



**UNIVERSITÀ
DEGLI STUDI
DI PADOVA**



DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

CORSO DI LAUREA IN: Ingegneria Informatica

“Servizi psicologici basati su AI: rischi etici e legali e possibili soluzioni tecniche per ridurli ”

Relatore: Prof. Antonio Rodà

Laureando: Tommaso Terrin

Correlatore: Prof.ssa Silvana Badaloni

ANNO ACCADEMICO 2024 – 2025

Data di laurea 17 Marzo 2025

Sommario

Grazie ai progressi nei modelli di linguaggio di grandi dimensioni (*Large Language Model - LLM*), si stanno delineando cambiamenti radicali in quasi tutti i settori della nostra società, tra cui psicologia, educazione e assistenza sanitaria. La capacità di questi modelli di analizzare, comprendere e rispondere al linguaggio naturale umano offre opportunità mai viste, ad esempio, nella terapia psicologica, dove possono supportare il trattamento di disturbi cognitivi e facilitare il benessere mentale. Tuttavia, l'adozione di LLM su larga scala impone anche la necessità di norme e regolamentazioni solide, come quelle proposte dall'AI Act, la normativa europea volta a regolamentare l'uso e lo sviluppo di intelligenze artificiali. L'AI Act rappresenta un tentativo di bilanciare innovazione e sicurezza, imponendo, ad esempio, controlli di trasparenza e rispetto della privacy per ogni livello di rischio associato all'uso di AI. Sebbene tali regolamentazioni siano cruciali per evitare discriminazioni o abusi, la loro applicazione resta una sfida importante. Tra gli aspetti da tenere maggiormente in considerazione c'è sicuramente l'affidabilità delle risposte generate da LLM, che rappresenta un aspetto cruciale per ridurre fenomeni come le allucinazioni e i bias cognitivi. Tra le tecniche emergenti, il Graph-RAG (Retrieval Augmented Generation tramite grafo) si distingue per la capacità di ampliare la base di conoscenza di un LLM integrandolo con un database a grafo. Questa metodologia si è dimostrata particolarmente promettente, offrendo risultati superiori rispetto all'approccio tradizionale basato su database vettoriali. In questo contesto una prima ricerca, presentata da Microsoft all'inizio dell'anno, ha avuto dei risultati molto interessanti in tal senso. Nell'ultimo capitolo di questo lavoro di tesi sarà anche analizzata l'implementazione di una soluzione software con lo scopo di affrontare una delle sfide più complesse di questa tipologia di approccio: la definizione di entità e relazioni tra i concetti per definire una solida base di conoscenza da utilizzare in un sistema RAG con grafo.

Contents

List of Figures	xiii
List of Tables	xiii
List of Code Snippets	xvii
1 Introduzione	1
1.1 L'AI Act	2
1.1.1 Principi dell'AI Act	2
1.1.2 Approccio basato sul rischio	3
1.2 Progetto Recognify	3
1.2.1 Cos'è Recognify e quali sfide vuole superare	4
2 Cos'è il RAG e come può rendere un sistema più conforme all'AI act	7
2.1 In che modo l'uso di RAG rende un sistema intelligente più conforme all'AI act	7
2.2 RAG naive, funzionamento e limiti	8
2.3 Cos'è il grafo di conoscenza e come utilizzarlo per il RAG	9
3 Framework GraphRAG e la sua applicazione per Recognify	11
3.1 Introduzione	11
3.2 Pipeline di GraphRAG	12
3.2.1 Suddivisione dei documenti in chunk di testo	13
3.2.2 Estrazione degli elementi del grafo	13
3.2.3 Sintesi delle entità	14
3.2.4 Creazione del grafo e partizionamento in comunità	14
3.2.5 Generazione della sintesi delle comunità	14
3.2.6 Generazione della risposta finale	14
3.3 Metodo per evaluation	15

CONTENTS

3.4	Risultati	15
3.5	Possibili sviluppi futuri	18
4	Sviluppo di un Sistema Multiagente per l'Analisi delle Distorsioni Cognitive	21
4.1	Introduzione	21
4.1.1	Distorsioni cognitive in ambito CBT	21
4.1.2	Implementazione	22
4.1.3	Evaluation	22
4.2	Implementazione	22
4.2.1	Agente 1: Definizione schema Entità-Relazione	23
4.2.2	Agente 2: Inserimento dello schema ER nel grafo e definizione di triplette	25
4.2.3	Agente 3: Analisi delle triplette per rilevamento delle Distorsioni	26
4.2.4	Utilizzo del Prompt Engineering nelle Fasi di Estrazione e Analisi	27
4.3	Risultati sperimentali	29
4.3.1	Descrizione dell'esperimento	29
4.3.2	Confronto con lo stato dell'arte	30
4.3.3	Analisi dei risultati	31
4.4	Possibili sviluppi futuri	31
	Bibliografia	33

List of Figures

10figure.caption.10

3.1	Schema che riassume le fasi per il funzionamento di GraphRAG. Immagine contenuta nella ricerca [11]	13
3.2	Esempio di valutazione di una domanda, confrontando le risposte di graphRAG e naiveRAG	16
3.3	Risultati completi sui 4 aspetti analizzati per tutte e 6 le metodolo- gie utilizzate confrontate tra loro	17

List of Tables

3.1	Numero di riassunti delle comunità per C0-C3 e di text chunk per TS. Ad ognuno viene associato il numero di token utilizzati per i riassunti durante la fase di map-reduce	17
3.2	Confronto tra GraphRAG e LazyGraphRAG. La tabella evidenzia le differenze nei processi di costruzione dell'indice, recupero delle informazioni e generazione delle risposte.	19
4.1	Confronto dei risultati (F1 Score) tra diversi modelli sul dataset <i>llm cbt-bench</i> [46].	30

