozioni, la gestione del contesto storico e l'implementazione di un sistema di sicurezza per la rilevazione automatica dei contenuti a rischio.

La fase di validazione condotta con gli esperti del settore ha confermato la validità dell'approccio metodologico: lo strumento è stato valutato come un **supporto efficace per il terapeuta**, in grado di offrire una visione d'insieme difficilmente ottenibile tramite strumenti tradizionali. Il sistema si configura attualmente come un prototipo avanzato e funzionale, ponendo basi solide per ulteriori iterazioni evolutive.

5.1 Sviluppi futuri

Al fine di rendere SoulDiaryConnect un prodotto maturo e usabile su larga scala, sono stati individuati diversi sviluppi futuri. Questi sono stati divisi tra le evoluzioni già pianificate e i suggerimenti emersi durante la fase di test.

Il passaggio naturale per il futuro del progetto potrebbe essere lo **sviluppo di un'applicazione mobile** (iOS e Android). Oltre a migliorare l'usabilità per il paziente, questa soluzione permetterebbe di risolvere il problema della tempestività emerso nei test: tramite l'app, si potrebbero infatti inviare **notifiche push** immediate sul dispositivo del medico in caso di emergenza.

Attualmente il sistema elabora solo le ultime 5 note per il contesto. Una possibile evoluzione riguarda l'**implementazione della tecnologia RAG** (Retrieval-Augmented Generation). Questo approccio permetterebbe al medico di "interrogare" l'intero diario del paziente tramite domande in linguaggio naturale (es. *"Il paziente ha mai parlato di insomnia negli ultimi sei mesi?"*), con il sistema in grado di cercare le risposte nel-

l'intero storico dei dati, superando i limiti della memoria attuale.

Per semplificare la procedura di registrazione, si potrebbe sostituire l'attuale sistema di codici identificativi con dei **link di invito generati dal medico**. In questo scenario ipotetico, il paziente riceverebbe un link diretto (via email o messaggistica) che, una volta cliccato, completerebbe automaticamente l'associazione con il terapeuta.

Si potrebbe infine valutare l'uso dei dati storici per tentare di **individuare pattern ricorrenti** (es. cali dell'umore nel weekend) e fornire piccoli alert predittivi. Inoltre, la funzionalità di dettatura vocale potrebbe essere arricchita analizzando, oltre al testo, anche le **caratteristiche vocali** (tono, pause, volume) per estrarre ulteriori indicatori emotivi.

5.1.1 Proposte derivanti dai test

Accogliendo i feedback, si potrebbe valutare di separare l'emozione *”ansia”* dalle altre macro-categorie. L'introduzione di un **”misurometro dell'ansia”** dedicato nei grafici offrirebbe al terapeuta uno strumento più granulare per monitorare questo specifico disturbo.

Per superare il limite di un'analisi clinica generica, una possibile integrazione riguarda l'implementazione di una sezione dedicata alla **definizione dei traguardi terapeutici** (es. *”Gestire la rabbia”*, *”Ridurre l'isolamento sociale”*). Tale modulo consentirebbe di definire obiettivi specifici concordati con il paziente e di istruire l'IA tramite prompt mirati, affinché generi analisi precise relative all'avvicinamento o all'allontanamento da tali traguardi.

Infine, per rispondere alle esigenze di trasparenza sollevate, sarebbe opportuno integrare nel flusso di registrazione una **richiesta di consenso esplicito**. Questo passaggio formale garantirebbe che l'utente sia pienamente consapevole dell'uso dell'IA nell'analisi dei propri dati.

