



UNIVERSITÀ DEGLI STUDI DI SALERNO

Dipartimento di Informatica

Corso di Laurea Triennale in Informatica

TESI DI LAUREA

*SoulDiaryConnect:*  
**Large Language models a supporto  
della psicoterapia**

RELATORE

Prof. Rocco Zaccagnino

Dott. Gerardo Benevento

Università degli Studi di Salerno

CANDIDATO

**Luca Casillo**

Matricola: 0512113696

Anno Accademico 2023-2024

*Questa tesi è stata realizzata nel*



*"A chi si trova perdendosi"*

*Nicholaus Von Letthus*

## Abstract

L'Intelligenza Artificiale (AI) sta rivoluzionando il settore della salute mentale, offrendo strumenti innovativi per il supporto terapeutico. Il *journaling digitale assistito da AI* si sta affermando come soluzione efficace per aiutare i pazienti ad esprimere emozioni e pensieri, migliorando il monitoraggio clinico e la personalizzazione dei trattamenti. Sono tuttavia presenti limitazioni sugli strumenti attuali, tra cui mancanza di personalizzazione, difficoltà nell'adattamento alle esigenze del paziente e integrazione limitata con il lavoro e l'esperienza del medico.

Si introduce in questa tesi *SoulDiaryconnect*, un sistema AI-based progettato per migliorare il supporto psicologico digitale. Sviluppato come un'estensione del sistema *MindfulDiary*, un diario digitale basato su AI che ha dimostrato di facilitare la narrazione terapeutica, *SoulDiaryconnect* utilizza l'AI per definire una connessione bidirezionale tra paziente e medico: da un lato permette al paziente di documentare esperienze quotidiane in modo intuitivo e discreto (*diario*), mentre dall'altro, offre risposte generate da un *Large Language Model (LLM)* personalizzabile dal medico e adattabile alle specifiche esigenze terapeutiche sia del medico che del paziente.

Per valutare l'efficacia del sistema, sono state fornite al sistema pagine di diario del mondo reale scritte da pazienti con necessità di supporto terapeutico, e valutate le risposte dell'LLM integrato da un'esperta psicoterapeuta. I risultati preliminari ottenuti hanno mostrato la capacità del sistema di facilitare la narrazione terapeutica, migliorare l'aderenza al trattamento e fornire un valido supporto ai professionisti della salute mentale, ma anche evidenziato delle limitazioni legate all'integrazione di LLM nel *journaling* terapeutico. In conclusione, noi riteniamo che questo lavoro di tesi possa offrire una prospettiva teorica e un'implementazione pratica per l'integrazione degli LLM nel supporto terapeutico, cercando di fare un passo avanti verso l'adozione di AI personalizzabili nella pratica clinica, garantendo un approccio flessibile e centrato sul paziente, pur mantenendo un collegamento costante con il medico, che ne supervisiona e personalizza l'intervento.

---

# Indice

---

<b>Elenco delle Figure</b>	<b>iii</b>
<b>1 Introduzione</b>	<b>1</b>
1.1 Journaling e AI nella Salute Mentale . . . . .	1
1.2 LLM e Conversational AI . . . . .	2
<b>2 Related Works</b>	<b>3</b>
2.1 Dagli anni '60 ad oggi . . . . .	3
2.2 Progressi Attuali . . . . .	3
<b>3 Background</b>	<b>5</b>
3.1 Mental Health . . . . .	5
3.1.1 Salute mentale: Un Problema Sociale e Tecnologico . . . . .	5
3.1.2 La comunicazione e il supporto psicologico . . . . .	6
3.1.3 Le tecnologie digitali e il supporto alla salute mentale . . . . .	6
3.2 Intelligenza Artificiale(AI) . . . . .	7
3.2.1 AI: Machine Learning e Deep Learning . . . . .	7
3.2.2 I tipi di dati . . . . .	7
3.2.3 I tipi di apprendimento . . . . .	8
3.2.4 Le Reti Neurali Artificiali . . . . .	10
3.2.5 Natural Language Processing (NLP) . . . . .	16

---

3.2.6	Principali Applicazioni del NLP . . . . .	17
3.2.7	Rappresentazione del Testo . . . . .	18
3.2.8	Sfide del NLP . . . . .	20
3.3	Large Language Models . . . . .	21
3.3.1	L'architettura Transformer . . . . .	21
3.3.2	Fine-Tuning e Transfer Learning . . . . .	25
3.3.3	Le sfide degli LLM . . . . .	25
3.4	Caso di studio: MindfulDiary . . . . .	26
3.4.1	Obiettivi del sistema . . . . .	26
3.4.2	Funzionamento del sistema . . . . .	27
3.4.3	Tecnologie utilizzate . . . . .	28
3.4.4	Risultati preliminari . . . . .	28
3.4.5	Limiti del sistema e sviluppi futuri . . . . .	28
<b>4</b>	<b>Application</b>	<b>29</b>
4.1	Architettura del Sistema . . . . .	29
4.1.1	Panoramica del Sistema per l'utente . . . . .	30
4.1.2	Gestione ed Elaborazione Dati (Backend) . . . . .	33
4.1.3	Modulo di NLP . . . . .	34
<b>5</b>	<b>Conclusioni</b>	<b>39</b>
5.1	Analisi . . . . .	39
5.1.1	Valutazioni di un esperto del dominio . . . . .	40
5.2	Sviluppi Futuri . . . . .	41
5.3	Conclusioni Finali . . . . .	41
	<b>Bibliografia</b>	<b>43</b>

---

## Elenco delle figure

---

3.1	Percentuali di persone con problemi di salute mentale negli USA (2022).	6
3.2	Esempio di architettura di una ANN . . . . .	11
3.3	Effetto del Learning Rate sull'ottimizzazione . . . . .	15
3.4	Convergenza con Learning Rate ideale . . . . .	16
3.5	Architettura Transformer ([1]) . . . . .	22
4.1	Schermata dell'applicativo lato paziente . . . . .	31
4.2	Schermata dell'applicativo lato medico . . . . .	32
4.3	Schermata per la personalizzazione della generazione del testo . . .	32
4.4	Diagramma Entità-Relazione del Sistema . . . . .	34

# CAPITOLO 1

---

## Introduzione

---

### 1.1 Journaling e AI nella Salute Mentale

Il *journaling* è da tempo riconosciuto come uno strumento efficace per il benessere psicologico, consentendo agli individui di esprimere pensieri ed emozioni in modo autentico [2, 3]. Pennebaker et al. [4, 5], affermano che scrivere sulle proprie esperienze, sia negative che positive, aiuta a elaborare stati emotivi complessi, riducendo stress, ansia e depressione. In ambito clinico, il journaling si è rivelato utile per raccogliere dati sulle condizioni quotidiane dei pazienti, migliorando il monitoraggio dei sintomi e il trattamento [6]. Tuttavia, molte persone trovano difficile strutturare una narrazione coerente dei propri pensieri e sentimenti, specialmente in contesti terapeutici [7]. L'avvento delle tecnologie digitali ha aperto nuove possibilità nel journaling, con strumenti come chatbot e assistenti AI che facilitano l'espressione personale e migliorano la raccolta di dati sullo stato emotivo [8].

Negli ultimi anni, l'*Intelligenza Artificiale* (AI) ha trasformato numerosi settori, dalla diagnosi medica ai sistemi decisionali intelligenti [9, 10]. Nella salute mentale, l'AI sta emergendo come una soluzione per affrontare la crescente domanda di servizi psicoterapeutici, spesso limitati da risorse insufficienti [11, 12]. Tradizionalmente, la cura della salute mentale si è basata su consulenze in presenza, ma questo modello



sta diventando meno sostenibile [13, 14]. Tecnologie AI come il *Machine Learning* e il *Natural Language Processing* (NLP) stanno rivoluzionando il settore, consentendo di sviluppare strumenti per la previsione degli esiti terapeutici, interventi automatizzati e sistemi di supporto per i clinici. Sistemi basati su AI, come chatbot terapeutici e assistenti virtuali, hanno dimostrato di poter alleviare i sintomi di ansia e depressione, offrendo uno spazio sicuro e sempre accessibile per l'espressione personale [15]. Tuttavia, la ricerca è ancora in evoluzione, con la necessità di studi su larga scala per validare l'efficacia di questi strumenti [16].

## 1.2 LLM e Conversational AI

I progressi nei modelli linguistici di grandi dimensioni (LLM), come GPT-4, LLaMA e PaLM-2, hanno rivoluzionato l'NLP, consentendo interazioni più fluide e personalizzate. LLM avanzati offrono nuove opportunità per il journaling digitale, permettendo a chatbot e assistenti AI di adattarsi dinamicamente alle esigenze degli utenti [17]. A differenza dei chatbot tradizionali, che utilizzano risposte predefinite, gli LLM possono generare conversazioni più naturali e contestualizzate. Questa caratteristica è particolarmente utile per il supporto psicologico, dove è fondamentale fornire un'interazione empatica e personalizzata [8]. Tuttavia, l'utilizzo degli LLM in ambito clinico presenta sfide significative, come il controllo della generazione delle risposte, la gestione di bias nei modelli AI e il rispetto di standard etici e normativi [18]. Chiaramente, collaborazioni tra esperti di AI e professionisti della salute mentale sono essenziali per sviluppare sistemi terapeutici che siano da un lato avanzati tecnologicamente, ma dall'altro anche affidabili e sicuri.

## CAPITOLO 2

---

### Related Works

---

#### 2.1 Dagli anni '60 ad oggi

L'integrazione dell'AI nella psicoterapia rappresenta una nuova frontiera nel settore della salute mentale, con il potenziale di aumentare l'efficacia e l'accessibilità dei trattamenti. Fin dagli anni '60, con il programma *ELIZA*<sup>1</sup>, l'AI ha mostrato il suo potenziale nelle applicazioni terapeutiche, simulando conversazioni umane per supportare i pazienti. Negli anni successivi, i sistemi AI hanno cercato di replicare l'expertise psicologica umana, portando alla creazione di strumenti diagnostici e terapeutici avanzati (Coppersmith et al., 2014) [19].

#### 2.2 Progressi Attuali

Con i progressi nel Machine Learning, in particolare nel Deep Learning, e nel Natural Language Processing (NLP), modelli di Large Language Model (LLM) come GPT-4, PaLM-2, LLaMA, e Mental-LLaMA sono stati testati per il supporto alla salute mentale (Singhal et al., 2023 [20]; Jiang et al., 2023 [21]). L'AI moderna è in grado di analizzare il linguaggio naturale, riconoscere emozioni e fornire risposte contestual-

---

<sup>1</sup>[https://it.wikipedia.org/wiki/ELIZA\\_\(chat\\_bot\)](https://it.wikipedia.org/wiki/ELIZA_(chat_bot))

mente appropriate, migliorando significativamente l'esperienza terapeutica digitale (Yang et al., 2023)[22, 23].

Esempi dei risultati raggiunti dagli LLM possiamo riscontrarli in recenti sistemi come SouLLMate [24] e MindGuard [25], che mirano a fornire supporto mentale accessibile, personalizzato e in tempo reale attraverso tecnologie AI all'avanguardia. SouLLMate integra modelli linguistici di grandi dimensioni con tecniche di Retrieval-Augmented Generation e *prompt engineering* per offrire funzionalità avanzate come il rilevamento del rischio di suicidio e dialoghi di guida proattiva. MindGuard, invece, è un sistema mobile progettato per fornire un primo soccorso mentale accessibile e privo di stigma, integrando dati sensoriali mobili con valutazioni ecologiche momentanee per offrire conversazioni di screening e intervento personalizzate.

Un altro esempio è il sistema MindfulDiary, proposto da Kim et al [26] nel Maggio 2024. Esso utilizza un LLM e il sentiment analysis per assistere i pazienti psichiatrici nella documentazione delle esperienze quotidiane mediante somministrazione guidata di note quotidiane. MindfulDiary, progettato in collaborazione con professionisti di salute mentale, adotta un approccio basato su stati per conformare il percorso del paziente in modo sicuro alle linee guida degli esperti, senza però permettere che essi possano interfacciarsi direttamente al paziente.

Tuttavia, nonostante i progressi, gli attuali modelli generativi mostrano ancora limitazioni nell'accuratezza delle risposte, nella personalizzazione dei consigli personali ai pazienti e terapeutici e nella comprensione del contesto clinico.

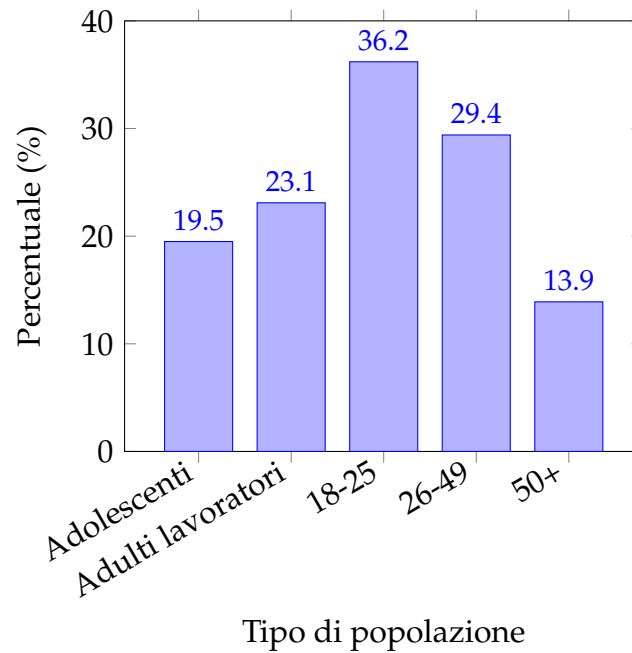
Al fine di supportare il medico nella personalizzazione dei consigli e migliorare la comprensione del contesto clinico da parte della componente AI, viene proposto **SoulDiaryConnect**. Questo sistema ha l'obiettivo di assistere il paziente nel suo percorso psicoterapeutico quotidiano, offrendo un supporto basato sul Natural Language Processing, potenzialmente fruibile anche offline, privo di stigma sociale e costantemente connesso al medico.

SoulDiaryConnect consente al professionista di personalizzare il metodo con cui l'AI genera il testo, adattandolo alla sua esperienza e alle esigenze del paziente, e di mantenere un contatto diretto con lui. Questo approccio offre al medico la possibilità di intervenire attivamente nel percorso terapeutico, favorendo un'interazione più efficace e mirata tra paziente e sistema di supporto.

### 3.1 Mental Health

#### 3.1.1 Salute mentale: Un Problema Sociale e Tecnologico

La salute mentale è una componente fondamentale del benessere umano. Essa infatti influenza la qualità della vita, la produttività lavorativa e le relazioni sociali. In base a quanto riportato dal report del 2023 di Angelini Pharma in partnership con The European House - Ambrosetti [27], circa il 20% della popolazione italiana in età lavorativa soffre di disturbi mentali. Dati analoghi ci arrivano dal 2022 National Survey on Drug Use and Health (NSDUH) [28] del Substance Abuse and Mental Health Services Administration (SAMHSA) negli Stati Uniti d’America, il quale afferma che anche nella popolazione adulta dei lavoratori americani, il 23,1% soffre di disturbi mentali, come mostrato in Figura 3.1. Tuttavia, l’accesso a cure adeguate resta, dal 1997 come riportato da Bland et al. [29], limitato a causa di barriere economiche, sociali e logistiche. Ciò, unito a una crescente consapevolezza di queste problematiche, ha portato alla ricerca di soluzioni digitali per colmare il divario tra domanda e offerta di servizi di supporto psicologico, oltre che offrire la possibilità ai pazienti di fornire un supporto sempre più costante.



**Figura 3.1:** Percentuali di persone con problemi di salute mentale negli USA (2022).

### 3.1.2 La comunicazione e il supporto psicologico

La comunicazione svolge un ruolo cruciale nel supporto psicologico. Studi scientifici hanno dimostrato che una buona comunicazione può migliorare significativamente il benessere e aiutare la gestione di condizioni come ansia e depressione, come riportato da Galderisi et al. nel 2017 [30]. Tuttavia, con patologie come depressione e disturbi alimentari, modelli di comunicazione inefficaci possono aggravare le condizioni dei pazienti (Chiariello e Orvaschel, 1995 [31]; Kiesler, 1979 [32]). Molte persone, inoltre, evitano di cercare aiuto a causa dello stigma sociale e degli elevati costi dei trattamenti (Bland et al. [29]), rendendo necessario lo sviluppo di strategie comunicative innovative.