List of Code Snippets

4.1	esempio di dati di input per il sistema	22
4.2	example of first agent's output	24
4.3	example of second agent's output	25
4.4	esempio di terzo outputt	27
4.5	esempio di prompt per la definizione di una delle 10 distorsioni .	27
4.6	Esempio di struttura di un elemento nel file utilizzato per l'evaluation	29

1

Introduzione

Negli ultimi anni, la crescente potenza computazionale e la disponibilità di enormi quantità di dati hanno reso possibile l'avvento e la rapida evoluzione dei Large Language Models (LLM). Moltissime aziende hanno rilasciato i loro modelli: gpt [33], llama [15], PaLm [9] sono alcuni tra i più famosi, ma anche dall'altra parte del mondo, in Cina, deepseek [16] e qwen [4] stanno avendo risultati interessanti. Questi LLM stanno rivoluzionando il nostro modo di interagire con le macchine poichè sono in grado di analizzare, comprendere e rispondere generando testi o immagini coerenti e informativi utilizzando linguaggio naturale, aprendo nuove frontiere in numerosi settori, da quello economico [45] a quello legale [7], dall'ambito dell'istruzione [14] a quello psicologico [32] [43]. Tuttavia la diffusione pervasiva dell'IA (*Intelligenza Artificiale*) ha anche fatto emergere diversi possibili rischi per un approccio non responsabile a questo strumento [35] e sollevato preoccupazioni sulla necessità di garantire trasparenza e affidabilità dei sistemi basati su LLM. Per rispondere a tale bisogno, l'Unione Europea ha proposto l'AI Act [12]. L'AI Act è una normativa volta a regolamentare lo sviluppo e l'utilizzo dell'intelligenza artificiale, ponendo però l'accento sulla sicurezza e sulla tutela dei diritti fondamentali dell'uomo, definendo anche diverse tipologie di sistema intelligente sulla base del livello di rischio che queste possono avere sotto questi due aspetti.

In questo contesto la presente tesi si concentra sullo sviluppo di tecniche innovative per migliorare l'affidabilità nello sviluppo di un'applicazione intelligente reale, "*Recognify*", in ambito di *mental health*, rendendola quindi anche maggiormente conforme alle normative europee.

1.1 L'AI ACT

L'espansione rapida e capillare dell'IA ha evidenziato sfide importanti, tra cui il rispetto della privacy, l'affidabilità delle risposte date, la prevenzione di bias discriminatori e la necessità di rendere i processi decisionali delle macchine più comprensibili e tracciabili. L'Unione Europea, per rispondere a queste esigenze, ha definito il primo quadro legislativo interamente dedicato a disciplinare l'utilizzo dell'IA e i rischi ad essa associati.

La normativa è entrata in vigore l'1 agosto 2024, anche se l'applicazione completa di alcune disposizioni avverrà in una fase successiva, con scadenza fissata entro il 2 agosto 2026.

1.1.1 PRINCIPI DELL'AI ACT

L'AI Act si basa su una struttura regolatoria che mira a promuovere l'uso dell'intelligenza artificiale in modo sicuro, etico e conforme alle leggi europee. In particolare, i principali **obiettivi** del regolamento sono:

- **Gestione dei rischi** [39]: Il regolamento introduce una classificazione dei sistemi di IA basata sul rischio, adottando diverse misure per mitigare i potenziali effetti negativi. I sistemi vengono suddivisi in categorie che vanno da rischio basso a rischio elevato, con requisiti di conformità più stringenti per le applicazioni che presentano rischi elevati.
- **Protezione della privacy e dei diritti fondamentali**: L'AI Act punta a garantire che i sistemi di IA rispettino i diritti fondamentali sanciti dall'Unione Europea, come il diritto alla privacy, seguendo i principi del GDPR (*General Data Protection Regulation*), la non discriminazione e la protezione dei dati personali.
- **Trasparenza**: La trasparenza è un altro principio fondamentale su cui si basa il regolamento. L'AI Act richiede che le applicazioni di IA siano comprensibili per gli utenti e che sia sempre chiaro quando un sistema utilizza l'IA, assicurando la possibilità di comprendere i processi decisionali. Questo comporta l'obbligo di minimizzare la raccolta e l'elaborazione dei dati e di utilizzare solo quelli strettamente necessari per il funzionamento del sistema intelligente. Inoltre, nei casi di utilizzi per riconoscimento facciale o biometrico, l'AI Act richiede che tali applicazioni siano strettamente regolamentate e, in molti casi, utilizzabili solo con consenso esplicito. Ad esempio è essenziale che il processo sia tracciabile e che si possano identificare eventuali bias, garantendo la possibilità di intervenire per ridurre i pregiudizi.