Bibliografia

- [1] Kaiser Family Foundation. Exploring the rise in mental health care use by demographics and insurance status, 2025. URL <https://www.kff.org/mental-health/exploring-the-rise-in-mental-health-care-use-by-demographics-and-insurance-status/>. [Visitato in data 25/01/2026].
- [2] Redazione PsicoHumanitas. Che cos'è la salute mentale? <https://psicohumanitas.it/che-cose-la-salute-mentale/>, October 2024. [Visitato in data 30/01/2026].
- [3] Luca Casillo. SoulDiaryconnect: Large language models a supporto della psicoterapia, 2024. Progetto di Tesi, Università degli Studi di Salerno.
- [4] World Health Organization. COVID-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide. [https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-a-nxiety-and-depression-worldwide](https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide), March 2022. [Visitato in data 25/01/2026].
- [5] Md Rezaul Kader, Mohammad Mahmudur Rahman, Piali Dey Bristi, and Foyez Ahmed. Change in mental health service utilization from pre-to post-covid-19 period in the united states. *Heliyon*, 10(22), 2024.
- [6] Mark Olfson, Chandler McClellan, Samuel H Zuvekas, Melanie Wall, and Carlos Blanco. Psychotherapy trends in the united states. *American Journal of Psychiatry*, 182(5):483–492, 2025.
- [7] Alex Green from Pexels. Study marks rise in psychotherapy outpatient visits and declines in medication use for mental health care. <https://medicalxpress.com/news/2025-05-psychotherapy-outpatient-declines-medication-mental.html>, May 2025. [Visitato in data 25/01/2026].
- [8] Columbia University's Mailman School of Public Health. Study shows significant rise in psychotherapy use among adults, but gains are uneven across socioeconomic groups. <https://www.sciencedaily.com/releases/2024/12/241204113537.htm>, December 2024. [Visitato in data 25/01/2026].

- [9] Paola Biasi, Maria De Paola, and Paolo Naticchioni. Il Disagio Psicologico in Italia: una fotografia basata sulle richieste del Bonus Psicologo. <https://eticaeconomia.it/il-disagio-psicologico-in-italia-una-fotografia-basata-sulle-richieste-del-bonus-psicologo/>, October 2023. [Visitato in data 25/01/2026].
- [10] Yining Hua, Hongbin Na, Zehan Li, Fenglin Liu, Xiao Fang, David Clifton, and John Torous. A scoping review of large language models for generative tasks in mental health care. *npj Digital Medicine*, 8(1):230, 2025.
- [11] Surendrabikram Thapa and Surabhi Adhikari. Gpt-4o and multimodal large language models as companions for mental wellbeing. *Asian Journal of Psychiatry*, 99:104157, 2024. ISSN 1876-2018. doi: <https://doi.org/10.1016/j.ajp.2024.104157>. URL <https://www.sciencedirect.com/science/article/pii/S1876201824002508>.
- [12] Sebastian Acevedo, Esha Aneja, Douglas J Opler, Pamela Valera, and Eric Jarmon. Evaluating the efficacy of chatgpt-3.5 versus human-delivered text-based cognitive-behavioral therapy: A comparative pilot study. *American Journal of Psychotherapy*, pages appi-psychotherapy, 2025.
- [13] Wanghao Dong, Weijun Wang, Xinheng Han, Junhao Huang, Lie Li, and Yinghui Huang. Can gpt-4 provide human-level emotion support? insights from machine learning-based evaluation framework. *Computers in Biology and Medicine*, 196: 110789, 2025. ISSN 0010-4825. doi: <https://doi.org/10.1016/j.combiomed.2025.110789>. URL <https://www.sciencedirect.com/science/article/pii/S0010482525011400>.
- [14] Nick Obradovich, Sahib S Khalsa, Waqas U Khan, Jina Suh, Roy H Perlis, Olusola Ajilore, and Martin P Paulus. Opportunities and risks of large language models in psychiatry. *NPP—Digital Psychiatry and Neuroscience*, 2(1):8, 2024.
- [15] Christopher Lalk, Kim Targan, Tobias Steinbrenner, Jana Schaffrath, Steffen Eberhardt, Brian Schwartz, Antonia Vehlen, Wolfgang Lutz, and Julian Rubel. Employing large language models for emotion detection in psychotherapy transcripts. *Frontiers in Psychiatry*, 16:1504306, 2025.
- [16] Akshat Santhana Gopalan. Enhancing Therapeutic Conversations with Sentiment Analysis in Natural Language Processing. <https://behavioralhealthnews.org/enhancing-therapeutic-conversations-with-sentiment-analysis-in-natural-language-processing/>, October 2024. [Visitato in data 25/01/2026].
- [17] Taewan Kim, Seolyeong Bae, Hyun Ah Kim, Su-Woo Lee, Hwajung Hong, Chanmo Yang, and Young-Ho Kim. Mindfuldiary: Harnessing large language model to support psychiatric patients' journaling. In *Proceedings of the 2024 CHI*