### 3.1.3 Le tecnologie digitali e il supporto alla salute mentale

Negli ultimi decenni, e soprattutto dall'avvento del COVID-19, le tecnologie digitali hanno rivoluzionato il mondo della comunicazione e del trattamento dei pazienti, compreso il settore della salute mentale. Infatti, sono stati introdotti strumenti accessibili e anonimi per la gestione del benessere psicologico. Social media, applicazioni mobile e chatbot terapeutici offrono soluzioni a basso costo e in tempo reale per il

supporto ai pazienti, come riportato nel 2020 da Naslund et al. [33], da Abd-Alrazaq et al. [34], e già dal 2018 da Neary e Schueller [35]. Questi strumenti aiutano a ridurre lo stigma, migliorando l'accessibilità ai servizi, ma presentano anche limitazioni. Le attuali piattaforme, al momento, faticano a garantire una comunicazione e un supporto psicologico personalizzato e affidabile, in quanto spesso mancano delle capacità di comprensione e di rispondere adeguatamente ai bisogni dei medici e dei pazienti, che nel dominio di interesse rappresentano gli utenti. Tali limitazioni potrebbero essere superate con l'integrazione dell'Intelligenza Artificiale (AI).

## **3.2 Intelligenza Artificiale(AI)**

### **3.2.1 AI: Machine Learning e Deep Learning**

L'Intelligenza Artificiale (AI), e in particolar modo il Machine Learning (ML), ha subito una crescita esponenziale negli ultimi anni, diventando sempre più utilizzata e fondamentale in svariati ambiti, come classificazione, calcolo avanzato, previsione di dati futuri ed elaborazione del linguaggio naturale. Il Machine Learning è un sottoinsieme dell'Intelligenza Artificiale, il quale fornisce ai sistemi la capacità di apprendere ed evolversi automaticamente dall'esperienza, senza un'esplicita programmazione a riguardo [36]. Il concetto di autoapprendimento si riferisce alla 'capacità' dei computer di apprendere dai dati similmente a come gli esseri umani acquisiscono conoscenze dall'esperienza. Le macchine non necessitano di essere programmate per affrontare ogni singola situazione; ma, partendo dallo studio dei numerosi esempi forniti, sono in grado di capire autonomamente come risolvere problemi dello stesso tipo in futuro.

### **3.2.2 I tipi di dati**

Uno degli aspetti cruciali di ogni sistema di ML o AI è l'analisi dei dati. Prima che i modelli possano imparare e fare previsioni, i dati devono essere acquisiti, pre-processati e organizzati in modo tale che possano essere facilmente interpretati dalle macchine [37]. In tale contesto, i dati possono essere classificati principalmente in 3 categorie:

1. **Dati strutturati:** si tratta di dati con una struttura rigida, tipicamente organizzati in tabelle con righe e colonne, come fogli di calcolo o database relazionali. Ogni elemento presenta una posizione specifica e un significato definito, il che rende tali dati facili da gestire e analizzare.
2. **Dati semi-strutturati:** questi non seguono uno schema tabellare rigido, avendo però una struttura interna. Esempi sono documenti JSON o XML, che contengono campi o tag predefiniti per facilitare l'organizzazione delle informazioni. Possono essere organizzati parzialmente per agevolare l'analisi all'AI.
3. **Dati non strutturati:** questi dati rappresentano la categoria più ampia di dati. Essi comprendono testi, immagini, foto, audio e video. Non è presente una struttura organizzata, e per questo l'elaborazione di dati non strutturati ha portato alla necessità di sviluppare tecniche avanzate di estrazione delle informazioni.

La scelta riguardante il tipo di dati da usare dipende dal compito specifico che si intende affrontare nel dominio di interesse. Solitamente i dati strutturati sono ideali per compiti che presentano una natura di per sé già organizzata, come nel caso di applicazioni gestionali e finanziarie. I dati semi-strutturati, d'altra parte, offrono una maggiore flessibilità per scenari in cui alcune informazioni sono organizzate ma richiedono comunque una certa elaborazione. Infine, i dati non strutturati, come testo, immagini o video, sono fondamentali per compiti di elaborazione avanzata, come l'analisi delle immagini o il riconoscimento del linguaggio naturale.

### 3.2.3 I tipi di apprendimento

Nel Machine Learning i modelli di AI devono imparare a riconoscere pattern e a prendere decisioni sulla base dei dati in loro possesso. Il modo in cui i modelli apprendono, però, può variare a seconda della struttura e della disponibilità dei dati, oltre che dal compito specifico da svolgere. L'apprendimento è solitamente diviso in tre principali tipi di approcci: *Predittivo* o *Supervisionato*, *Descrittivo* o *Non Supervisionato*, e *Per Rinforzo*. Verranno di seguito illustrate le differenze tra questi approcci utilizzando la spiegazione fornita da Kevin P. Murphy nel suo libro "Machine Learning: A Probabilistic Perspective" [38].

**Apprendimento Supervisionato** In questo approccio l'obiettivo è imparare una mappatura dagli input  $x$  agli output  $y$ , dato un insieme di coppie input-output etichettate  $D = \{(x_i, y_i)\}_{i=1}^N$ . Qui,  $D$  è chiamato *Training Set* e  $N$  è il numero di esempi di addestramento. Si consideri ogni input di addestramento  $x_i$  come un vettore di numeri a  $D$  dimensioni, che rappresentano, per esempio, l'altezza e il peso di una persona. Questi sono chiamati *Feature*, *Attributi* o *Covariate*. In generale,  $x_i$  potrebbe essere un oggetto strutturato complesso come un'immagine, una frase, una serie temporale, etc. Analogamente anche l'output  $y_i$  può essere in linea di principio qualsiasi cosa, ma di solito si assume che sia una variabile categorica nominale proveniente da un insieme finito. Quando  $y_i$  è categorico, il problema è noto come *Classificazione* o *Riconoscimento di Pattern*, quando  $y_i$  assume valori reali, il problema è noto come *Regressione*. Un'altra variante nota come *Ordinal Regression*, si verifica quando lo spazio delle etichette  $Y$  ha un ordinamento naturale.

**Apprendimento Non Supervisionato** Nell'approccio Non Supervisionato vengono forniti soltanto gli input  $D = \{x_i\}_{i=1}^N$  e l'obiettivo è trovare pattern interessanti nei dati. Si tratta di un problema molto meno definito poiché non viene specificato quale tipo di pattern cercare siccome i dati non sono etichettati. Una tecnica comune che fa utilizzo di questo tipo di apprendimento è il *Clustering* dove il modello raggruppa i dati in categorie o cluster basati su somiglianze. Algoritmi comuni includono:

- *K-means*: divide i dati in  $K$  gruppi, cercando di minimizzare la distanza tra i punti e il centroide del gruppo.
- *DBSCAN*: identifica cluster di densità, individuando punti che formano aree densamente popolate.

**Apprendimento Per Rinforzo** Questo tipo di apprendimento è utile per insegnare al modello come agire o comportarsi quando vengono forniti segnali occasionali di ricompensa o punizione. L'obiettivo è massimizzare una ricompensa cumulativa nel tempo, scegliendo le azioni che portano ai risultati migliori.



I diversi tipi di apprendimento offrono quindi strumenti flessibili per affrontare una vasta gamma di problemi di Machine Learning, ciascuno con le proprie caratteristiche e applicazioni ottimali.

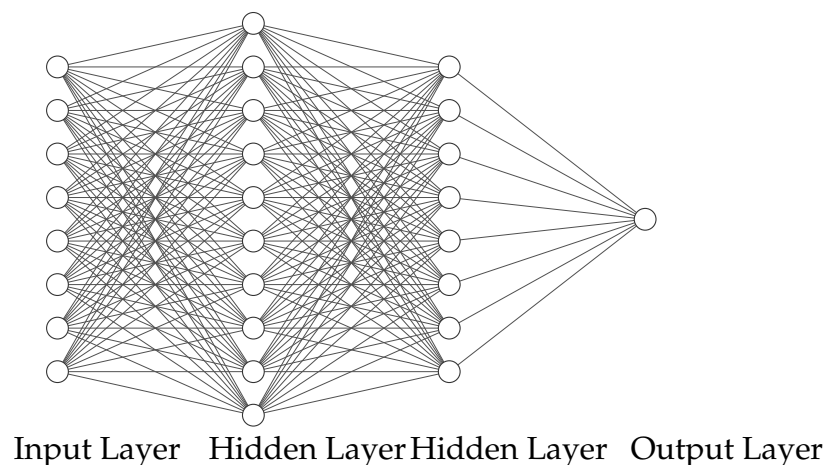
### 3.2.4 Le Reti Neurali Artificiali

Le prestazioni degli algoritmi di ML più semplici dipendono molto dalla rappresentazione dei dati (sottosezione 3.2.2) che ricevono[39]. Si consideri un obiettivo come quello di effettuare la diagnosi precoce dell'Alzheimer attraverso tomografie computerizzate (TAC). Un algoritmo di ML semplice non esamina le immagini TAC del cervello del paziente direttamente, bensì riceve informazioni strutturate fornite dal medico, come la presenza di specifiche atrofie cerebrali o riduzioni del volume dell'ippocampo (caratteristiche comuni nei pazienti affetti da Alzheimer). Queste informazioni, chiamate *features*, vengono usate dall'algoritmo per fare previsioni sulla probabilità che quel determinato paziente sviluppi la malattia. Tuttavia, se il sistema ricevesse come dati in input la TAC, formata da migliaia di pixel non etichettati, esso non sarebbe in grado di fornire predizioni. I singoli pixel, da soli, non hanno una correlazione comprensibile con l'insorgenza della malattia, pertanto l'algoritmo dipende completamente dalle features selezionate e date dal medico. In maniera simile, anche operazioni 'semplici' come la classificazione delle email in *spam* o *non spam* possono risultare difficili per algoritmi di ML se i dati in input non sono rappresentati correttamente, rendendo complicato ottenere previsioni con un buon grado di accuratezza. Per altri compiti, invece, risulta estremamente complesso determinare in anticipo le features da dover estrarre ed utilizzare. Questo ha portato alla necessità di algoritmi più avanzati, come quelli di Deep Learning (DL), i quali possono apprendere automaticamente le rappresentazioni rilevanti dai dati grezzi, senza il bisogno che esse siano progettate manualmente.

Il DL, facendo riferimento al testo di Ian Goodfellow et al. dal titolo "Deep Learning"[39], si basa su una gerarchia di concetti, con i concetti più semplici che formano quelli più complessi. Questa struttura gerarchica viene rappresentata attraverso le *Reti Neurali Artificiali* o *Artificial Neural Network (ANN)*, che consistono in più strati di neuroni artificiali. Differentemente dai modelli tradizionali, le ANN hanno

la capacità di apprendere rappresentazioni sempre più astratte dei dati, permettendo la risoluzione di compiti complessi come il riconoscimento delle immagini e la traduzione automatica, segnando un'evoluzione e un cambiamento radicale nell'AI. I sistemi di DL, a differenza di quelli di ML, non richiedono più una programmazione esplicita di regole complesse, ma acquisiscono la conoscenza necessaria attraverso l'addestramento su grandi moli di dati, permettendo loro di generalizzare bene anche compiti che richiedono una difficile formalizzazione.

Le ANN, alla base del DL, rappresentano un modello computazionale ispirato alla struttura e al funzionamento del cervello umano. Esse sono costituite da strati di unità di elaborazione denominate *neuroni artificiali*, organizzate in un'architettura a più livelli: partendo da uno strato di input, passando per uno o più strati nascosti, arrivando a uno strato di output. Ogni neurone presente in un determinato strato è collegato attraverso dei pesi ai neuroni dello strato successivo, che rappresentano l'importanza delle informazioni trasmesse [40]. Per guidare il processo di addestramento si utilizza una *Funzione di Perdita*, la quale quantifica la differenza tra il valore predetto dal modello e il valore reale che viene osservato. L'obiettivo principale è rendere minima questa differenza, essendo una delle principali valutazioni delle prestazioni del modello. Di seguito, in Figura 3.2 viene riportato un esempio di architettura di una ANN, con uno strato composto da 8 nodi, due strati nascosti rispettivamente di 10 e 8 nodi, e uno strato di output con un singolo nodo.



**Figura 3.2:** Esempio di architettura di una ANN

Il processo di elaborazione delle informazioni in una ANN si basa sulla trasforma-

zione dei dati in input attraverso i vari strati, dove ogni neurone calcola una somma pesata degli input applicando una *Funzione di Attivazione* non lineare. Tra le varie Funzioni di Attivazione, due delle più usate nelle ANN sono la *Sigmoide* e la *Rectified Linear Unit (ReLU)*.

**Sigmoide** La funzione Sigmoide è definita come segue:

$$f(z) = \frac{1}{1 + e^{-z}} \quad (3.2.1)$$

Questa funzione mappa ogni input in un intervallo compreso tra 0 e 1. È quindi particolarmente utile per problemi di classificazione binaria. Tuttavia, una delle sue limitazioni è il *Vanishing Gradient Problem*, in cui i gradienti nelle fasi successive dell'addestramento diventano molto piccoli, rallentando l'aggiornamento dei pesi.

**ReLU** La funzione ReLU è definita come segue:

$$f(z) = \max(0, z) \quad (3.2.2)$$

Essa restituisce l'input se è positivo, zero altrimenti. La funzione ReLU è comunemente utilizzata negli ultimi anni perché i problemi di *Vanishing Gradient* sono meno frequenti rispetto alle Funzioni di Attivazione Sigmoide che si saturano in entrambe le direzioni. Uno dei principali problemi della ReLU è che può presentare il *Dying ReLU Problem*, il quale si verifica quando i neuroni ricevono input negativi o nulli per un lungo periodo di tempo. Dato che la funzione ReLU restituisce 0 per input negativi, i neuroni possono entrare in uno stato in cui non producono alcun output (venendo definiti "morti") e non partecipano più al processo di apprendimento, perché i gradienti non possono più fluire attraverso di loro per aggiornarne i pesi.

L'aggiornamento dei pesi della ANN avviene in due fasi: il *Forward Pass*, in cui i dati in input attraversano la Rete e generano una previsione, e la *Backpropagation*, che propaga l'errore dalla fine della Rete, ovvero dallo strato di output, agli strati iniziali, ovvero lo strato di input. Verrà di seguito fornita, sulla base di quanto riportato da Yann LeCun et al. in "Deep Learning"[40], una spiegazione sintetica del funzionamento di questo processo, prendendo in considerazione una ANN con due hidden layers.

**Forward Pass** In una ANN con due hidden layers, ogni unità riceve degli input  $x_i$ . Questi vengono moltiplicati per i rispettivi pesi  $w_{ij}$  e sommati per ottenere il valore netto di input di ogni nodo del livello successivo:

$$z_j = \sum_{i \in \text{Input}} w_{ij} x_i \quad (3.2.3)$$

Il nodo  $j$  nello strato nascosto  $H1$  riceverà  $z_j$  come input netto. Dopo aver calcolato la somma pesata degli input, si applica una Funzione di Attivazione per ottenere l'output  $y_j$ :

$$y_j = f(z_j) \quad (3.2.4)$$

Questo è il valore di output per il nodo  $j$  nello strato  $H1$ . Successivamente si ha la propagazione verso il secondo hidden layer, dove gli output del primo vengono utilizzati come input per  $H2$ , con i relativi pesi  $w_{jk}$

$$z_k = \sum_{j \in H1} w_{jk} y_j \quad (3.2.5)$$

Il valore netto viene nuovamente attivato con una funzione non lineare per ottenere l'output  $y_k$  del secondo hidden layer  $H2$ . L'output degli strati viene infine utilizzato per calcolare l'output della Rete con pesi  $w_{kl}$ :

$$z_l = \sum_{k \in H2} w_{kl} y_k \quad (3.2.6)$$

L'output finale è  $y_l = f(z_l)$ , dove  $f$  è la Funzione di Attivazione finale.