1.1.2 APPROCCIO BASATO SUL RISCHIO

Uno dei punti principali trattati dall'*AI Act* è l'**approccio basato sul rischio**, infatti nell'articolo 5 esso definisce quattro tipologie di Intelligenze artificiali e, nei successivi, regole di vigilanza e conformità proporzionali al rischio che queste comportano. Vengono definite **4 tipologie** di rischio:

1. **Rischio inaccettabile:** Il regolamento vieta gli utilizzi dell'IA che presentino un rischio inaccettabile per i diritti fondamentali, come i sistemi che utilizzano tecniche di manipolazione subliminale, l'assegnazione di punteggi sociali e alcune tipologie di riconoscimento facciale in tempo reale per scopi di sorveglianza. Questi usi sono considerati incompatibili con i valori europei e, pertanto, vietati.
2. **Rischio alto:** I sistemi classificati come ad alto rischio includono quelli utilizzati in settori critici come la sanità, l'istruzione, l'occupazione e la pubblica sicurezza. Per queste applicazioni, l'*AI Act* prevede che siano rispettati criteri di conformità specifici, come la valutazione dell'impatto, la documentazione della trasparenza e la supervisione umana, con l'obiettivo di limitare i rischi di errore e pregiudizio. Il sistema di *Recognify* rientra quindi in questa categoria.
3. **Rischio limitato:** Le applicazioni a rischio limitato, come i chatbot che forniscono supporto clienti, richiedono solo che gli utenti vengano informati di interagire con un sistema di IA. In questi casi, l'obbligo è quello di informare chiaramente l'utente e non è necessario seguire protocolli rigidi di conformità.
4. **Rischio minimo:** Le applicazioni considerate a rischio minimo, come i filtri anti-spam, non richiedono requisiti specifici per la conformità, permettendo l'uso libero dell'IA in ambiti dove non sono previsti rischi significativi.

1.2 PROGETTO RECOGNIFY

Perchè Recognify? Nel 2019 è stato riportato che circa 84 milioni di cittadini dell'Unione Europea soffrono di problemi psicologici, circa un sesto della popolazione collettiva [13]. Inoltre, l'avvento della pandemia COVID-19, nel 2020, ha esacerbato in modo significativo la prevalenza di problemi di salute mentale, soprattutto tra i giovani [44] [1]. Moltissime persone non accedono alle cure per due ragioni principali:

- lo *stigma culturale* relativo alla cura psicologica [10].

1.2. PROGETTO RECOGNIFY

- le *barriere di accesso*, legate soprattutto ai costi e alle lunghe file di attesa [3].

Questi problemi critici sottolineano l'urgente necessità di un supporto per la salute mentale accessibile e conveniente su scala globale, con una soluzione che affronti direttamente le principali barriere: costi, accesso e stigma. I risultati sui primi studi che riguardano l'utilizzo di LLM in ambito psicologico sono molto promettenti [31] [21]. Addirittura, nell'esperimento [2], risulta che i modelli sono riusciti a superare il livello di intelligenza emotiva recepita da un interlocutore, rispetto a quella di studenti e ricercatori di psicologia.

1.2.1 COS'È RECOGNIFY E QUALI SFIDE VUOLE SUPERARE

L'**obiettivo** di *Recognify* è quello di creare un'applicazione che integri tecniche di intelligenza artificiale e metodologie psicologiche per fornire un supporto personalizzato nell'ambito della salute mentale e dell'auto-miglioramento. Il sistema utilizza l'AI per avere un approccio olistico e un'esperienza fortemente personalizzata sulla base delle varie esigenze individuali.

L'approccio psicoterapeutico utilizzato è la **CBT** [5] (*Cognitive Behavioural Therapy, Terapia Cognitivo Comportamentale*), che mira a modificare schemi di pensiero e comportamenti disfunzionali attraverso l'identificazione e la ristrutturazione delle distorsioni cognitive. Diversi studi hanno già avuto risultati promettenti sull'utilizzo di questa metodologia psicologica tramite risposte generate da LLM [40] [22].

Recognify non mira a sostituire professionisti umani (cosa peraltro impossibile con le tecnologie attuali), ma a colmare il divario tra la terapia tradizionale e le soluzioni di auto-aiuto, fornendo uno strumento accessibile ed efficace per il miglioramento del benessere emotivo e della resilienza psicologica. La piattaforma si basa su modelli scientificamente validati, con un'attenzione particolare all'identificazione e alla correzione delle distorsioni cognitive. Inoltre propone esercizi mirati, atti a ricercare un processo di cambiamento duraturo, adattandosi dinamicamente per ottimizzare l'intervento in base ai progressi dell'utente.

Le funzionalità di *Recognify* sono molteplici e si articolano su diversi livelli di interazione:

- **Interazione con il Paziente:** L'utente interagisce con l'applicazione tramite un'interfaccia conversazionale intuitiva. Tramite questa interfaccia il sistema raccoglie informazioni sullo stato emotivo e sulle esperienze quotidiane del paziente, individuando automaticamente situazioni critiche e relativi pensieri disfunzionali. Il percorso di miglioramento inoltre è caratterizzato da un'interfaccia grafica simbolica, denominata *mind garden*, che rappresenta visivamente i progressi compiuti dall'utente: ogni esercizio completato e ogni pensiero riformulato fa "fiorire" il giardino virtuale, fornendo un feedback immediato e motivante.
- **Interazione con il Terapeuta:** Pur essendo pensato come strumento di auto-aiuto, *Recognify* supporta anche un'integrazione con il lavoro del terapeuta. Il sistema raccoglie e organizza dati strutturati relativi alle sessioni. Questi possono essere schemi di pensiero, distorsioni identificate o i progressi fatti e possono essere consultati e analizzati dal professionista. In questo modo il terapeuta avrà uno strumento per monitorare in modo continuo l'evoluzione del paziente.

Infine *Recognify* si pone come obiettivo quello di avere un ruolo di appoggio a pazienti e a centri per la cura della salute mentale dando un sostegno durante il periodo di attesa in lista di accesso alle cure. Il sistema dà l'opportunità di fornire un periodo di supporto e di raccogliere dati preliminari sulle problematiche dei futuri pazienti.