- Conference on Human Factors in Computing Systems, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400703300. doi: 10.1145/3613904.3642937. URL <https://doi.org/10.1145/3613904.3642937>.
- [18] Qiming Guo, Jinwen Tang, Wenbo Sun, Haoteng Tang, Yi Shang, and Wenlu Wang. Soullmate: An adaptive llm-driven system for advanced mental health support and assessment, based on a systematic application survey, 2025. URL <https://arxiv.org/abs/2410.11859>.
 - [19] Huachuan Qiu and Zhenzhong Lan. PsyDial: A large-scale long-term conversational dataset for mental health support. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, editors, Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 21624–21655, Vienna, Austria, jul 2025. Association for Computational Linguistics. ISBN 979-8-89176-251-0. doi: 10.18653/v1/2025.acl-long.1049. URL <https://aclanthology.org/2025.acl-long.1049/>.

Ringraziamenti

Desidero innanzitutto rivolgere un sentito ringraziamento al mio relatore, il Professor Rocco Zaccagnino, per la disponibilità dimostrata durante lo svolgimento di questo lavoro. Un ringraziamento speciale va anche ai dottorandi Alessia e Dino per il supporto tecnico, i consigli e il tempo che mi hanno dedicato nel corso di questi mesi.

Un grazie immenso va ai miei genitori. Grazie per avermi sempre assecondata e per non avermi fatto mancare mai nulla, ma soprattutto grazie per la vostra presenza costante. Non dimenticherò mai i chilometri fatti solo per venirmi a prendere per pochi decimi di febbre, pur di non lasciarmi sola o evitarmi lo stress dei mezzi pubblici. Se oggi ho raggiunto questo traguardo lo devo in primis a voi: spero di non deludervi mai e di rendervi sempre orgogliosi (tranquilli, la villa con piscina è sempre più vicina!). Grazie a mia sorella Dora, che ho importunato ripetutamente nei momenti di noia; grazie per le abbuffate offerte di sushi ogni volta che tornavo a casa: la dose massiccia di Omega-3 è stata fondamentale per dare la carica al mio cervello e studiare meglio. A proposito, quando mi porti?! Infine, un pensiero al piccolo Max, disturbato a distanza anche solo per un cup cup. Grazie per aver creduto sempre in me.

A Giuseppe (alias Pepe, Pepitos, Patatino, Brokky, Trokky, Puzeta, ... e la lista dei nomi potrebbe continuare all'infinito). Probabilmente dovrebbero farti una statua per la pazienza che hai ogni giorno nel sopportarmi. Ci siamo trovati in un momento complicato per entrambi e ci siamo aiutati a vicenda a venirne fuori; il tuo supporto costante è stato fondamentale, specialmente per tutte quelle volte in cui hai creduto in me molto più di quanto non facessi io stessa. Grazie <3.

A Sara, la mia spalla destra. Questo è un altro traguardo raggiunto insieme e spero che la lista continui ad allungarsi, sempre con te al mio fianco a sostenermi. Come recita quella canzone, *”sempre sarai nella tasca destra in alto”*.

Ai miei amici fidati: Roberto, Manu, Dex, Mary, Alex e Marco. Non importa se ci sentiamo ogni giorno o una volta al mese, so con certezza che potrò sempre contare su di voi in qualsiasi momento.

A Francesco (CiscoFran, Francisco). Che dire, ce l'abbiamo fatta nonostante i problemi, i disagi e le persone a "84728 facce" (ogni riferimento a Voldy è puramente voluto). Grazie per avermi aiutata a fare pulizia nella mia vita, per le sessioni "anti-depressione" e per i momenti di sano procrastinamento a suon di ping pong, bowling, Mario Kart e Mario Party. Un grazie speciale anche a Poldino, anche lui (in)giustamente importunato nei momenti bui. Grazie anche a Chiara e Luca (Lucas Vázquez) per avermi accolta nel gruppo senza mai farmi sentire fuori posto: senza voi tre, gli esaurimenti nervosi sarebbero stati molti di più.

Un pensiero speciale va come sempre ai miei nonni Paolo e Maria e a mio zio Mimmo: anche se non possono essere qui a festeggiare con me, so che non hanno mai smesso di accompagnarmi e che oggi sarebbero stati orgogliosi di questo risultato.

Pagina bianca per chiunque voglia lasciarmi un segno, una dedica o un consiglio su come affrontare il post-laurea senza impazzire.