**Calcolo dell'Errore** Per il calcolo del gradiente dell'errore nello strato di output, si utilizza la derivata parziale della Funzione di Perdita rispetto all'output  $y_l$ . L'errore per l'unità  $l$  dell'output è dato dalla differenza tra l'output  $y_l$  e il valore target  $t_l$ :

$$\frac{\partial E}{\partial y_l} = y_l - t_l \quad (3.2.7)$$

Successivamente, il gradiente dell'errore rispetto all'input netto  $z_l$  all'unità  $l$  è dato dalla derivata parziale della Funzione di Attivazione  $f$  rispetto all'input netto, moltiplicata per il gradiente dell'errore rispetto all'output:

$$\frac{\partial E}{\partial z_l} = \frac{\partial E}{\partial y_l} \frac{\partial y_l}{\partial z_l} \quad (3.2.8)$$

**Backpropagation** Dopo aver calcolato l'errore nello strato di output, l'errore viene propagato indietro attraverso i due hidden layers. Per ciascun nodo  $k$  in un hidden layer, l'errore è calcolato sommando l'errore proveniente dagli strati successivi. In particolare, l'errore per il nodo  $k$  dello strato nascosto  $H2$  è dato da:

$$\frac{\partial E}{\partial z_k} = \sum_{l \in out} w_{kl} \frac{\partial E}{\partial z_l} \quad (3.2.9)$$

Lo stesso processo di Backpropagation viene applicato per calcolare l'errore nei nodi  $j$  del layer  $H1$ , sommando l'errore dai nodi  $k$  di  $H2$

$$\frac{\partial E}{\partial z_j} = \sum_{k \in H2} w_{jk} \frac{\partial E}{\partial z_k} \quad (3.2.10)$$

**Aggiornamento dei pesi** Una volta calcolati i gradienti per tutte le unità nella Rete, si può aggiornare ogni peso  $w_{ij}$  utilizzando una regola di ottimizzazione come il *Gradient Descent*. Questo processo viene gestito da un *Optimizer*, che determina in che modo i pesi devono essere modificati per ridurre l'errore complessivo del modello. Consideriamo l'utilizzo del Gradient Descent:

$$w_{ij} = w_{ij} - \eta \frac{\partial E}{\partial w_{ij}} \quad (3.2.11)$$

Dove:  $w_{ij}$  Peso che collega l'unità  $i$  all'unità  $j$

$\eta$  Tasso di Apprendimento (*Learning Rate*)

$\frac{\partial E}{\partial w_{ij}}$  Gradiente dell'errore rispetto al peso  $w_{ij}$

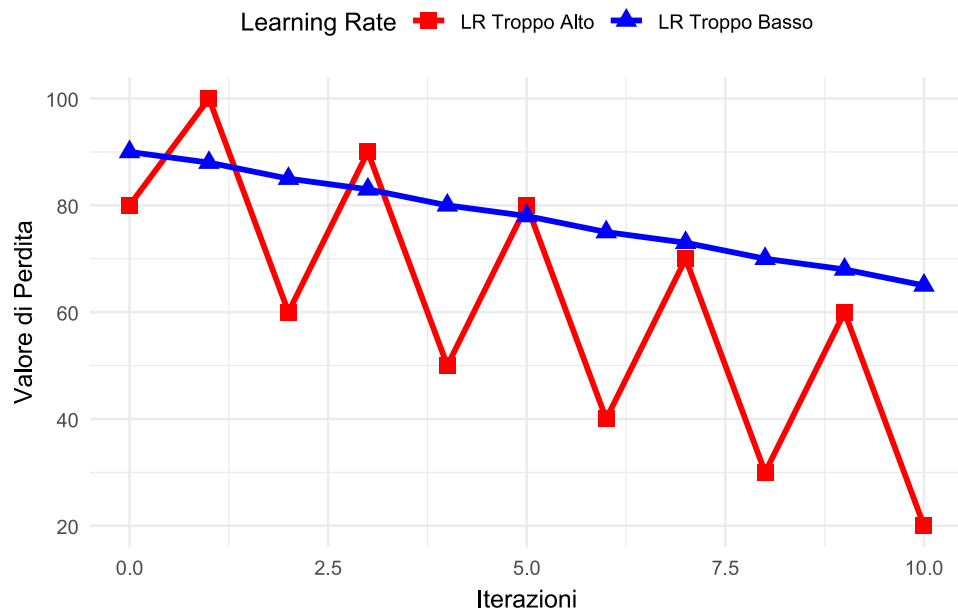
Grazie a questo meccanismo iterativo, una ANN è in grado di apprendere e migliorare le sue prestazioni, adattandosi progressivamente ai dati forniti e rendendo possibile l'addestramento di modelli complessi in grado di risolvere una vasta gamma di problemi.

Nella formula 3.2.11 è stato menzionato il *Tasso di Apprendimento* o *Learning Rate* (*LR*). Esso viene utilizzato dall'*Optimizer* e controlla la dimensione dei passi che l'algoritmo compie verso il minimo della Funzione di Perdita. Determina quindi quanto velocemente o lentamente il modello aggiorna i suoi pesi in risposta all'errore osservato durante l'addestramento.

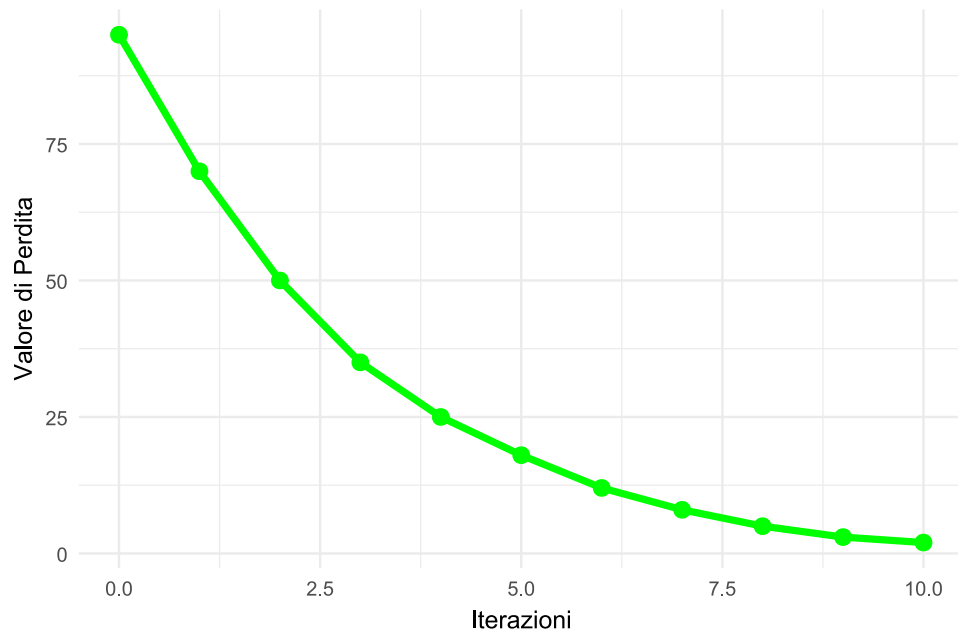
- Se LR è troppo alto, l'*Optimizer* potrebbe "saltare" oltre il minimo, portando a una convergenza instabile o addirittura a una divergenza, dove l'errore di addestramento inizia ad aumentare invece di diminuire.

- Al contrario, se è troppo basso, l'addestramento diventa molto lento perché l'Optimizer fa passi molto piccoli. Ciò può portare a un lungo tempo di addestramento e c'è il rischio che l'addestramento si arresti in un minimo locale anziché globale, a seconda della natura della Funzione di Perdita.

È quindi importante mantenere il LR sufficientemente basso da evitare oscillazioni eccessive ma abbastanza alto da garantire una convergenza efficiente senza rimanere bloccati in minimi locali o richiedere troppe iterazioni.



**Figura 3.3:** Effetto del Learning Rate sull'ottimizzazione



**Figura 3.4:** Convergenza con Learning Rate ideale

### 3.2.5 Natural Language Processing (NLP)

Negli ultimi anni, il *Natural Language Processing (NLP)* ha ottenuto un ruolo centrale nell'AI, permettendo lo sviluppo di modelli in grado di comprendere e generare il linguaggio umano. Il NLP si occupa di trasformare il linguaggio naturale in una forma che possa essere processata dalle macchine, permettendo applicazioni come la traduzione automatica, l'analisi del sentiment e la generazione di testo. Grazie agli avanzamenti nei modelli di linguaggio su larga scala, oggi il NLP ha raggiunto un livello di avanzamento tale che i testi generati da questi modelli risultano spesso indistinguibili da quelli scritti da esseri umani[41].

Le difficoltà principali nel NLP derivano dalla complessità del linguaggio umano, la quale presenta ambiguità, variazioni contestuali e strutture sintattiche articolate[42]. A differenza degli approcci tradizionali basati su regole, i modelli moderni di NLP usano rappresentazioni vettoriali (embedding) per mappare le parole in uno spazio continuo, catturando le relazioni semantiche. Grazie a queste rappresentazioni vetto-

riali, i modelli di NLP possono gestire la complessità del linguaggio umano in modo più efficace, superando i limiti degli approcci basati su regole rigide.

### 3.2.6 Principali Applicazioni del NLP

Il NLP ha portato a una vasta gamma di applicazioni che spaziano dai sistemi di comprensione del linguaggio a modelli in grado di soddisfare diversi tipi di richieste. Tali progressi hanno reso possibile l'uso del NLP in molteplici settori, che vanno dal marketing all'assistenza sanitaria, fino ai servizi finanziari.

**Traduzione Automatica** La traduzione automatica è una delle applicazioni più conosciute del NLP. Inizialmente sviluppata utilizzando modelli basati su regole e successivamente modelli statistici, questa tecnologia ha compiuto un enorme salto qualitativo con l'introduzione delle Reti Neurali Ricorrenti (RNN) utilizzate da Google Translate [43]. Successivamente, l'introduzione dei Transformer ha permesso un ulteriore miglioramento della qualità delle traduzioni, rendendole più fluide e contestuali e ponendo le basi per l'evoluzione verso i moderni modelli di linguaggio di grandi dimensioni [41].

**Analisi del Sentiment** L'analisi del Sentiment è una tecnica utilizzata per estrarre opinioni e stati d'animo da testi, come recensioni dei clienti o commenti sui social media.

**Chatbot e Assistenti Virtuali** I chatbot e gli assistenti virtuali, come Siri di Apple, Alexa di Amazon, Google Assistant e Gemini di Google, sono diventati strumenti sempre più popolari per l'interazione tra macchine e utenti [44]. Grazie ai progressi dei modelli di NLP, i chatbot sono ora in grado di comprendere e rispondere a una gamma di richieste complesse, migliorando l'esperienza dell'utente e aumentando l'efficienza delle interazioni.

**Riconoscimento di Entità Nominate (NER)** Il Riconoscimento di Entità Nominate (NER) è una tecnica che permette di identificare e classificare entità, come nomi di persone, luoghi, e organizzazioni, all'interno di testi. Modelli avanzati come SpaCy e BERT sono spesso utilizzati per migliorare la precisione del NER, permettendo di estrarre informazioni strutturate da testi non strutturati.



**Question Answering (QA)** Negli ultimi anni, il campo del Question Answering (QA) ha compiuto enormi passi avanti grazie all'evoluzione dei modelli di linguaggio di grandi dimensioni, in particolare con la famiglia di modelli GPT, che ha raggiunto livelli di comprensione del linguaggio tali da permettere risposte a domande in modo quasi indistinguibile da un esperto umano [41]. Uno degli aspetti chiave per sfruttare queste capacità è il Prompt Engineering, una tecnica che consente di guidare il modello verso risposte più accurate e inerenti formulando strategicamente l'input [45]. Grazie al Prompt Engineering, i modelli possono essere adattati a una varietà di task di QA, permettendo di ottenere risposte dettagliate e rilevanti senza la necessità di ulteriori fasi di addestramento.

**Moderazione dei Contenuti** Il NLP è diventato fondamentale per la moderazione dei contenuti sui social media, in particolare per il rilevamento delle fake news. Con l'aumento della disinformazione, i modelli di NLP sono stati impiegati per identificare testi ingannevoli o offensivi. Studi recenti hanno utilizzato BERT e RoBERTa per rilevare pattern linguistici indicativi di disinformazione o contenuti dannosi [46].

**Riassunto Automatico del Testo** Il riassunto automatico è diventato uno strumento importante per estrarre i contenuti rilevanti da documenti lunghi, come articoli scientifici e libri. Modelli come T5 [47], BART [48] e Mistral-7B [49] hanno raggiunto risultati all'avanguardia in questa applicazione, facilitando l'accesso rapido alle informazioni.

### 3.2.7 Rappresentazione del Testo

Uno dei primi passi nell'elaborazione del linguaggio naturale è la rappresentazione del testo in una forma numerica comprensibile dai modelli di AI. Nel tempo, sono state sviluppate diverse tecniche per catturare al meglio il significato e il contesto linguistico.

Tra le prime tecniche vi è il *Bag of Words (BoW)*, che converte un documento in un vettore basato sulla frequenza delle parole. Tuttavia, BoW ignora l'ordine delle parole

e le relazioni semantiche, limitandone l'efficacia. Per migliorarne le prestazioni, è stato introdotto *TF-IDF* (*Term Frequency-Inverse Document Frequency*), che riduce il peso delle parole comuni e enfatizza quelle più distintive per ogni documento [42].

Con l'avvento delle reti neurali, sono emerse tecniche più avanzate come Word2Vec (Mikolov et al., 2013), che apprende rappresentazioni vettoriali continue delle parole, posizionando termini semanticamente simili vicino nello spazio vettoriale. Word2Vec include due modelli principali: CBOW, che prevede una parola in base al contesto, e Skip-Gram, che usa una parola per prevedere il contesto circostante. Parallelamente, GloVe (Global Vectors for Word Representation) (Pennington et al., 2014) sfrutta le co-occorrenze globali per catturare relazioni semantiche con ottimi risultati in classificazione testuale e riconoscimento di entità [50].

Più recentemente, i modelli basati su Transformer hanno introdotto embeddings contestuali, superando i limiti delle rappresentazioni statiche come Word2Vec e GloVe. Questi modelli, grazie al pre-addestramento su grandi quantità di dati e al successivo *fine-tuning*, offrono rappresentazioni più dinamiche e adattabili al contesto, migliorando significativamente le prestazioni nei compiti di NLP (sottosezione 3.3.2).

Strumenti fondamentali per l'elaborazione e generazione del testo nell'ambito del DL e del NLP sono le architetture Encoder, Decoder ed Encoder-Decoder.

**Encoder** Trasforma una sequenza di token in una rappresentazione latente, catturando relazioni semantiche e contesto testuale. Questa struttura è ideale per attività di comprensione come classificazione e riconoscimento di entità. Modelli come BERT e RoBERTa sono esempi di architetture Encoder progettate per la comprensione del linguaggio[51, 52].

**Decoder** Genera testo a partire da una rappresentazione latente, operando in modo autoregressivo (predizione token per token). Questa architettura è alla base dei modelli GPT, ottimizzati per generare testo contestualmente coerente [53, 41].

**Encoder-Decoder** Combina entrambe le strutture, rendendola ideale per compiti di traduzione automatica e sintesi testuale. L'Encoder elabora l'input e il Decoder genera l'output. Modelli come T5 e BART si basano su questa architettura per migliorare la trasformazione del testo [47, 48].

L'integrazione di queste architetture con i meccanismi di Attention ha reso possibili modelli NLP avanzati, migliorando la comprensione del contesto e la generazione di risposte più precise [41].

### 3.2.8 Sfide del NLP

Nonostante i progressi nel Natural Language Processing (NLP), in gran parte dovuti al Deep Learning e agli LLM, permangono sfide complesse legate alla natura del linguaggio umano [54].