Cos'è il RAG e come può rendere un sistema più conforme all'AI act

Il **RAG** (*Retrieval Augmented Generation - Generazione Aumentata dal Recupero*) è una tecnica che combina il **recupero di informazioni da fonti esterne** con la **generazione testuale di un modello**, consentendo ai modelli di linguaggio di fornire risposte più accurate basandosi su dati aggiornati o specifici. Le informazioni recuperate vengono fornite come contesto aggiuntivo al modello linguistico, che genera una risposta più accurata e contestualizzata.

Viene utilizzato per migliorare l'accuratezza delle risposte e ridurre allucinazioni del modello AI. Inoltre permette l'utilizzo di informazioni non contenute all'interno del modello, che possono quindi essere facilmente aggiornate e contestualizzate.

2.1 IN CHE MODO L'USO DI RAG RENDE UN SISTEMA INTEL-LIGENTE PIÙ CONFORME ALL'AI ACT

Il RAG si pone come uno strumento fondamentale nello sviluppo di sistemi IA più responsabili e conformi alle attuali normative europee sotto i seguenti aspetti:

- *Miglior Governance dei Dati* (Art.10, IA act): RAG permette di recuperare informazioni da fonti esterne, integrando dati accurati e aggiornati, riducendo dipendenza da dati di training potenzialmente obsoleti o incompleti.

2.2. RAG NAIVE, FUNZIONAMENTO E LIMITI

- **Migliore Trasparenza** (Art.13, IA act): Le reti neurali sono black-box, quindi è molto difficile capire come, a partire da un certo input, il modello arrivi a una certa risposta. Viceversa le fonti utilizzate nel RAG sono contenute in database, facilitandone il reperimento. Collegare le risposte alle fonti rende più trasparente il processo decisionale di un sistema AI
- **Maggiore Supervisione umana** (Art.14, IA act): RAG permette di supervisionare le fonti e correggere errori, facilitando l'intervento umano quando necessario. Garantisce quindi una costante supervisione umana sul sistema, aspetto fondamentale per le questioni etiche legate all'AI.

2.2 RAG NAIVE, FUNZIONAMENTO E LIMITI

Il *RAG naive* [24] è il primo, e ad oggi più utilizzato, approccio base di Retrieval-Augmented Generation. Recupera documenti rilevanti attraverso una semplice ricerca semantica. Oggi viene definito "*naive*" (ingenuo) proprio per la sua semplicità, distinguendolo dagli altri approcci utilizzati.

Il processo si suddivide in tre fasi:

1. **Indicizzazione dei dati:** Il corpus dei dati viene suddiviso in porzioni di dimensioni ottimali, dette "*chunks*". Ogni chunk viene trasformato in un *embedding*, ovvero una rappresentazione vettoriale del dato, utilizzando modelli appositi. Questo viene poi inserito in appositi database vettoriali.
2. **Ricerca della query:** quando arriva una query dall'utente, questa viene convertita in un embedding dallo stesso modello utilizzato per i documenti. Successivamente l'embedding della query viene confrontato tramite similarità semantica (solitamente approssimata tramite la distanza del coseno nello spazio vettoriale dell'embedding [37]), facendo un confronto geometrico tra il vettore query e quelli ottenuti dal corpus di dati.
3. **Generazione condizionata:** i documenti recuperati vengono concatenati alla query (aggiungendo dati nel contesto in input) e passati a un modello generativo che, utilizzando questi due elementi, offre una risposta più accurata e contestualizzata.

Nonostante la sua semplicità implementativa, il naive RAG presenta i seguenti limiti:

1. Difficoltà nel gestire contesti complessi interconnessi
2. Recupero poco selettivo di informazioni

3. Scarsa comprensione delle relazioni tra documenti
4. Limitata capacità di sintesi globale

2.3 COS'È IL GRAFO DI CONOSCENZA E COME UTILIZZARLO PER IL RAG

I **KG**, **Knowledge graphs** (*Grafi di conoscenza*) [18] sono un modo per rappresentare le informazioni in cui i dati sono organizzati all'interno di grafi. I nodi rappresentano **entità**, ovvero concetti, mentre gli archi rappresentano le **relazioni** che esistono tra due entità. Un esempio di grafo è presente alla Figura 2.1. Questi grafi aiutano a dare significato ai dati collegando i diversi concetti in modo logico e contestuale.

Nel 2012 Google ha iniziato a utilizzare questo strumento [41] per strutturare i dati a cui ricorre il suo celebre motore di ricerca perchè presentava innumerevoli vantaggi. Innanzitutto permette di superare la semplice ricerca basata sulla corrispondenza tra stringhe attraverso una comprensione legata al significato semantico delle query. Inoltre comporta importanti miglioramenti nella comprensione delle entità, analizzando come queste sono connesse e non considerandole solo come termini isolati.

Grafi e LLM. Per queste sue caratteristiche, in molti studi si è cercato di combinare le potenzialità dei *KG* con quelle degli *LLM* [26]. Da un lato, grazie alla loro capacità di comprensione dei testi, gli *LLM* possono estrarre le entità fondamentali e le relazioni presenti tra loro [48] utilizzando semplici tecniche di *prompt-engineering* [6]. D'altro canto i *KG* possono essere usati dai modelli come base di conoscenza, grazie alla loro capacità di arricchire la comprensione delle relazioni tra concetti, migliorando la precisione delle risposte e l'affidabilità delle informazioni fornite. Questo aiuta a rendere i modelli più consapevoli delle connessioni tra le informazioni, permettendo loro di offrire risposte più strutturate e coerenti, anche lavorando su diversi **domini specifici** [29]. Inoltre, la struttura di un *KG* permette di ricostruire in maniera precisa il ragionamento sui dati, permettendo anche la presenza di sistemi di *fact-checking*, come evidenziato nelle ricerche [23] e [34]. Questo rende il sistema più sicuro e conforme ai principi dell'AI act.

2.3. COS'È IL GRAFO DI CONOSCENZA E COME UTILIZZARLO PER IL RAG

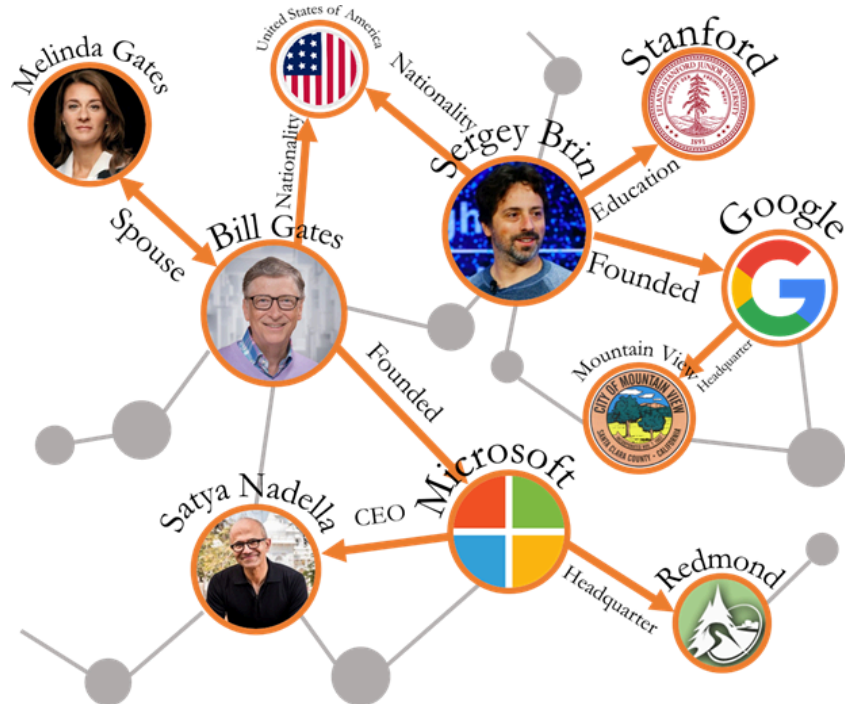


Figure 2.1: Esempio di grafo di conoscenza. (Immagine presa da un articolo sul web¹)

A livello accademico già diversi framework di RAG su KG sono stati testati: *GraphRAG* [11], che sarà approfondito nel capitolo 3, *KAG* [27], *SURGE* [20], *ToG 2.0* [30], *GRAG* [19], *HippoRAG* [17] e *DALK* [25]. I test sui seguenti approcci porterebbero i seguenti **vantaggi**:

1. Migliori **capacità di ricerca globale**: i database vettoriali sono insiemi di documenti, quindi i dati sono strutturati e prelevati secondo un principio di località locale. Viceversa l'entità di un grafo può essere collegata tramite arco ad una qualsiasi altra entità sulla base di una qualsiasi relazione tra le due. In questo modo è possibile effettuare una ricerca che possa analizzare l'intero database nel suo contesto e non documento per documento.
2. Miglior **gestione delle disambiguità**: miglior diversificazione dei vari concetti, non solo sulla base delle entità ma anche sulla base delle relazioni tra di esse.
3. Miglior **capacità nell'effettuare ricerche specifiche**: il retrieval di dati può essere fatto su entità specifiche, non necessariamente su chunk (di dimensioni fisse). In questo modo si ha una maggiore granularità nell'utilizzo dei dati, ricercando solo quelli che effettivamente si vogliono utilizzare.

3

Framework GraphRAG e la sua applicazione per Recognify

Abbiamo visto finora i vantaggi nell'uso RAG e di come questo permetta ad un sistema di essere maggiormente conforme alle norme dell'AI act. Tuttavia, le tecniche di RAG utilizzate oggi spesso falliscono su domande globali. Per questa funzionalità complementare si utilizzano tecniche di *summarization guidato da query* (QFS) [47], tecniche che hanno però limiti relativi alla scalabilità su grandi quantità di dati.

Per combinare i punti di forza di questi due metodi, a luglio del 2024 Microsoft ha pubblicato una ricerca riguardante la creazione di GraphRAG, un framework opensource con un approccio innovativo per il question answering. Combina la generazione di grafi di conoscenza, RAG e la QFS. Ha lo scopo di migliorare il retrieval di informazioni per rispondere a domande più complesse, in particolare se fatte su un contesto globale.

3.1 INTRODUZIONE

Nella sezione introduttiva la ricerca definisce le sfide legate all'elaborazione di grandi quantità di dati e all'automazione del *sensemaking* (processo di comprensione di connessioni tra persone, luoghi ed eventi per anticiparne gli sviluppi). In particolare si descrivono **due limiti** principali dei sistemi RAG tradizionali:

1. **Inadeguatezza per domande globali:** il recupero di chunk testuali non

3.2. PIPELINE DI GRAPHRAG

supporta sintesi complessive, necessarie per analisi tematica e identificazione di trend.

2. Gli LLM faticano a gestire corpus di testo che superano la lunghezza massima della finestra di contesto del modello, con conseguente rischio di perdita di informazioni.