Una delle principali difficoltà riguarda la comprensione contestuale a lungo termine. Modelli come BERT, GPT e T5 hanno migliorato la capacità di catturare relazioni semantiche, ma faticano a gestire testi lunghi o strutture narrative complesse. La limitazione della lunghezza della sequenza nei modelli Transformer influisce sulla qualità della rappresentazione del contesto, a causa della crescita quadratica della complessità computazionale dell'Attention [1]. Studi recenti hanno introdotto varianti più efficienti dei Transformer per ridurre questa complessità [55], ma spesso con compromessi sulla precisione.

Un'altra sfida critica riguarda i bias nei modelli NLP. Essendo addestrati su grandi dataset estratti da Internet, tendono a riflettere pregiudizi culturali e sociali, portando a potenziali risultati discriminatori [56]. La mitigazione di questi bias è oggi una priorità, con metodi per rilevare e ridurre le distorsioni nei dati e nelle risposte dei modelli [57].

Anche la generalizzazione dei modelli rappresenta una questione aperta. Sebbene le prestazioni siano elevate nei contesti di addestramento, i modelli spesso faticano ad adattarsi a dati nuovi, generando frasi semanticamente incoerenti. Approcci come il *Multi-Task Learning* e il *Domain Adaptation* mirano a migliorare la capacità di adattamento a contesti specifici [58].

Infine, un ostacolo significativo è la mancanza di interpretabilità dei modelli NLP, spesso considerati delle "scatole nere". Questa opacità limita l'adozione in settori critici come la medicina e la finanza, dove è essenziale giustificare le decisioni prese [59]. Nonostante i progressi, rendere interpretabili i modelli di Deep Learning resta una sfida aperta [60].

In conclusione, affrontare le sfide del NLP non significa solo migliorare la tecnologia, ma anche promuovere un approccio etico e trasparente. Lo sviluppo di sistemi NLP affidabili e responsabili è essenziale per garantire un utilizzo sicuro e conforme ai valori sociali.

## 3.3 Large Language Models

L'evoluzione delle ANN e del Deep Learning ha portato allo sviluppo dei *Large Language Models (LLM)*, modelli su larga scala in grado di comprendere e generare testo con un livello di accuratezza senza precedenti. Questi modelli si distinguono per l'enorme quantità di parametri, che possono raggiungere miliardi o addirittura bilioni, come nel caso di GPT-4o e Switch Transformer [61, 62].

I parametri rappresentano le componenti del modello apprese durante l'addestramento e sono essenziali per catturare contesti complessi e sfumature linguistiche. Mentre i modelli tradizionali di Deep Learning operano con milioni di parametri, gli LLM ne utilizzano quantità significativamente superiori, migliorando la capacità di interpretare il linguaggio umano [63]. Modelli come GPT-3, BERT e T5 hanno dimostrato un notevole impatto in molteplici applicazioni, dalla sanità all'educazione, ridefinendo le modalità di interazione uomo-macchina.

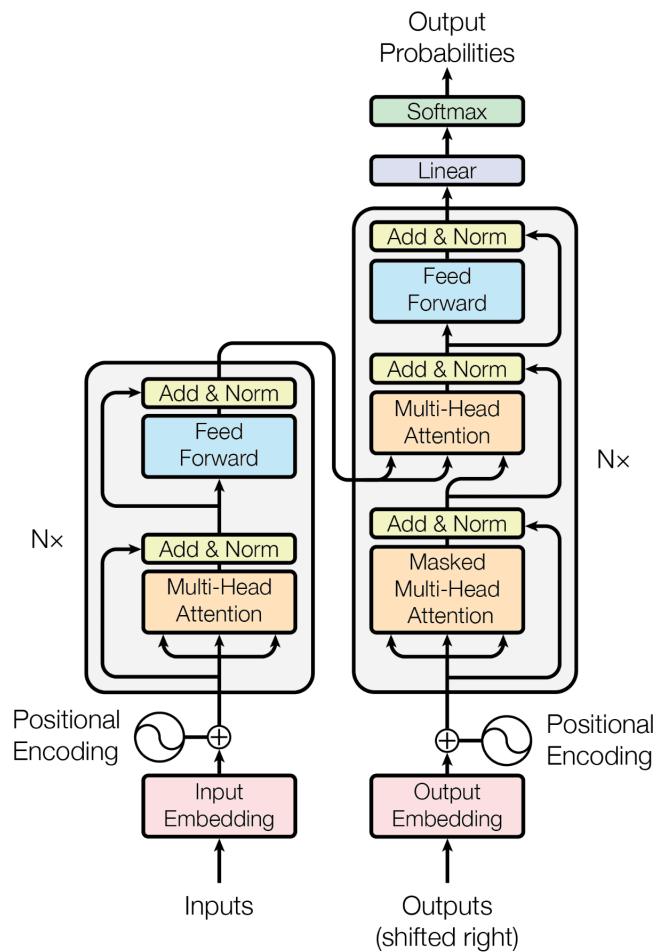
Nonostante i progressi, gli LLM affrontano ancora sfide tecniche come l'overfitting, la gestione di parole rare o mai viste e la capacità di rappresentare in modo efficace la struttura complessa del linguaggio. Miglioramenti nelle architetture e metodologie di addestramento sono essenziali per superare questi limiti e ottimizzare le prestazioni dei modelli di AI [41].

### 3.3.1 L'architettura Transformer

I meccanismi di Attention sono diventati essenziali nella modellazione delle sequenze, consentendo di catturare le dipendenze tra parole indipendentemente dalla loro distanza nel testo. Prima dell'introduzione dei *Transformer*, i modelli basati su *Reti Ricorrenti (RNN)*, come *LSTM* e *GRU*, erano ampiamente utilizzati per la gestione delle dipendenze temporali [64, 65]. Tuttavia, presentavano limiti significativi, tra

cui difficoltà di parallelizzazione e problemi nella gestione delle dipendenze a lungo termine [66].

Per superare queste limitazioni, è stata introdotta l'architettura Transformer, che sfrutta il meccanismo di *Self-Attention* per modellare le relazioni tra elementi di una sequenza in modo più efficiente, migliorando così l'elaborazione del linguaggio naturale rispetto ai modelli di Reti Ricorrenti.



**Figura 3.5:** Architettura Transformer ([1])

L'architettura Transformer è diventata la base dei moderni LLM grazie alla sua capacità di processare dati in parallelo e gestire grandi quantità di testo, rendendola ideale per applicazioni avanzate di Natural Language Processing (NLP) [67]. Come mostrato nella Figura 3.5, il Transformer adotta un'architettura Encoder-Decoder, eliminando la necessità delle Reti Ricorrenti e sfruttando il meccanismo di Attention per catturare il contesto globale [1].

L'Encoder è costituito da 6 layer identici, ciascuno composto da due sub-layer: Multi-Head Self-Attention e una rete Feed-Forward completamente connessa. Ogni sub-layer è avvolto da una connessione residuale e seguito da una normalizzazione layer-wise, garantendo stabilità e una migliore propagazione dei gradienti [51]. Tutti i sub-layer producono output di dimensione fissa 512.

Il Decoder ha una struttura simile all'Encoder, ma include un terzo sub-layer di Multi-Head Attention, che esegue un'operazione sull'output dell'Encoder. Inoltre, il masking applicato all'Attention del Decoder assicura che la generazione del testo sia autoregressiva, ossia che ogni token venga generato in base ai token precedenti.

Alla base del Transformer vi è il meccanismo di Attention, in particolare la Scaled Dot-Product Attention e la Multi-Head Attention. Il concetto di Attention può essere descritto come una funzione che associa una query a una serie di coppie chiave-valore, restituendo un'output basato su una somma pesata dei valori. Il peso assegnato a ciascun valore dipende dalla compatibilità tra la query e la chiave corrispondente. Questo processo consente ai Transformer di catturare in modo efficace relazioni semantiche complesse tra i token di una sequenza.

**Scaled Dot-Product Attention** La Scaled Dot-Product Attention riceve in input le query e le chiavi di dimensione  $d_k$ , e i valori di dimensione  $d_v$ . Il processo consiste nel calcolare i prodotti scalari tra le query e tutte le chiavi, dividere ciascun prodotto per  $\sqrt{d_k}$ , e applicare una funzione *Softmax* per ottenere i pesi associati ai valori:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3.3.1)$$

Questa forma di Attention è molto efficiente poiché può essere implementata tramite moltiplicazioni di matrici ottimizzate. A confronto, l'Additive Attention utilizza una rete Feed-Forward con un hidden layer per calcolare la compatibilità tra le parole. Sebbene entrambe le tecniche abbiano complessità teorica simile, la Scaled Dot-Product Attention è preferita per velocità ed efficienza in termini di memoria, specialmente per dimensioni più grandi di  $d_k$ .

**Multi-Head Attention** La Multi-Head Attention permette al modello di eseguire più operazioni di Attention in parallelo, proiettando query, chiavi e valori su diversi

sotto-spazi tramite trasformazioni lineari. Ogni testa di Attention lavora su una proiezione distinta, e i risultati vengono concatenati e proiettati nuovamente per ottenere l'output finale:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O \quad (3.3.2)$$

Dove ogni testa è definita come:

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (3.3.3)$$

con matrici di proiezione:

$$W_i^Q \in \mathbb{R}^{d_{\text{model}} \times d_k}, \quad W_i^K \in \mathbb{R}^{d_{\text{model}} \times d_k}, \quad W_i^V \in \mathbb{R}^{d_{\text{model}} \times d_v}, \quad W^O \in \mathbb{R}^{hd_v \times d_{\text{model}}} \quad (3.3.4)$$

Il vantaggio principale della Multi-Head Attention è la capacità di 'catturare relazioni multiple tra i token' all'interno di una sequenza, migliorando la comprensione contestuale e riducendo il rischio di perdita di informazioni cruciali.

Esistono tre applicazioni principali della **Multi-Head Attention** nel Transformer:

1. **Self-Attention nell'Encoder-Decoder:** In questo caso, le query provengono dal layer precedente del Decoder, mentre chiavi e valori derivano dall'output dell'Encoder. Questo meccanismo permette a ogni posizione nel Decoder di focalizzarsi su tutte le posizioni della sequenza di input, facilitando il trasferimento delle informazioni dall'input all'output.
2. **Self-Attention nell'Encoder:** Ogni posizione dell'Encoder può "attendere" su tutte le altre posizioni della sequenza di input, migliorando la capacità del modello di catturare relazioni semantiche e dipendenze contestuali. Qui, le query, le chiavi e i valori provengono dallo stesso layer dell'Encoder, consentendo una rappresentazione più ricca del testo.

3. **Self-Attention nel Decoder:** Simile al caso dell'Encoder, ma con una differenza chiave: ogni posizione può "attendere" solo sulle posizioni precedenti, evitando l'uso di informazioni future durante la generazione del testo. Questo è garantito attraverso un *masking*, che imposta a  $-\infty$  le connessioni non valide, preservando così la sequenzialità autoregressiva.

Grazie a questa struttura, i *Transformer* permettono di affrontare con efficienza compiti linguistici complessi come *traduzione automatica*, *riassunto* e *question answering*. La loro flessibilità e adattabilità ai diversi contesti linguistici favoriscono un'interazione sempre più naturale tra uomo e macchina, aprendo nuove frontiere nel *Natural Language Processing (NLP)* [68].

### 3.3.2 Fine-Tuning e Transfer Learning

Una delle caratteristiche fondamentali degli LLM è la loro capacità di adattarsi a diversi compiti tramite il Transfer Learning e il Fine-Tuning [69]. Dopo un pre-addestramento su vasti dataset generici, questi modelli possono essere adattati a compiti specifici con quantità limitate di dati.

Il Transfer Learning permette di riutilizzare un modello pre-addestrato per compiti correlati, migliorando l'efficienza senza dover ripartire da zero. Il Fine-Tuning, invece, consente di perfezionare il modello su un dataset più piccolo e specifico. In *Universal Language Model Fine-tuning (ULMFiT)* [70] ha dimostrato l'efficacia di questa tecnica, riducendo l'errore fino al 24% in compiti di classificazione testuale. Modelli come BERT [51], GPT [53], T5 [47] e Mistral-7B [49] hanno consolidato questo approccio, migliorando le prestazioni in NLP con un adattamento efficiente ai diversi contesti.

### 3.3.3 Le sfide degli LLM

L'adozione degli LLM pone diverse sfide, non solo in termini di complessità computazionale, ma anche di gestione delle risorse e sostenibilità.

- **Efficienza del Fine-tuning:** L'addestramento e l'adattamento degli LLM richiedono ingenti risorse computazionali. LoRA (Low-Rank Adaptation) [71] riduce



la quantità di parametri da ottimizzare, abbattendo il consumo di memoria GPU fino al 66% e riducendo lo spazio per i checkpoint da 350GB a 35MB.

- **Quantizzazione:** Tecnica che riduce la precisione numerica dei parametri (es. da 32 bit a 8 bit), diminuendo la memoria richiesta e migliorando l'efficienza computazionale [72]. Questo metodo consente l'inferenza su dispositivi con risorse limitate senza compromettere significativamente le prestazioni.
- **Impatto ambientale:** L'addestramento di modelli come GPT-3 (175 miliardi di parametri) ha richiesto un consumo energetico di 1.287 MWh, con un'impronta di carbonio pari a 502 tonnellate di CO<sub>2</sub> [73]. Ridurre il consumo energetico è una priorità nella ricerca sui modelli di nuova generazione.
- **Bias e interpretabilità:** Gli LLM tendono a riflettere i bias presenti nei dataset di addestramento, generando risposte potenzialmente discriminatorie. Inoltre, la loro scarsa interpretabilità ne limita l'uso in contesti critici come la medicina e la finanza.

Sebbene tecniche come *LoRA* e *Quantizzazione* abbiano migliorato l'efficienza degli LLM, restano sfide cruciali: sostenibilità energetica, adattabilità a contesti diversi e mitigazione dei bias. Affrontare questi problemi è essenziale per garantire un utilizzo sicuro, etico e responsabile dei modelli di linguaggio su larga scala.

## 3.4 Caso di studio: MindfulDiary

MindfulDiary è un sistema progettato per supportare i pazienti nel loro percorso psicoterapeutico attraverso l'integrazione di un *Large Language Model (LLM)* per facilitare la documentazione delle esperienze quotidiane. Il sistema è stato sviluppato in collaborazione con professionisti della salute mentale, con l'obiettivo di fornire un supporto continuo, anonimo e personalizzato ai pazienti, riducendo le barriere legate allo stigma sociale e migliorando il monitoraggio da parte del medico.

### 3.4.1 Obiettivi del sistema

MindfulDiary è stato sviluppato per:

- **Migliorare il monitoraggio del paziente** attraverso la registrazione e l'analisi delle sue esperienze quotidiane.
- **Fornire supporto psicologico** utilizzando un modello AI addestrato per riconoscere schemi emotivi.
- **Mantenere una connessione tra medico e paziente**, consentendo al professionista di monitorare il percorso terapeutico, senza la possibilità di intervenire quando necessario.
- **Eliminare lo stigma sociale**, offrendo uno strumento sempre a portata di mano di auto-monitoraggio accessibile e discreto.

### 3.4.2 Funzionamento del sistema

MindfulDiary si basa su un'interfaccia conversazionale, in cui il paziente può descrivere le proprie esperienze giornaliere e ricevere feedback di supporto generato dall'AI. Il funzionamento può essere suddiviso in tre fasi principali:

#### 1. Inserimento delle annotazioni giornaliere.

- Il paziente interagisce con l'applicazione tramite testo o voce, registrando le sue esperienze quotidiane.
- L'AI analizza il contenuto, identificandone le emozioni.

#### 2. Analisi e generazione di risposte.

- Il modello NLP genera un feedback di supporto basato sulle emozioni rilevate nel testo.

#### 3. Monitoraggio e iterazione con il medico.

- Il medico può accedere ai dati raccolti e visualizzare tendenze nel benessere del paziente.
- In base ai dati, il professionista può, dal vivo, personalizzare il percorso terapeutico e intervenire attivamente in caso di necessità.