I ricercatori di Microsoft hanno dunque introdotto GraphRAG, un approccio innovativo basato su:

1. **Costruzione di un grafo**, estraendo dai dati i seguenti elementi:
 - Nodi: contengono entità (concetti relativi al testo)
 - Archi: relazioni tra le entità
 - Covariate: sono informazioni contestuali (ad es. affermazioni o intervalli temporali)
2. **Partizionamento del grafo in comunità**: dopo aver costruito il grafo si suddivide il grafo in comunità (insieme di più nodi correlati) gerarchiche. Ogni comunità viene poi sintetizzata in modo indipendente.
3. **Map-reduce** per la sintesi globale: ad ogni richiesta viene costruita una sintesi unica per la risposta a partire dalla sintesi delle comunità

Il framework è stato testato su due dataset reali: **trascrizioni di podcast** in ambito tecnologico e un insieme di **articoli giornalistici** di differenti argomenti. I risultati (che verranno analizzati nella sezione finale del capitolo) sono molto promettenti per quanto riguarda la completezza e la diversità nelle risposte date e una migliore gestione di corpus fino ad un milione di token.

3.2 PIPELINE DI GRAPHRAG

La pipeline di graphRAG si articola su **sei fasi**, ciascuna ottimizzata per bilanciare precisione, efficienza e scalabilità.

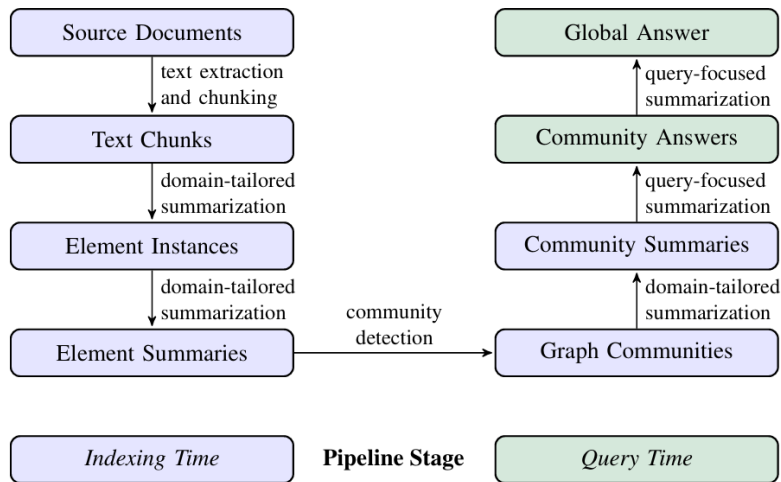


Figure 3.1: Schema che riassume le fasi per il funzionamento di GraphRAG. Immagine contenuta nella ricerca [11]

3.2.1 SUDDIVISIONE DEI DOCUMENTI IN CHUNK DI TESTO

I documenti sorgente vengono suddivisi in segmenti di testo ricercando una granularità ottimale. Come si può vedere dalla figura 3.1 utilizzando chunk di dimensioni ridotte (600 token) invece che chunk di dimensioni maggiori (2400 token) è possibile estrarre un maggior numero di riferimenti entitari ma, di contro, si richiede un maggior numero di chiamate al modello LLM. L'obiettivo di questa fase è quindi quello di bilanciare recall e precisione nell'estrazione delle informazioni.

3.2.2 ESTRAZIONE DEGLI ELEMENTI DEL GRAFO

Ogni chunk viene processato mediante prompt multipart, ovvero un tipo di prompt strutturato che suddivide la richiesta in più richieste distinte. Nella prima fase vengono individuate le diverse entità e, successivamente, le relazioni presenti tra queste. Le istanze vengono quindi restituite come una lista di tuple. Inoltre si utilizzano diversi round di "gleaning" (processo iterativo per estrarre informazioni aggiuntive dai dati) per rilevare eventuali entità mancanti nelle estrazioni iniziali.

3.2. PIPELINE DI GRAPHRAG

3.2.3 SINTESI DELLE ENTITÀ

Le istanze estratte vengono successivamente sintetizzate in blocchi di testo descrittivi per ciascun elemento del grafo (nodi, archi e covariate). Questo processo mira a creare descrizioni autonome dei concetti, anche se questi non sono esplicitamente enunciati nel testo originale. In questo modo si riducono le duplicazioni e si garantisce una rappresentazione uniforme, nonostante possibili variazioni nel modo in cui il LLM identifica le stesse entità.

3.2.4 CREAZIONE DEL GRAFO E PARTIZIONAMENTO IN COMUNITÀ

I riassunti degli elementi vengono utilizzati per costruire un grafo: i nodi rappresentano le entità e gli archi le relazioni, i cui pesi sono determinati in base al numero normalizzato delle occorrenze rilevate. Viene utilizzato l'algoritmo Leiden [42] per rilevare comunità (insiemi di nodi più densamente connessi tra loro) gerarchiche, consentendo una sintesi globale dei dati.

3.2.5 GENERAZIONE DELLA SINTESI DELLE COMUNITÀ

Per ogni comunità individuata si generano riassunti che sintetizzano la struttura dei dati. Questi riassunti hanno una duplice utilità: da un lato forniscono una visione d'insieme del corpus, dall'altro sono utilizzati come parte integrante del grafo per rispondere alle query globali. Nel caso in cui la quantità di informazioni superi il limite di token del contesto del LLM, si procede iterativamente sostituendo riassunti dettagliati con versioni più compatte, fino a ottenere un blocco testuale che rientra nei vincoli di token.

3.2.6 GENERAZIONE DELLA RISPOSTA FINALE

Quando viene posta una query, le risposte vengono elaborate in due fasi:

1. *Fase map*: i riassunti delle comunità vengono elaborati in parallelo per fornire delle risposte parziali alla domanda. Ad ognuna di queste viene inoltre associato un punteggio (da 0 a 100) sulla base di quanto la risposta contribuisca a risolvere effettivamente la query.
2. *Fase reduce*: Le risposte parziali sono ordinate sulla base del loro punteggio (scartando quelle con punteggio nullo o molto basso) e, sulla base di queste si riscrive un riassunto che si utilizza per ottenere un'unica risposta globale.

3.3 METODO PER EVALUATION

La valutazione del sistema è stata condotta su due dataset: trascrizioni di podcast e articoli giornalistici, ciascuno composto da circa 1 milione di token. Per testare i risultati sono state generate domande globali che richiedevano una comprensione complessiva del corpus dei dati. Per il confronto dei dati sono state usate 6 condizioni:

- **C0-C4:** Quattro livelli gerarchici di comunità ottenuti con *graphRAG*. C0 livello root, con meno comunità contenenti dati sintetizzati con un livello di astrazione maggiore. I livelli successivi con sub-comunità ottenute dal livello precedente, quindi di numero maggiore e un livello di astrazione minore, per ottenere risposte più specifiche all'interno dei dati.
- **TS (Text Summarization):** approccio alternativo a *graphRAG*, in cui si riassumono direttamente i chunk di testo, senza la fase di definizione di un grafo.
- **SS (Semantic Search):** RAG tradizionale basato su ricerca semantica.

Tutte queste metodologie sono state confrontate direttamente tra loro sulla base delle risposte date a 125 domande, valutando 4 metriche nelle risposte:

1. *Completezza:* misura di quanto la risposta copre tutti gli aspetti della domanda, fornendo dettagli completi.
2. *Diversità:* valutazione della varietà e ricchezza di prospettive incluse nella risposta.
3. *Empowerment:* quanto la risposta aiuta l'utente a comprendere l'argomento
4. *Chiarezza:* definisce quanto specificatamente e chiaramente la risposta affronta la domanda

3.4 RISULTATI

Il processo di indicizzazione ha prodotto un grafo composto da 8.564 nodi e 20.691 archi per il dataset Podcast e un grafo più grande con 15.754 nodi e 19.520 archi per il dataset News. La tabella 3.1 mostra il numero di riassunti delle comunità ai diversi livelli della gerarchia delle comunità del grafo.

3.4. RISULTATI

Question	Which public figures are repeatedly mentioned across various entertainment articles?
Graph RAG	<p>Overview of Prominent Public Figures in Entertainment</p> <p>The entertainment industry is vast and diverse, encompassing film, television, music, sports, and digital media. Certain public figures stand out due to their significant contributions and influence across these sectors. The following summary highlights key individuals who are repeatedly mentioned in various entertainment articles, reflecting their impact and presence within the industry.</p> <p>Actors and Directors [...] Public Figures in Controversy [...] Musicians and Executives [...] Athletes and Coaches [...] Influencers and Entrepreneurs [...]</p> <p>The repeated mention of these figures in entertainment articles signifies their ongoing relevance and the public's interest in their work. Their influence spans across various aspects of entertainment, from shaping cultural narratives in film and television to driving trends in music and digital media. These individuals not only contribute to their respective fields but also influence the broader cultural landscape, often becoming central figures in social discussions and public discourse.</p>
Naïve RAG	<p>Public figures who are repeatedly mentioned across various entertainment articles include Taylor Swift, Travis Kelce, Britney Spears, and Justin Timberlake. These individuals have been highlighted for various reasons, ranging from their professional achievements to their personal lives.</p> <p>Taylor Swift [...] Travis Kelce [...] Britney Spears [...] Justin Timberlake [...]</p> <p>These figures are frequently covered due to their high-profile status and the public's interest in their careers and personal lives. Their activities, whether in music, sports, or personal relationships, have significant cultural and economic impacts, as evidenced by the media coverage and public reactions.</p>
LLM Decision	<p>Comprehensiveness: Winner=1 (Graph RAG)</p> <p>Answer 1 is better because it provides a more comprehensive and detailed list of public figures from a wider range of entertainment sectors, including film, television, music, sports, gaming, and digital media. It also includes specific examples of their contributions and the impact they have on their respective fields, as well as mentions of controversies and their implications. Answer 2, while detailed in its coverage of a few individuals, is limited to a smaller number of public figures and focuses primarily on their personal lives and relationships rather than a broad spectrum of their professional influence across the entertainment industry.</p> <p>Diversity: Winner=1 (Graph RAG)</p> <p>Answer 1 is better because it provides a more varied and rich response by covering a wide range of public figures from different sectors of the entertainment industry, including film, television, music, sports, gaming, and digital media. It offers insights into the contributions and influence of these figures, as well as controversies and their impact on public discourse. The answer also cites specific data sources for each mentioned figure, indicating a diverse range of evidence to support the claims. In contrast, Answer 2 focuses on a smaller group of public figures, primarily from the music industry and sports, and relies heavily on a single source for data, which makes it less diverse in perspectives and insights.</p> <p>Empowerment: Winner=1 (Graph RAG)</p> <p>Answer 1 is better because it provides a comprehensive and structured overview of public figures across various sectors of the entertainment industry, including film, television, music, sports, and digital media. It lists multiple individuals, providing specific examples of their contributions and the context in which they are mentioned in entertainment articles, along with references to data reports for each claim. This approach helps the reader understand the breadth of the topic and make informed judgments without being misled. In contrast, Answer 2 focuses on a smaller group of public figures and primarily discusses their personal lives and relationships, which may not provide as broad an understanding of the topic. While Answer 2 also cites sources, it does not match the depth and variety of Answer 1.</p> <p>Directness: Winner=2 (Naïve RAG)</p> <p>Answer 2 is better because it directly lists specific public figures who are repeatedly mentioned across various entertainment articles, such as Taylor Swift, Travis Kelce, Britney Spears, and Justin Timberlake, and provides concise explanations for their frequent mentions. Answer 1, while comprehensive, includes a lot of detailed information about various figures in different sectors of entertainment, which, while informative, does not directly answer the question with the same level of conciseness and specificity as Answer 2.</p>