### 3.4.3 Tecnologie utilizzate

In MindfulDiary sono state utilizzate le seguenti tecnologie:

- Modelli NLP avanzati, tra cui LLM come GPT-4, LLaMA e Mistral-7B, per la generazione di testo personalizzato.
- Tecniche di Fine-Tuning per adattare l'IA ai bisogni specifici della psicoterapia.
- Privacy e sicurezza dei dati, garantendo la conformità alle normative sulla protezione dei dati sanitari.

### 3.4.4 Risultati preliminari

Studi sperimentali hanno dimostrato che MindfulDiary può:

- Aumentare l'aderenza terapeutica fornendo un metodo intuitivo per la registrazione delle esperienze.
- Ridurre il senso di isolamento nei pazienti, offrendo un supporto costante.
- Migliorare l'efficacia delle sedute terapeutiche, permettendo al medico di avere un quadro più dettagliato della condizione del paziente tra una sessione e l'altra.

### 3.4.5 Limiti del sistema e sviluppi futuri

MindfulDiary rappresenta un passo significativo verso l'integrazione dell'AI nella psicoterapia, offrendo un supporto continuo ai pazienti e facilitando il lavoro dei terapeuti. Tuttavia, restano aperte alcune sfide, tra cui:

- Migliorare la comprensione contestuale dei modelli AI per garantire risposte più accurate.
- Rendere personalizzata la generazione da parte dei modelli AI.
- Ottimizzare il Fine-Tuning del modello per adattarsi meglio a esigenze individuali.
- Valutare l'impatto clinico a lungo termine attraverso studi su larga scala.

## CAPITOLO 4

---

### Application

---

*SoulDiaryConnect*[74] è un sistema progettato per supportare il paziente nel percorso psicoterapeutico. Si distingue da *MindfulDiary* per l'introduzione di una personalizzazione avanzata del modello AI da parte del medico. A differenza dei sistemi tradizionali di journaling assistito, dove il paziente riceve solo un feedback statico dall'AI, *SoulDiaryConnect* consente al medico di configurare e modificare attivamente i parametri di generazione del testo, garantendo risposte più mirate e coerenti con il percorso terapeutico. Inoltre, introduce un sistema di interazione diretta tra medico e paziente, superando il limite del monitoraggio passivo presente in *MindfulDiary*.

#### 4.1 Architettura del Sistema

L'architettura di *SoulDiaryConnect* segue uno stile simil-modulare, diviso in diversi componenti principali, basato sul modello *Model Template View* (MTV). Esso è un modello architetturale derivante dal modello *Model View Controller* (MVC). Nel caso del modello MTV la **view** viene sostituita dal **template**, basato sull'idea di *how you can see it*, e il **controller** dal **view**. Questo cambio di prospettiva, invece, si basa

sul considerare la view come *which data you see* e non *how you see it* <sup>1</sup> da parte degli sviluppatori di Django <sup>2</sup>, ovvero il framework utilizzato.

Django è un framework web ad alto livello di Python <sup>3</sup>. Si è scelto per la sua sicurezza, in quanto permette di evitare problematiche comuni come le *SQL injection*, i *cross-site scripting* e la gestione dell'autenticazione e dei dati persistenti. È inoltre estremamente scalabile, così da poter supportare grandi flussi di dati.

### 4.1.1 Panoramica del Sistema per l'utente

Il sistema si presenta composto da poche pagine user-friendly - in accordo con le linee guida di Badashian et al. [75] - le quali danno la possibilità all'utente di avere un'esperienza completa e guidata. Le tecnologie utilizzate per le pagine da visualizzare sono: HTML <sup>4</sup>, CSS3 <sup>5</sup> e JavaScript <sup>6</sup>.

Nel dettaglio, il sistema:

- Permette ai nuovi utenti di registrarsi come medico o paziente, facendo selezionare a quest'ultimo il proprio medico.
- Per chi è già registrato, mediante un'interfaccia utente unificata dà la possibilità di effettuare login ridirezionandolo alla sua area privata.

#### Area privata del paziente

Questa si presenta come un'area semplice ed intuitiva per il paziente, per il quale è possibile vedere, come si evince dalla Figura 4.1, i dati relativi al suo medico in alto a sinistra, lo storico delle note aggiunte al suo diario digitale ed i relativi commenti rispettivamente generato dall'intelligenza artificiale e dal medico a destra. Inoltre, in basso a sinistra dà la possibilità di aggiungere nuove note autobiografiche.

---

<sup>1</sup><https://docs.djangoproject.com/en/5.1/faq/general/#does-django-scale>

<sup>2</sup><https://www.djangoproject.com>

<sup>3</sup><https://www.python.org>

<sup>4</sup><https://html.spec.whatwg.org>

<sup>5</sup><https://www.w3.org/Style/CSS/Overview.en.html>

<sup>6</sup><https://developer.mozilla.org/en-US/docs/Web/JavaScript>

**Soul Diary**

**Benvenuto, Antonio Manghi** Logout

**Scheda del tuo medico**

**Nome:** Mario Rossi  
**Indirizzo:** Via Roma, 10 - Milano  
**Email:** mario.rossi@medico.com  
**Numero Studio:** 0245789632  
**Numero Cellulare:** 3487654321

**Aggiungi una nuova nota**

Scrivi la tua nota...

Aggiungi Nota

I'm in Lecce. In a hypothetical present, we could have been together.

**Supporto:** Response: "It sounds like you're going through a difficult time, trying to process your feelings and emotions. It's natural to have these thoughts and questions, especially when you're away from the person you care about. Remember that you deserve to enjoy your vacation and take care of yourself during this period. It's important to give yourself the space and time to heal and grow, even if it might feel challenging at times. Trust that you can find a way to move forward and create a positive future for yourself. Take things one day at a time and be gentle with yourself in the process."

**Data:** Feb. 4, 2025

**Nota:** While looking at different gym courses, I had a brief thought: I'll never meet anyone. And whenever I have these kinds of thoughts, I latch onto the idea of Elena. It seems like I'm repeating the same pattern—difficulty → comfort zone—just like I did a few months ago when things were going badly with Elena, and I would go back to the idea of Eleonora. Which, of course, leads to the question: Do I even like Elena?

**Supporto:** Response: It's normal to have these thoughts and feelings when facing new situations. Everyone has their unique journey and experiences. It's important to reflect on your feelings towards Elena. If you find yourself questioning your feelings and repeatedly returning to thoughts of Eleonora, it could be a sign that there may be some unresolved feelings or emotions. You could try to explore these feelings by talking to a friend, writing in a journal, or seeking professional help if needed. Remember, it's okay to feel uncertain, but take the time to understand yourself better.

**Data:** Feb. 4, 2025

**Nota:** 've been thinking about how my relationship with Elena is similar to a drug addiction, and honestly, it seems pretty accurate—the dynamic is the same. So many times, I've reached out to her even though I rationally knew it would hurt me. So many other times, I managed to resist the temptation to seek her out, but I wasn't strong enough when she was the one reaching out. Now, I have two problems: This is a drug that's constantly within reach—I can run into her even when I don't want to. There's a very real chance I'll go through withdrawal. I'm scared.

**Supporto:** Response: I understand how challenging it can be to navigate through a difficult relationship. It's essential to prioritize your well-being and establish healthy boundaries with Elena. Reach out to a trusted friend, family member, or professional to discuss your feelings and help you find ways to manage the situation. Remember, you're stronger than you think and deserve to be in a healthy relationship. Take it one day at a time and continue working on your self-care.

**Cosa ne pensa il medico:** Don't give up. Think about your relationship with Elena

Figura 4.1: Schermata dell'applicativo lato paziente

### Area privata del medico

Anch'essa si presenta come un'area semplice, come si evince in Figura 4.2, attraverso la quale il medico ha la possibilità di analizzare i diari dei suoi pazienti, selezionandoli nella colonna di sinistra e di personalizzare la generazione del testo da parte dell'AI cliccando il pulsante dedicato, riportandolo alla schermata in Figura 4.3. Il funzionamento di tale area verrà approfondito nella Sottosezione 4.1.3).

The screenshot shows the medical interface of the Soul Diary application. At the top, a blue header bar contains the 'Soul Diary' logo, the user name 'Benvenuto, Dr. Mario Rossi', and buttons for 'Personalizza Generazione' and 'Logout'. On the left, a sidebar titled 'Lista dei Pazienti' lists four patients: Anna Bianchi, Carlo Verdi, Alessandro Gialli, and Antonio Manghi, with Antonio Manghi selected. The main content area is titled 'Dati del Paziente' and displays information for Antonio Manghi, including his name, fiscal code, and date of birth. Below this, there is a section for a journal entry dated February 4, 2025. The entry includes a 'Nota' (note) about a relationship with Elena, a 'Supporto' (support) section with advice, and a 'Testo per il medico' (text for the doctor) section with a clinical analysis and emotional validation. A 'Cosa ne pensa il medico' (what does the doctor think) section is also present, with a text input field and a 'Salva' button.

Figura 4.2: Schermata dell'applicativo lato medico

The screenshot shows the 'Personalizza Generazione' (Personalize Generation) screen. The header bar includes the 'Soul Diary' logo, the title 'Personalizza Generazione', and buttons for 'Torna Indietro' and 'Logout'. The main content area is divided into two columns: 'Tipo di Nota' (Note Type) and 'Lunghezza della Nota' (Note Length). Under 'Tipo di Nota', there are two radio buttons: 'Commento Strutturato' (selected) and 'Commento Non Strutturato'. Under 'Lunghezza della Nota', there are two radio buttons: 'Commento Lungo' (selected) and 'Commento Breve'. Below these, a section titled 'Parametri Personalizzati' (Personalized Parameters) instructs the user to respond in a parametrized mode to a test. The test text is 'Today I failed my exam and feel like giving up.' Below this, there is a table with three rows, each representing a parameter. The first row is 'Clinical Analysis' with the text 'The individual expresses hopelessness and failure, likely reflecting cognitive distortions like catastrophizing.' The second row is 'Emotional Validation' with the text 'It's natural to feel upset after a setback. These feelings are valid and show your care for your goals.' The third row is 'Therapeutic Suggestions' with the text 'Reframe this as an opportunity to improve. Focus on actionable steps like revising your study plan.' Each row has a 'Rimuovi' (Remove) button. At the bottom, there are two buttons: 'Aggiungi Parametro' (Add Parameter) and 'Conferma' (Confirm).

Figura 4.3: Schermata per la personalizzazione della generazione del testo

### 4.1.2 Gestione ed Elaborazione Dati (Backend)

Nella parte di backend, il sistema coordina la gestione dei dati persistenti e la comunicazione tra questi ultimi, l'utente e il modulo di AI. Per la gestione dei dati persistenti si è scelto di utilizzare un database di tipo relazionale in quanto, tramite le proprietà ACID vengono garantiti:

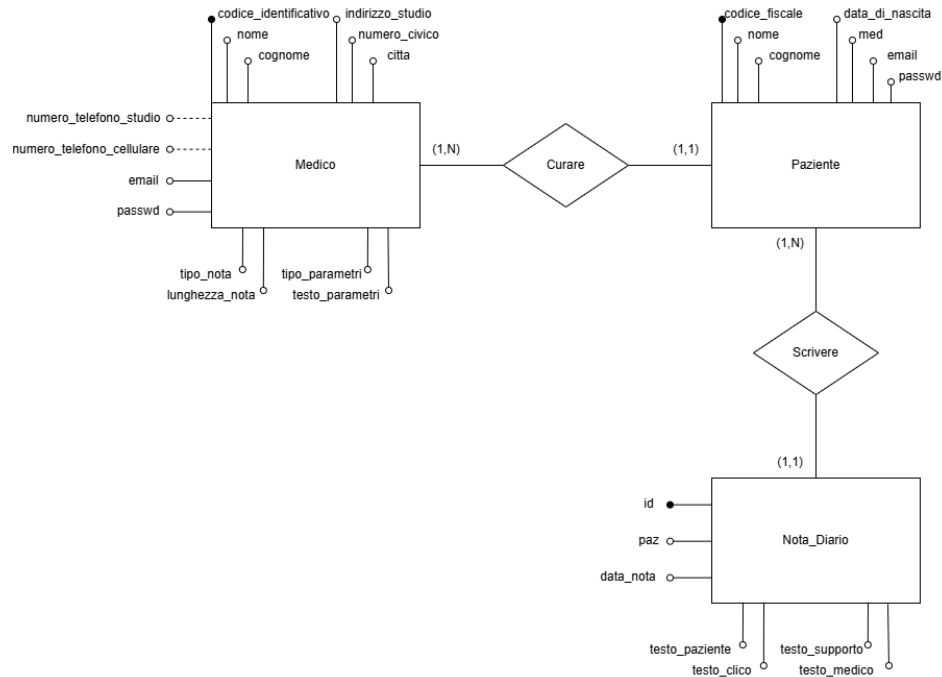
- **Atomicità delle operazioni:** Permette di effettuare transazioni atomiche, ciò vuol dire che la transazione deve essere eseguita interamente o non essere eseguita affatto.
- **Coerenza dei dati:** Permette di mantenere una coerenza tra i dati, in vista anche dell'utilizzo del sistema su diverse architetture hardware.
- **Isolamento delle operazioni:** Permette ad ogni utente di eseguire transazioni isolate assicurando che queste non interferiscano tra di loro.
- **Durabilità:** Le modifiche apportate ai dati tramite transazioni eseguite con successo vengono salvate anche in caso di guasto.

Le proprietà elencate assicurano il massimo livello di affidabilità e integrità dei dati, garantendo che essi non si trovino mai in una condizione di incoerenza. Inoltre, questa scelta permette di:

- Avere un modello semplice dei dati.
- Facilitare l'evoluzione del progetto tramite la separazione tra la logica del database e la struttura del database stesso (di cui è riportato in Figura 4.4 il Diagramma Entità-Relazione).
- Avere un'alta sicurezza, tramite l'accesso protetto ai dati.

A tal proposito, si è deciso di utilizzare PostgreSQL, un DBMS open source con alti livelli di resilienza, integrità e correttezza.





**Figura 4.4:** Diagramma Entità-Relazione del Sistema

### 4.1.3 Modulo di NLP

La componente di Natural Language Processing di SoulDiaryConnect rappresenta il cuore del sistema, gestendo l'elaborazione del linguaggio naturale e la generazione di risposte personalizzate. Essa si basa su Mistral-7B [49], caricato attraverso la libreria `llama_cpp`<sup>7</sup> per un'inferenza locale ottimizzata. Il sistema prevede due livelli di generazione del testo:

- **Feedback di supporto emotivo**, basato su un prompt ingegnerizzato per fornire messaggi motivazionali e rassicuranti.
- **Analisi clinica personalizzata**, generata sulla base delle impostazioni definite dal medico.

Il flusso di elaborazione può essere suddiviso nelle seguenti fasi:

1. **Ricezione dell'input**, ovvero quando il paziente inserisce una nota nel proprio diario, che viene inviata al backend.

<sup>7</sup><https://github.com/ggerganov/llama.cpp>

2. **Generazione del feedback di supporto:** Il sistema richiama la funzione `genera_frase_di_supporto(testo)`, che invia il testo a Mistral-7B con un prompt predefinito per ottenere una risposta empatica e motivazionale. L'output viene restituito all'utente come parte del diario.

```

1  from llama_cpp import Llama
2
3  model_path = "/path/to/mistral-7b-openorca.Q8_0.gguf"
4  llama_model = Llama(model_path=model_path, n_ctx=2048)
5
6  ...
7
8  def genera_frase_di_supporto(testo):
9      try:
10         prompt = f"""
11         You are a supportive assistant. Use the
12         following example to craft your response.
13
14         Example:
15         Text: "I failed my exam and feel like giving up."
16         Response: "I'm so sorry to hear about your exam.
17         It's okay to feel disappointed, but this doesn't define
18         your worth. Consider revising your study strategy and
19         asking for help. You've got this!"
20
21         Now, respond to the following text:
22         {testo}
23         """
24
25         result = llama_model(prompt, max_tokens=150)
26         return result['choices'][0]['text'].strip()
27     except Exception as e:
28         return f"Errore durante la generazione: {e}"

```

3. **Generazione del commento clinico personalizzato:** Il sistema identifica il medico curante del paziente e ne estrae le impostazioni personalizzate. In base alla configurazione salvata dal medico (`tipo_nota`, `lunghezza_nota`, `tipo_parametri`, `testo_parametri`), viene costruito un prompt specifico per `genera_frase_cliniche(testo, medico)`. Il modello AI genera una risposta strutturata o libera, rispettando le direttive personalizzate dal medico. Il risultato viene salvato come testo clinico generato, visibile nel pannello medico.

```

1 def genera_frase_cliniche(testo, medico):
2     try:
3
4         tipo_nota = medico.tipo_nota
5         # True per "strutturato", False per "non strutturato"
6         lunghezza_nota = medico.lunghezza_nota
7         # True per "lungo", False per "breve"
8         tipo_parametri =
9             medico.tipo_parametri.split(".;!")
10            if medico.tipo_parametri else []
11         testo_parametri =
12             medico.testo_parametri.split(".;!")
13            if medico.testo_parametri else []
14
15         # Determina il max_tokens in base alla lunghezza_nota
16         max_tokens = 250 if lunghezza_nota else 150
17
18         if tipo_nota:
19             # Genera il prompt strutturato con parametri
20             parametri_strutturati = "\n".join(
21                 [f"{tipo}: {testo}" for tipo,
22                  testo in zip(tipo_parametri, testo_parametri)]
23             )
24
25             prompt = f"""
26                 You are a psychotherapist specializing in CBT.
27                 Analyze the following text and provide a clinical
28                 assessment.
29
30                 Example:
31                 Text: "Today I failed my exam and
32                 feel like giving up."
33                 Response:
34                 {parametri_strutturati}
35
36                 Parameters:
37                 {tipo_parametri}
38
39                 Now analyze this text:
40                 {testo}
41
42                 Respond in the format of the example response:
43                 """
44         else:
45             # Genera il prompt non strutturato
46             prompt = f"""
47                 You are a psychotherapist specializing in CBT.

```

```

48         Analyze the following text and provide a clinical
49         assessment. The text is: {testo}
50         """
51
52         # Genera il risultato usando il modello
53         result = llama_model(prompt, max_tokens=max_tokens)
54         return result['choices'][0]['text'].strip()

```

4. **Monitoraggio e Interazione Medico-Paziente** Il medico può accedere alle annotazioni del paziente attraverso la dashboard. Se necessario, può modificare o integrare manualmente il commento clinico, grazie alla funzione **modifica\_testo\_medico(request, nota\_id)**.

```

1  def modifica_testo_medico(request, nota_id):
2      if request.method == 'POST':
3          nota = get_object_or_404(NotaDiario, id=nota_id)
4          testo_medico = request.POST.get('testo_medico', '').strip()
5          nota.testo_medico = testo_medico
6          nota.save()
7          return redirect(f'/medico/home/?paziente_id=
8                          {nota.paz.codice_fiscale}')

```

Il sistema permette inoltre la personalizzazione dinamica dell'AI attraverso **personalizza\_generazione(request)**, consentendo al medico di modificare i parametri delle risposte AI in base all'evoluzione del paziente.

```

1  def personalizza_generazione(request):
2      if request.session.get('user_type') != 'medico':
3          return redirect('/login/')
4
5      medico_id = request.session.get('user_id')
6      medico = Medico.objects.get(codice_identificativo=medico_id)
7
8      if request.method == 'POST':
9          # Tipo di Nota
10         tipo_nota = request.POST.get('tipo_nota')
11         medico.tipo_nota =
12             True if tipo_nota == 'strutturato' else False
13
14         # Lunghezza della Nota

```

```

15     lunghezza_nota = request.POST.get('lunghezza_nota')
16     medico.lunghezza_nota =
17         True if lunghezza_nota == 'lungo' else False
18
19     # Concatenazione di tipo_parametri e testo_parametri
20     tipo_parametri = request.POST.getlist('tipo_parametri')
21     testo_parametri = request.POST.getlist('testo_parametri')
22     medico.tipo_parametri = ";;;".join(tipo_parametri)
23     medico.testo_parametri = ";;;".join(testo_parametri)
24
25     medico.save()
26     return redirect('medico_home')
27
28     # Suddivide i parametri già salvati in liste
29     # per visualizzarli nella tabella
30     tipo_parametri = medico.tipo_parametri.split(";;;")
31     if medico.tipo_parametri else []
32     testo_parametri = medico.testo_parametri.split(";;;")
33     if medico.testo_parametri else []
34
35     return render(request, 'SoulDiaryConnectApp/'
36         + 'personalizza_generazione.html', {
37         'medico': medico,
38         'tipo_parametri': zip(tipo_parametri, testo_parametri),
39     })

```

Per garantire efficienza, sicurezza e flessibilità, il sistema AI di SoulDiaryConnect impiega:

- **Modelli di Natural Language Processing (NLP):**
  - Mistral-7B (Quantizzato ad 8 bit in GGUF) per un’inferenza locale veloce ed efficiente.
  - llama\_cpp per il caricamento e la gestione del modello.
- **Contestualizzazione:**
  - Prompt Engineering avanzato per ottimizzare la generazione delle risposte.
  - Configurazione personalizzabile dei parametri dell’AI, salvata direttamente nel profilo del medico.

#### 5.1 Analisi

L'integrazione dell'Intelligenza Artificiale (AI) nella salute mentale rappresenta un'innovazione significativa, con il potenziale di migliorare l'accessibilità, il monitoraggio e il supporto terapeutico. Sistemi come MindfulDiary e quello proposto, SoulDiaryConnect, dimostrano come gli LLM (Large Language Models) ed in particolare gli NLP possano essere utilizzati per facilitare il journaling terapeutico, migliorando la qualità e la quantità delle annotazioni dei pazienti e fornendo ai medici una visione più dettagliata del loro stato emotivo e cognitivo.

Tuttavia, l'uso dell'AI nella salute mentale presenta sfide e limitazioni. Gli attuali modelli NLP, sebbene avanzati, non possiedono empatia reale e possono generare risposte prive di un reale giudizio clinico contestualizzato. Inoltre, le problematiche legate alla privacy, al bias algoritmico e alla personalizzazione restano aperte. È fondamentale garantire che i sistemi AI non sostituiscano il terapeuta umano, ma piuttosto lo supportino nel fornire un trattamento più mirato. Come evidenziato da Fusar-Poli et al. (2022)[76], la supervisione umana rimane essenziale per preservare il valore del rapporto terapeutico, evitando una medicalizzazione eccessiva dell'interazione con l'AI.

### 5.1.1 Valutazioni di un esperto del dominio

A seguito dello sviluppo di SoulDiaryConnect, le prestazioni del sistema sono state valutate da un esperto nel settore della salute mentale, il quale ha fornito un'analisi critica dell'integrazione dell'AI nei processi di journaling terapeutico. Tra gli aspetti positivi, è stato evidenziato il valore del supporto che tali strumenti possono offrire al terapeuta, in particolare attraverso l'analisi automatizzata del testo. L'identificazione di parole chiave e temi ricorrenti può infatti agevolare la comprensione dello stato emotivo del paziente, fornendo al clinico una base informativa più ampia e dettagliata.

Tuttavia, emergono alcune considerazioni di rilievo per garantire un utilizzo efficace e sicuro di questi sistemi. In primo luogo, il rinforzo fornito dall'AI al paziente può essere utile, rimanendo però su un piano generale. Un supporto troppo specifico potrebbe risultare problematico per alcuni pazienti, come coloro con disturbi ossessivi, i quali potrebbero non essere in grado di gestire un'informazione nuova senza un adeguato contesto terapeutico. La relazione con il terapeuta rimane quindi, come garantito da SoulDiaryConnect, un elemento imprescindibile per l'assimilazione e l'elaborazione di nuovi messaggi.

Infine, l'esperto ha sottolineato alcune criticità da considerare nell'adozione di queste tecnologie. Tra i principali rischi vi è la possibile razionalizzazione eccessiva da parte del paziente, che potrebbe focalizzarsi su un'elaborazione puramente logica delle proprie emozioni, trascurando il coinvolgimento affettivo. Inoltre, l'assenza di comunicazione non verbale rappresenta un limite significativo, dato che molte informazioni clinicamente rilevanti emergono attraverso il linguaggio del corpo e l'intonazione. Infine, se non gestito correttamente sia dal paziente che dal terapeuta, il diario digitale potrebbe diventare uno strumento statico, perdendo la sua funzione dinamica di supporto all'esplorazione e alla crescita personale.

Queste valutazioni suggeriscono che, affinché strumenti come SoulDiaryConnect siano realmente efficaci, sia necessaria una progettazione attenta che tenga conto di tali implicazioni, integrando e confermando l'AI come supporto e non come sostituto dell'intervento umano.

## 5.2 Sviluppi Futuri

Per migliorare l'integrazione dell'AI nella psicoterapia, è necessario un approccio multidisciplinare, coinvolgendo esperti di psichiatria, etica e ingegneria AI. Alcuni sviluppi futuri chiave per SoulDiaryConnect includono:

- **Valutazione clinica a larga scala:** Studi su campioni più ampi e periodi prolungati per valutare l'efficacia del Sistema nel supporto terapeutico.
- **Maggior controllo sulla generazione del testo:** Implementare algoritmi di controllo etico per prevenire risposte AI potenzialmente dannose o inappropriate.
- **Sviluppo di una chat con un chatbot e una chat con il medico:** Implementare tali chat permetterebbero al paziente di avere un canale unico dove caricare le note autobiografiche e comunicare direttamente con il medico.
- **Utilizzo del modello non quantizzato o di modelli più performanti.**
- **Miglioramento della personalizzazione lato medico:** Potenziare il controllo medico sulle risposte, affinché i modelli possano adattarsi dinamicamente alle esigenze specifiche di ogni paziente.
- **Valutazione del bias ed eventuale mitigazione:** Implementare dataset più rappresentativi, tecniche di fine-tuning mirate e strategie di moderazione delle risposte AI.
- **Integrazione di segnali multimodali:** Oltre ai dati testuali, l'integrazione di altri tipi di segnali del paziente, come tono della voce, espressioni facciali e/o parametri fisiologici potrebbe migliorare l'accuratezza del supporto terapeutico fornito dall'AI.

## 5.3 Conclusioni Finali

L'AI sta rapidamente emergendo come strumento complementare alla psicoterapia tradizionale, offrendo un supporto accessibile e scalabile in un contesto in cui la domanda di servizi di salute mentale supera l'offerta. Tuttavia, per garantire



un'implementazione etica ed efficace, è essenziale adottare un approccio prudente e basato sull'evidenza.

SoulDiaryConnect rappresenta un esempio concreto di come l'AI possa essere personalizzata e controllata dal medico, riducendo il rischio di risposte standardizzate, come presente in sistemi come MindfulDiary, e aumentando la rilevanza clinica dell'intervento.

A differenza di MindfulDiary e di altri sistemi di journaling AI, SoulDiaryConnect introduce una svolta nel supporto terapeutico digitale: il medico diventa parte attiva nel processo, garantendo che i feedback AI siano in linea con il percorso terapeutico personalizzato del paziente. Questa innovazione migliora il monitoraggio clinico, riduce il rischio di risposte generiche e crea un sistema in cui AI e professionista collaborano per offrire un'assistenza più efficace e sicura.

Tuttavia, l'uso dell'AI in ambito terapeutico deve essere guidato da principi di trasparenza, equità e supervisione umana, per evitare di compromettere l'integrità del rapporto tra paziente e terapeuta.

In definitiva, l'AI non deve sostituire la figura del terapeuta, ma piuttosto estendere e potenziare il suo lavoro, facilitando un monitoraggio continuo e rendendo più accessibili le risorse di supporto psicologico. Solo attraverso un'integrazione responsabile, lo sfruttamento di queste tecnologie potrà essere considerato come una strada da prendere e accettare, e non qualcosa da temere e dal quale scappare.

---

## Bibliografia

---

- [1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2023. [Online]. Available: <https://arxiv.org/abs/1706.03762> (Citato alle pagine iii, 20 e 22)
- [2] C. Travers, "Unveiling a reflective diary methodology for exploring the lived experiences of stress and coping," *Journal of Vocational Behavior*, vol. 79, no. 1, pp. 204–216, 2011. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0001879110001934> (Citato a pagina 1)
- [3] P. M. Ullrich and S. K. Lutgendorf, "Journaling about stressful events: Effects of cognitive processing and emotional expression," *Annals of Behavioral Medicine*, vol. 24, no. 3, pp. 244–250, Aug 2002. [Online]. Available: [https://doi.org/10.1207/S15324796ABM2403\\_10](https://doi.org/10.1207/S15324796ABM2403_10) (Citato a pagina 1)
- [4] J. W. Pennebaker, "Writing about emotional experiences as a therapeutic process," *Psychological Science*, vol. 8, no. 3, pp. 162–166, 1997. [Online]. Available: <https://doi.org/10.1111/j.1467-9280.1997.tb00403.x> (Citato a pagina 1)
- [5] J. W. Pennebaker and C. K. Chung, "Expressive writing: Connections to physical and mental health," in *The Oxford Handbook of Health Psychology*. Oxford University Press, 08 2011. [Online]. Available: <https://doi.org/10.1093/oxfordhb/9780195342819.013.0018> (Citato a pagina 1)

- [6] M. C. Figueiredo and Y. Chen, "Patient-generated health data: Dimensions, challenges, and open questions," *Foundations and Trends® in Human-Computer Interaction*, vol. 13, no. 3, pp. 165–297, 2020. [Online]. Available: <http://dx.doi.org/10.1561/11000000080> (Citato a pagina 1)
- [7] D. S. Weiss, "Structured clinical interview techniques: An overview," *Journal of Clinical Psychology*, vol. 55, no. 10, pp. 1243–1261, 1999. (Citato a pagina 1)
- [8] G. Cameron, D. Cameron, G. Megaw, R. Bond, M. Mulvenna, S. O'Neill, C. Armour, and M. McTear, "Assessing the usability of a chatbot for mental health care," in *Internet Science - INSCI 2018 International Workshops, Revised Selected Papers*, ser. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), A. Følstad, H. Halpin, H. Niedermayer, S. Bodrunova, A. Smoliarova, O. Koltsova, P. Kolaridi, and L. Yuldashev, Eds. Germany: Springer Verlag, Jan. 2019, pp. 121–132, 5th International Conference on Internet Science, INSCI 2018 ; Conference date: 24-10-2018 Through 26-10-2018. (Citato alle pagine 1 e 2)
- [9] J. Howard, "Artificial intelligence: Implications for the future of work," *American journal of industrial medicine*, vol. 62, no. 11, pp. 917–926, 2019. (Citato a pagina 1)
- [10] C. Morosan, "The role of artificial intelligence in decision-making," *Procedia Computer Science*, vol. 91, pp. 1036–1042, 2016. (Citato a pagina 1)
- [11] E. E. Lee, J. Torous, M. De Choudhury, C. A. Depp, S. A. Graham, H.-C. Kim, M. P. Paulus, J. H. Krystal, and D. V. Jeste, "Artificial intelligence for mental health care: clinical applications, barriers, facilitators, and artificial wisdom," *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, vol. 6, no. 9, pp. 856–864, 2021. (Citato a pagina 1)
- [12] G. Espejo, W. Reiner, and M. Wenzinger, "Exploring the role of artificial intelligence in mental healthcare: progress, pitfalls, and promises," *Cureus*, vol. 15, no. 9, 2023. (Citato a pagina 1)
- [13] M. L. Wainberg, P. Scorza, J. M. Shultz, L. Helpman, J. J. Mootz, K. A. Johnson, Y. Neria, J.-M. E. Bradford, M. A. Oquendo, and M. R. Arbuckle, "Challenges

- and opportunities in global mental health: a research-to-practice perspective," *Current psychiatry reports*, vol. 19, pp. 1–10, 2017. (Citato a pagina 2)
- [14] X. Qin and C.-R. Hsieh, "Understanding and addressing the treatment gap in mental healthcare: economic perspectives and evidence from china," *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, vol. 57, p. 0046958020950566, 2020. (Citato a pagina 2)
- [15] H. Liu, H. Peng, X. Song, C. Xu, and M. Zhang, "Using ai chatbots to provide self-help depression interventions for university students: A randomized trial of effectiveness," *Internet Interventions*, vol. 27, p. 100495, 2022. (Citato a pagina 2)
- [16] P. Nilsen, P. Svedberg, J. Nygren, M. Frideros, J. Johansson, and S. Schueller, "Accelerating the impact of artificial intelligence in mental healthcare through implementation science," *Implementation research and practice*, vol. 3, p. 26334895221112033, 2022. (Citato a pagina 2)
- [17] S. Bae, D. Kwak, S. Kim, D. Ham, S. Kang, S.-W. Lee, and W. Park, "Building a role specified open-domain dialogue system leveraging large-scale language models," 2022. [Online]. Available: <https://arxiv.org/abs/2205.00176> (Citato a pagina 2)
- [18] F. Farhat, "Chatgpt as a complementary mental health resource: A boon or a bane," 07 2023. (Citato a pagina 2)
- [19] G. A. Coppersmith, C. Harman, and M. Dredze, "Measuring post traumatic stress disorder in twitter," *Proceedings of the International AAAI Conference on Web and Social Media*, 2014. [Online]. Available: <https://api.semanticscholar.org/CorpusID:14612598> (Citato a pagina 3)
- [20] K. Singhal, T. Tu, J. Gottweis, R. Sayres, E. Wulczyn, L. Hou, K. Clark, S. Pfohl, H. Cole-Lewis, D. Neal, M. Schaeckermann, A. Wang, M. Amin, S. Lachgar, P. Mansfield, S. Prakash, B. Green, E. Dominowska, B. A. y Arcas, N. Tomasev, Y. Liu, R. Wong, C. Semturs, S. S. Mahdavi, J. Barral, D. Webster, G. S. Corrado, Y. Matias, S. Azizi, A. Karthikesalingam, and V. Natarajan, "Towards

- expert-level medical question answering with large language models,” 2023. [Online]. Available: <https://arxiv.org/abs/2305.09617> (Citato a pagina 3)
- [21] Z. Jiang, S. I. Levitan, J. Zomick, and J. Hirschberg, “Detection of mental health from Reddit via deep contextualized representations,” in *Proceedings of the 11th International Workshop on Health Text Mining and Information Analysis*, E. Holderness, A. Jimeno Yepes, A. Lavelli, A.-L. Minard, J. Pustejovsky, and F. Rinaldi, Eds. Online: Association for Computational Linguistics, Nov. 2020, pp. 147–156. [Online]. Available: <https://aclanthology.org/2020.louhi-1.16/> (Citato a pagina 3)
- [22] K. Yang, S. Ji, T. Zhang, Q. Xie, and S. Ananiadou, “On the evaluations of chatgpt and emotion-enhanced prompting for mental health analysis,” 04 2023. (Citato a pagina 4)
- [23] K. Yang, T. Zhang, Z. Kuang, Q. Xie, J. Huang, and S. Ananiadou, “Mentallama: Interpretable mental health analysis on social media with large language models,” in *Proceedings of the ACM Web Conference 2024*, ser. WWW ’24. ACM, May 2024, p. 4489–4500. [Online]. Available: <http://dx.doi.org/10.1145/3589334.3648137> (Citato a pagina 4)
- [24] Q. Guo, J. Tang, W. Sun, H. Tang, Y. Shang, and W. Wang, “Soullmate: An adaptive llm-driven system for advanced mental health support and assessment, based on a systematic application survey,” 2024. [Online]. Available: <https://arxiv.org/abs/2410.11859> (Citato a pagina 4)
- [25] S. Ji, X. Zheng, J. Sun, R. Chen, W. Gao, and M. Srivastava, “Mindguard: Towards accessible and stigma-free mental health first aid via edge llm,” 2024. [Online]. Available: <https://arxiv.org/abs/2409.10064> (Citato a pagina 4)
- [26] T. Kim, S. Bae, H. A. Kim, S.-W. Lee, H. Hong, C. Yang, and Y.-H. Kim, “Mindfuldiary: Harnessing large language model to support psychiatric patients’ journaling,” in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, ser. CHI ’24. ACM, May 2024, p. 1–20. [Online]. Available: <http://dx.doi.org/10.1145/3613904.3642937> (Citato a pagina 4)

- [27] A. Pharma. (2023) Rapporto headway sulla salute mentale: presentati alla camera i dati italiani. Accessed: 2025-01-31. [Online]. Available: <https://www.angelinipharma.it/media/comunicati-stampa/rapporto-headway-sulla-salute-mentale-presentati-alla-camera-i-dati-italiani/> (Citato a pagina 5)
- [28] S. Abuse and M. H. S. Administration. (2022) 2022 national survey on drug use and health (nsduh) releases. Accessed: 2025-01-31. [Online]. Available: <https://www.samhsa.gov/data/release/2022-national-survey-drug-use-and-health-nsduh-releases> (Citato a pagina 5)
- [29] R. C. Bland, S. C. Newman, and H. Orn, "Help-seeking for psychiatric disorders," *The Canadian Journal of Psychiatry*, vol. 42, no. 9, pp. 935–942, 1997, pMID: 9429063. [Online]. Available: <https://doi.org/10.1177/070674379704200904> (Citato alle pagine 5 e 6)
- [30] S. Galderisi, A. Heinz, M. Kastrup, J. Beezhold, and N. Sartorius, "A proposed new definition of mental health," *Psychiatria Polska*, vol. 51, pp. 407–411, 06 2017. (Citato a pagina 6)
- [31] M. A. Chiariello and H. Orvaschel, "Patterns of parent-child communication: Relationship to depression," *Clinical Psychology Review*, vol. 15, no. 5, pp. 395–407, 1995, the Impact of the Family on Child Adjustment and Psychopathology. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/027273589500022H> (Citato a pagina 6)
- [32] D. J. Kiesler, "An interpersonal communication analysis of relationship in psychotherapy," *Psychiatry*, vol. 42, no. 4, pp. 299–311, 1979, pMID: 504511. [Online]. Available: <https://doi.org/10.1080/00332747.1979.11024034> (Citato a pagina 6)
- [33] J. A. Naslund, A. Bondre, J. Torous, and K. A. Aschbrenner, "Social media and mental health: Benefits, risks, and opportunities for research and practice," *Journal of Technology in Behavioral Science*, vol. 5, no. 3, pp. 245–257, Sep

2020. [Online]. Available: <https://doi.org/10.1007/s41347-020-00134-x> (Citato a pagina 7)
- [34] A. A. Abd-alrazaq, M. Alajlani, N. Ali, K. Denecke, B. M. Bewick, and M. S. Househ, "Perceptions and opinions of patients about mental health chatbots: Scoping review," *Journal of Medical Internet Research*, vol. 23, 2020. [Online]. Available: <https://api.semanticscholar.org/CorpusID:226161018> (Citato a pagina 7)
- [35] M. Neary and S. M. Schueller, "State of the field of mental health apps," *Cognitive and Behavioral Practice*, vol. 25, no. 4, pp. 531–537, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1077722918300233> (Citato a pagina 7)
- [36] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *SN Computer Science*, vol. 2, no. 3, p. 160, Mar 2021. [Online]. Available: <https://doi.org/10.1007/s42979-021-00592-x> (Citato a pagina 7)
- [37] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, ser. The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, 2011. [Online]. Available: <https://books.google.it/books?id=pQws07tdpjoC> (Citato a pagina 7)
- [38] K. P. Murphy, *Machine learning: a probabilistic perspective*, Cambridge, MA, 2012. (Citato a pagina 8)
- [39] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>. (Citato a pagina 10)
- [40] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015. [Online]. Available: <https://doi.org/10.1038/nature14539> (Citato alle pagine 11 e 12)
- [41] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter,

- C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, "Language models are few-shot learners," *CoRR*, vol. abs/2005.14165, 2020. [Online]. Available: <https://arxiv.org/abs/2005.14165> (Citato alle pagine 16, 17, 18, 19, 20 e 21)
- [42] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge University Press, 2008. (Citato alle pagine 16 e 19)
- [43] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah, M. Johnson, X. Liu, Łukasz Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, K. Stevens, G. Kurian, N. Patil, W. Wang, C. Young, J. Smith, J. Riesa, A. Rudnick, O. Vinyals, G. Corrado, M. Hughes, and J. Dean, "Google's neural machine translation system: Bridging the gap between human and machine translation," 2016. [Online]. Available: <https://arxiv.org/abs/1609.08144> (Citato a pagina 17)
- [44] D. Perez-Marin and I. Pascual-Nieto, *Conversational Agents and Natural Language Interaction: Techniques and Effective Practices: Techniques and Effective Practices*, ser. Premier reference source. Information Science Reference, 2011. [Online]. Available: <https://books.google.it/books?id=2nUcqtbcOBcC> (Citato a pagina 17)
- [45] L. Reynolds and K. McDonell, "Prompt programming for large language models: Beyond the few-shot paradigm," 2021. [Online]. Available: <https://arxiv.org/abs/2102.07350> (Citato a pagina 18)
- [46] R. Zellers, A. Holtzman, H. Rashkin, Y. Bisk, A. Farhadi, F. Roesner, and Y. Choi, "Defending against neural fake news," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019. [Online]. Available: [https://proceedings.neurips.cc/paper\\_files/paper/2019/file/3e9f0fc9b2f89e043bc6233994dfcf76-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2019/file/3e9f0fc9b2f89e043bc6233994dfcf76-Paper.pdf) (Citato a pagina 18)
- [47] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text



- transformer,” 2023. [Online]. Available: <https://arxiv.org/abs/1910.10683> (Citato alle pagine 18, 19 e 25)
- [48] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, “BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault, Eds. Online: Association for Computational Linguistics, Jul. 2020, pp. 7871–7880. [Online]. Available: <https://aclanthology.org/2020.acl-main.703> (Citato alle pagine 18 e 19)
- [49] A. Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. de las Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier, L. R. Lavaud, M.-A. Lachaux, P. Stock, T. L. Scao, T. Lavril, T. Wang, T. Lacroix, and W. E. Sayed, “Mistral 7b,” 2023. [Online]. Available: <https://arxiv.org/abs/2310.06825> (Citato alle pagine 18, 25 e 34)
- [50] J. Pennington, R. Socher, and C. Manning, “GloVe: Global vectors for word representation,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, A. Moschitti, B. Pang, and W. Daelemans, Eds. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1532–1543. [Online]. Available: <https://aclanthology.org/D14-1162> (Citato a pagina 19)
- [51] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding.” (Citato alle pagine 19 e 25)
- [52] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, “Roberta: A robustly optimized bert pretraining approach,” 2019. [Online]. Available: <https://arxiv.org/abs/1907.11692> (Citato a pagina 19)

- [53] A. Radford and K. Narasimhan, “Improving language understanding by generative pre-training,” 2018. [Online]. Available: <https://api.semanticscholar.org/CorpusID:49313245> (Citato alle pagine 19 e 25)
- [54] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*, 3rd ed., 2024, online manuscript released August 20, 2024. [Online]. Available: <https://web.stanford.edu/~jurafsky/slp3/> (Citato a pagina 20)
- [55] Y. Tay, M. Dehghani, D. Bahri, and D. Metzler, “Efficient transformers: A survey,” 2022. [Online]. Available: <https://arxiv.org/abs/2009.06732> (Citato a pagina 20)
- [56] T. Bolukbasi, K.-W. Chang, J. Zou, V. Saligrama, and A. Kalai, “Man is to computer programmer as woman is to homemaker? debiasing word embeddings,” 2016. [Online]. Available: <https://arxiv.org/abs/1607.06520> (Citato a pagina 20)
- [57] T. Sun, A. Gaut, S. Tang, Y. Huang, M. ElSherief, J. Zhao, D. Mirza, E. Belding, K.-W. Chang, and W. Y. Wang, “Mitigating gender bias in natural language processing: Literature review,” in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, A. Korhonen, D. Traum, and L. Màrquez, Eds. Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 1630–1640. [Online]. Available: <https://aclanthology.org/P19-1159> (Citato a pagina 20)
- [58] S. Ruder, “Neural transfer learning for natural language processing,” *PhD Thesis, National University of Ireland, Galway*, 2019. [Online]. Available: <https://researchrepository.universityofgalway.ie/server/api/core/bitstreams/db70a7a4-836b-4161-9269-e979efdd01ef/content> (Citato a pagina 20)
- [59] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, D. Pedreschi, and F. Giannotti, “A survey of methods for explaining black box models,” 2018. [Online]. Available: <https://arxiv.org/abs/1802.01933> (Citato a pagina 20)

- [60] A. B. Arrieta, N. Díaz-Rodríguez, J. D. Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai," 2019. [Online]. Available: <https://arxiv.org/abs/1910.10045> (Citato a pagina 20)
- [61] W. Fedus, B. Zoph, and N. Shazeer, "Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity," 2022. [Online]. Available: <https://arxiv.org/abs/2101.03961> (Citato a pagina 21)
- [62] R. Islam and O. M. Moushi, "Gpt-4o: The cutting-edge advancement in multimodal llm," Jul. 2024. [Online]. Available: <http://dx.doi.org/10.36227/techrxiv.171986596.65533294/v1> (Citato a pagina 21)
- [63] B. M. Gurusamy, P. K. Rangarajan, P. Krishh, A. Keerthinathan, G. Lavanya, M. Meghana, S. Sulthana, and S. Doss, "An analysis of large language models: their impact and potential applications," *Knowledge and Information Systems*, vol. 66, pp. 1–24, 05 2024. (Citato a pagina 21)
- [64] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, pp. 1735–80, 12 1997. (Citato a pagina 21)
- [65] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," 2014. [Online]. Available: <https://arxiv.org/abs/1406.1078> (Citato a pagina 21)
- [66] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994. (Citato a pagina 22)
- [67] D. Rothman, *Transformers for Natural Language Processing: Build innovative deep neural network architectures for NLP with Python, PyTorch, and TensorFlow*. Packt Publishing, 2020. (Citato a pagina 22)
- [68] S. Islam, H. Elmekki, A. Elsebai, J. Bentahar, N. Drawel, G. Rjoub, and W. Pedrycz, "A comprehensive survey on applications of transformers for

- deep learning tasks,” *Expert Systems with Applications*, vol. 241, p. 122666, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417423031688> (Citato a pagina 25)
- [69] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, “Language models are unsupervised multitask learners,” 2019. [Online]. Available: <https://api.semanticscholar.org/CorpusID:160025533> (Citato a pagina 25)
- [70] J. Howard and S. Ruder, “Universal language model fine-tuning for text classification,” 2018. [Online]. Available: <https://arxiv.org/pdf/1801.06146> (Citato a pagina 25)
- [71] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, “Lora: Low-rank adaptation of large language models,” 2021. [Online]. Available: <https://arxiv.org/abs/2106.09685> (Citato a pagina 25)
- [72] B. Jacob, S. Kligys, B. Chen, M. Zhu, M. Tang, A. Howard, H. Adam, and D. Kalenichenko, “Quantization and training of neural networks for efficient integer-arithmetic-only inference,” 2017. [Online]. Available: <https://arxiv.org/abs/1712.05877> (Citato a pagina 26)
- [73] A. S. Luccioni, S. Vigui er, and A.-L. Ligozat, “Estimating the carbon footprint of bloom, a 176b parameter language model,” 2022. [Online]. Available: <https://arxiv.org/abs/2211.02001> (Citato a pagina 26)
- [74] L. Casillo. (2025) SoulDiaryConnect. Accessed: 2025-02-09. [Online]. Available: <https://github.com/FLaTNNBio/SoulDiaryConnect> (Citato a pagina 29)
- [75] A. S. Badashian, M. Mahdavi, A. Pourshirmohammadi, and M. M. nejad, “Fundamental usability guidelines for user interface design,” in *2008 International Conference on Computational Sciences and Its Applications*, 2008, pp. 106–113. (Citato a pagina 30)
- [76] P. Fusar-Poli, M. Manchia, N. Koutsouleris, D. Leslie, C. Woopen, M. E. Calkins, M. Dunn, C. L. Tourneau, M. Mannikko, T. Mollema, D. Oliver, M. Rietschel, E. Z. Reininghaus, A. Squassina, L. Valmaggia, L. V. Kessing,

E. Vieta, C. U. Correll, C. Arango, and O. A. Andreassen, "Ethical considerations for precision psychiatry: A roadmap for research and clinical practice," *European Neuropsychopharmacology*, vol. 63, pp. 17–34, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0924977X22008276> (Citato a pagina 39)

---

## Ringraziamenti

---

Desidero ringraziare il Professor Rocco Zaccagnino per la sua guida preziosa, i suoi incoraggiamenti e l'ottima musica ascoltata in questi mesi.

Ringrazio inoltre il Dottor Gerardo Benevento per essermi stato accanto nella stesura di questa tesi non solo come tutor, ma come compagno e supporter; e il Dottor Nicola Lettieri per tutti i consigli e i confronti avuti in questo percorso. Grazie a Lei ho avuto la possibilità di avere un altro orizzonte su molti argomenti, non solo quelli affrontati in questo progetto.

Un ringraziamento speciale alla Dottoressa Maria Mirra, specialista in ambito psicoterapeutico, senza la quale questa tesi non avrebbe visto le fondamenta, i dati reali, e le valutazioni finali.

---

Ora vorrei fare alcuni ringraziamenti personali, quindi:

**Premessa:** se li leggo tutti è colpa vostra, quindi ascoltate e soffrite.

Se attualmente sto leggendo queste righe sicuramente è grazie a tutti voi. Spero non baderete all'ordine perché sennò è la fine.

Non sono la persona più indicata a esprimere i sentimenti, né quella più adatta a dimostrarli molto spesso, ma ho capito il senso dei ringraziamenti in questo documento nel momento in cui ho iniziato a scriverli. "Perché ringraziare per il lavoro di tesi svolto?" mi chiedevo. La risposta però è che la domanda che mi ponevo era sbagliata. Quante volte, dopo una partita, si ringrazia il proprio avversario o, dopo un percorso, si ringrazia chi lo ha condiviso con te? Ecco, siamo nel secondo caso. Vorrei quindi ringraziare chi mi ha dato la possibilità di fare questo percorso, la mia famiglia, poi tutti voi.

Papà, grazie per essere il mio "fratello maggiore", per guidarmi nella vita senza mai forzare il percorso, per quei silenzi la sera che valgono più di tante parole. Grazie per farmi compagnia in quei momenti dove nessun altro riuscirebbe senza darmi fastidio. Grazie per esserci senza importi, e grazie anche per i tuoi punti di vista ferrei. Odio ammetterlo ma a volte mi danno punti di vista diversi su cose che credevo assolute.

Mamma, grazie per esserci sempre. Nonostante i nostri caratteri simili, e quindi i nostri "Dialoghi sopra i due massimi sistemi del mondo", giusto per citare Galilei, riusciamo sempre a ritornare come prima e porre le basi per altri dialoghi. Grazie per avermi sempre aiutato, per ricordarmi costantemente le cose (Luca Demented or not demented? Obviously demented), e per cercare di prenderti cura di me.

---

Vincenzo, è difficile parlare del nostro rapporto. Sarebbe come dimostrare che  $P=NP$ . Nonostante non te lo dica spesso e nonostante tu non mi faccia dormire quando torni a casa nelle varie festività, mi piacciono le giocate insieme e le sgattaiolate a San Giuseppe per le parigine. Grazie per il supporto quasi sempre costante e per esserci sempre stato, anche se non riesci proprio a ricordare che il Corso di Laurea si chiama solo "Informatica".

Antonio, sei per me come un fratello. Ti ho conosciuto tra quei banchi a P4 il primo anno e da allora ti ho sempre considerato un grande amico. Ci sei sempre, soprattutto la notte, per supportarmi e aiutarmi. Sono felice di poter far parte della tua vita e sono orgoglioso della persona che stai diventando. C'è chi guarda gli uccelli in cielo, tu guardi le mango loco e va bene così, anche se, quando dici casotto più di 8 volte (volevo mettere 3 ma avresti detto tu 8), diventa too much.

Rosa, sei la dimostrazione che Piazzollesi e Torresi possono essere amici. Sei la prima persona che ho conosciuto in università ed è grazie a te che il primo giorno sono arrivato in aula senza perdermi. Nonostante ci scontriamo molto spesso, so di poter contare sempre su di te, sia a livello personale, che universitario. Per quanto a volte la tua puntualità sia eccessiva, sei un'ottima compagna di progetto. Non mi mancheranno le ore passate insieme a programmare, ma spero non mi mancheranno mai le risate fatte insieme enoenoeno.

Matteo, piccolo panda e grande rappresentante, grazie per tutte le minchiate fatte insieme in questi anni. Mi hai insegnato insieme a Massimo il calcolo della probabilità applicata al blackjack e hai sempre dato voce ai miei scleri. Quando diventeremo seri UniSa cadrà. Grazie per aver condiviso con me questo percorso, e ChatGPT.

Cocco, figlio ingrato e tanto amato. Grazie per essere un amico con cui condividere tutto e soprattutto la passione per gli anime. Grazie a te, o per colpa tua, decidi tu, ho ritrovato la passione per Dokkan. Sono poche le persone a cui affiderei tutto, e tu sei una di esse. Ti voglio bene a papà.



---

Peppe, o forse dovrei chiamarti Signor Giuseppe Pino Raiòla Paduano, se ti aspetti che ti faccio la solita cacciata sul cornetto, ti prometto che non la farò più (forse) se mangerai i dolcini al banchetto, dopotutto li ha fatti tuo padre. Grazie a te mi sono avvicinato ad ASS e grazie a te ho scoperto lati del mio autismo che ignoravo. Sei un'ottima persona e un'ottimo presidente, anche se spesso, o sempre, ti sottovaluti. Grazie a te ho superato PD senza far arrabbiare troppo Chiara (le bestemmie Chiara...). Grazie per il disagio condiviso, i momenti seri e tutti i passaggi che mi hai dato. Spero che le chiavi non ricompariranno più dentro le mie tasche.

Aru, anche se ci conosciamo da relativamente poco, sei una sorta di fratello maggiore. Sei la dimostrazione che l'età è solo un numero e che i trentenni non sono tutti da buttare. Grazie per avermi accompagnato in questo percorso con una ferma delicatezza. Non hai mai avuto una parola di troppo, ma non hai mai detto troppo poco. Grazie per avermi accompagnato in questo percorso, oltre che in quest'avventura che è il direttivo. Spero di continuare ad essere in auto dietro di te mentre Peppe guida.

Giorgina, salvatrice dei miei bus serali. Sei una grande amica e la più buona dei nani malefici. Grazie per gli inciuci, per le chiacchiere e per le giornate passate insieme al desk, spero di continuare ad avere questo rapporto anche al di fuori di queste quattro mura azzurre napoli.

Francesco, o meglio noto come Fattorusso o Ffaffeffo Faffofuffo. Scusa per aver sbagliato a dire il tuo nome e aver creato questo trend. Grazie per essere il mio compagno di ingiarmi e pezzotti. Se Bob l'aggiustatutto esistesse, avrebbe la barba e direbbe "SalveSalve". Insieme a Peppe sei la persona che mi ha fatto scoprire l'ebbrezza di essere passeggero dalla ridentissima Fisciano all'ancora più ridentissima Piazzolla. Spero di mangiare altri kebab da 43cm insieme.

---

Chiara, tesoriera dei conti, sei stata una grande risorsa per le procedure da fare per questa tesi. Grande inciucessa, quasi ai livelli di Giorgina, sai dare sempre news ed allegria alle mie giornate. Grazie per essere sempre così solare e memosa. Grazie per accettare la mia ironia e il mio affetto pungente, visto che praticamente non ti chiamo mai per nome. Sarai più spirito che santa, ma sei anche una cucciolina a cui voglio molto bene.

Enzo, nonostante il nostro rapporto sia molto legato all'Associazione, sei una persona che stimo molto e che ha fomentato in me la passione per i pc retrò. So che non potrò mai avere un museo buono quanto la tua stanza, ma grazie a te ora ho questo desiderio. Grazie inoltre per aiutarci nel percorso intrapreso. La tua opinione per me è importante non solo a livello associativo, ma anche personale.

Luigi, Gianfranco e Tullio, vi ho conosciuto realmente solo lo scorso anno. Grazie per esserci stati da allora. Grazie per tutte le risate, i consigli, i kebab, le pizze e le birre insieme. Probabilmente senza di voi molti momenti sarebbero stati più difficili. Siete rispettivamente il rosso, il riccio e lo zio barbuto che tutti vorrebbero avere.

Valeria, non so come fai a sopportare il riccio, anche se ti conosco poco, grazie per i pareri durante lo sviluppo della tesi e per le risate fatte.

Tea e Rara, siete la non quota rosa degli informatici. C'è chi vi invidia per come avete preparato gli esami di matematica, e chi mente. Anche se ci sono periodi in cui non ci sentiamo, vi reputo amiche preziose e insostituibili. Rara, il tuo notion è stata la salvezza che non ci meritavamo sui vari esami. Le playlist abbinate un grande tocco di nonsense. I meme nei direct grande grande cosa. Tea, spero di vederci più spesso e non ogni 4 mesi, così almeno sfottiamo Peppe insieme.

Elvira, ti conosco da pochissimo, ma esiesiesiesiesi. Grazie per avermi ricordato ogni giorno quanto mi mancava alla scadenza, e grazie per le memate dell'ultimo mese. Ti voglio bene.

---

Daniele, Daniele Daniele. Forse sei il mio unico ingegnere informatico amico. Dire di aver condiviso tanto vissuto in questi anni sarebbe riduttivo, però ti ringrazio per esserci nei momenti importanti e nelle serate un po' meh. Grazie per gli sfoghi, le risate, gli insulti e il disagio, in senso positivo, che contraddistinguono le volte in cui ci vediamo.

Carlotta, Cristiano e Fonzis, grazie per le serate di gioco passate insieme. Carlotta, grazie per le *candele*, gli sfottò su Peppe, le risate e i livelli su New Super Mario Bros. Cri, grazie per aver avuto un'idea perfetta per il babbo natale segreto. Non userò mai quel taccuino ma lo custodirò molto gelosamente. Grazie inoltre per essere una persona con la quale si può parlare di tutto. Senza vederti arrivare e prendere il succhino le mattine saranno diverse. Fonzis, grazie per avermi insegnato il treasure hunting e per tutti i giochi. Grazie per la risata contagiosa e per la tua sincerità d'animo.

Dottor Gerardo Benevento, o forse dovrei dire Dino, amore mio. Grazie per essere diventato mio amico oltre che mio tutor. Sei stato capace di darmi supporto sempre e di rialzarmi il morale nei momenti più impensabili. Grazie per la musica, le risate, la lotta greco-romana e la Sagra delle Sagre. Spero di rilavorare in futuro con te e di visitare il Monte Gelbison insieme.

Alessio, o forse dovrei dire 'O polà. In questi ringraziamenti sei la persona che mi conosce da più tempo dopo i miei genitori. Grazie per esserci stato in questi 9 anni. So che non ci sentiamo spesso, ma sei una presenza fissa nella mia vita. Dicono che chi conosci per più di dieci anni sarà tuo amico per tutta la vita. Spero di restare amici altri due anni allora, così da superare la quota minima.

Denise, rappresentante del mio cuore e del cuore della panterona. Grazie per avermi aiutato nel MEP. Grazie anche per sopportarmi continuamente senza mai mandarmi affanculo. Ormai si può dire che ti ho visto crescere, e sono molto fiero della donna che stai diventando. Resta sempre così, delicata ma allo stesso tempo forte e dinamica. Tivibi.

---

Seb, non so se te l'aspettavi di essere qui. Grazie anche a te per avermi aiutato negli ultimi tempi nel MEP. Se dovessi scegliere un'erede come ha fatto AllMight col OneForAll, sceglierei te, anche se non abbiamo molto tempo, visto che i capelli stanno finendo.

MangoCrew, aka Tony, Grazia, Antonio, Grazia, Giovanni Esposito, Raffaele, Alessandro, Eleonora ed Alva (dai fa figo il nome d'arte), grazie per esserci quotidianamente. Dall'attesa per il bus la mattina, passando per le risate in uni, finendo con la mango loco la sera. Grazie a voi ho iniziato ad avere un'allergia per il mango, si scherza, e grazie a voi, oltre che alle cacate, ho avuto modo di avere punti diversi su tante cose. C'è chi crede in #freebossetti, io credo in #freecasottomlb grazie a voi.

Michela, infermiera di notte, grazie per esserci il giusto. Mai troppo, mai troppo poco. Il nostro rapporto alla fine va così, un filo che ci collega, e che anche se a volte non si vede resta sempre lì senza mai spezzarsi. Grazie per la tua ironia, la tua gentilezza e la tua timidezza.

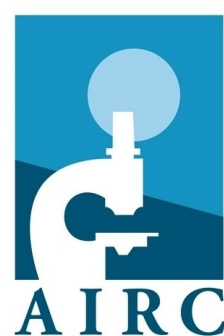
Cristina, o forse dovrei dire Cri. Prometto di provare a chiamarti così d'ora in poi. Senza di te l'università, e specialmente il primo anno, sarebbe stata monotona. Per fortuna o sfortuna sono successe così tante cose da non ricordarmele tutte. Anche se non te lo dico spesso, ti voglio bene. Se a volte sono duro con te, è solo perché vedo la bella persona che potresti essere dietro la maschera che a volte metti. Grazie per esserci quando c'è bisogno.

Chiara, anche se ci conosciamo relativamente poco e non ci vediamo spesso, o forse dovrei dire che io conosco te meno di quanto tu conosca me (grazie a marygio), hai avuto sempre parole gentili per sollevarmi il morale. So che purtroppo sei di parte su chi dei due tra me e marygio è meno cattivo, ma vabbè, su quello posso farci poco. Grazie per tutte le chiacchierate e per avermi insegnato che Kb non sono solo kilo-byte, ma anche kilo-basi.

---

Vorrei ringraziare inoltre tutti i ragazzi che ho conosciuto in questi anni. Purtroppo non posso mettervi tutti che sennò non finirei più, ma in particolare quelli della casa che mi ha accolto per prima, l'Associazione Coscienze, e la casa dove ho trovato un mio posto dove voler stare, l'Associazione Studenti Scienze. Grazie a tutti voi ho sconfitto la noia di tutti i giorni, creato legami, fatto inciuci e superato esami. Un ringraziamento speciale a Marisa, la moglie di mio marito; Benedetto, colui che mi ha salvato l'ultimo esame; Albertone, la pelata studentesca più bella di UniSa, AlfonsoMaurizio, la seconda pelata studentesca più bella di UniSa, Giovanni (Carbone e Tufano), Grazia, Giuseppe Farina, Simone Florio, Gedi, Biagio, Alessia, Tess, Daria e Ragoz.

Amoo, te sei volontariamente messa per ultima, perchè sei la più difficile da ringraziare. Vorrei non fare un'altra tesi, quindi cercherò di essere breve e coinciso (non circonciso Peppe). Ci siamo conosciuti quattro anni fa. Tra un dispetto e un altro, tra un confessionale e un altro, ci accorgemmo veramente di chi avevamo di fronte, forse perchè destinati ad essere qui. Sicuramente non è stato un percorso rettilineo, ma dopo tutto pure i ciclisti pedalano in salita. Nonostante tu sia una gran testarda e parta molto facilmente con la quinta, grazie per esserci giorno dopo giorno, più di ieri e meno di domani. Probabilmente sei tra i pochi, se non l'unica, che sa tutto quello che mi è passato per testa, e sai quanti sia difficile per me aprirmi realmente. Sono molto contento di averti al mio fianco, anche se mi fai arrabbiare un sacco quando non credi in te stessa o quando non vuoi che gli altri lo facciano per te. Citando Kamina, *se non credi in te stessa, credi nella parte di me che crede in te*. Ti amo.



*Questa tesi ha contribuito alla ricerca sul cancro tramite  
l'Associazione Italiana per la Ricerca sul Cancro.*