Figure 3.2: Esempio di valutazione di una domanda, confrontando le risposte di graphRAG e naïveRAG

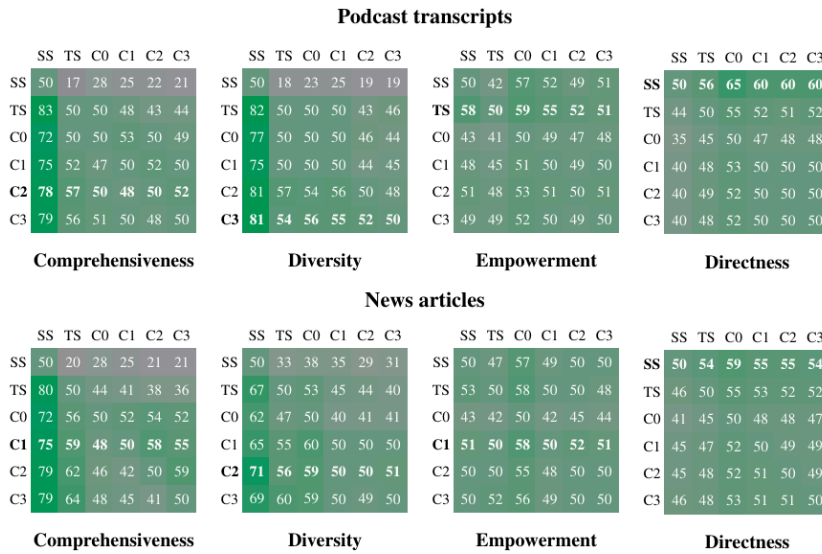


Figure 3.3: Risultati completi sui 4 aspetti analizzati per tutte e 6 le metodologie utilizzate confrontate tra loro

	Podcast Transcripts					News Articles				
	C0	C1	C2	C3	TS	C0	C1	C2	C3	TS
Units	34	367	969	1310	1669	55	555	1797	2142	3197
Tokens	26657	222756	565720	746100	1014611	39770	352641	980898	1140266	1707694
% Max	2.6	22.2	55.8	73.5	100	2.3	20.7	57.4	66.8	100

Table 3.1: Numero di riassunti delle comunità per C0-C3 e di text chunk per TS. Ad ognuno viene associato il numero di token utilizzati per i riassunti durante la fase di map-reduce

Approccio globale vs RAG semantico. Come mostrato nella Figura 3.3, gli approcci globali hanno costantemente superato l'approccio della semplice ricerca semantica (SS) sia in termini di completezza (tra il 72% e l'83% per le trascrizioni dei Podcast e tra il 72% e l'80% per gli articoli di News) che di diversità (tra il 75% e l'82% e tra il 62% e il 71%, rispettivamente).

Riassunti delle comunità vs testi originali. Come visto nella Figura 3.3, quando si confrontano i riassunti delle comunità con i riassunti dei testi originali (TS), i primi hanno generalmente fornito un miglioramento piccolo ma costante nella completezza e diversità delle risposte, tranne per i riassunti di livello radice (C0). La Tabella 3.1 evidenzia anche i vantaggi di scalabilità di Graph RAG rispetto alla sintesi diretta dei testi originali. Per i riassunti delle comunità di basso livello (C3), Graph RAG ha richiesto dal 26% al 33% in meno di token di contesto. Per i riassunti delle comunità di livello radice (C0), la riduzione è stata superiore al 97%. A fronte quindi di una leggera perdita di prestazioni, il graphRAG a

3.5. POSSIBILI SVILUPPI FUTURI

livello C0 rappresenta un metodo altamente efficiente per le risposte a livello globale. Inoltre, mantiene un vantaggio nella completezza (72% di vittoria) e nella diversità (62% di vittoria) rispetto al RAG naive (SS).

Potenziamento (Empowerment) . Come mostrato nella Figura 3.3, i confronti relativi al potenziamento hanno mostrato risultati simili per tutti i sei metodi utilizzati. Ottimizzare i prompt di estrazione degli elementi chiave potrebbe aiutare a mantenere più dettagli nell'indice di *GraphRAG*.

3.5 POSSIBILI SVILUPPI FUTURI

Il 25 novembre 2024, Microsoft ha descritto sul suo blog un nuovo approccio rispetto al graphRAG "tradizionale" denominato **LazygraphRAG** [36]. Come mostrato nella Tabella 3.2 la differenza sostanziale risiede nella fase di riassunto delle comunità: mentre GraphRAG pre-calcola tutti i riassunti delle comunità durante la fase di indicizzazione, , LazyGraphRAG posticipa questa operazione solo alle comunità rilevanti per la query, generando i riassunti su richiesta. Questo riduce drasticamente l'utilizzo di token e il carico computazionale, mantenendo risultati comparabili in termini di completezza e diversità.

In particolare l'articolo enuncia i seguenti risultati:

- Costi comparabili a quelli di un sistema RAG vettoriale e pari allo 0,1% dei costi di GraphRAG completo.
- A parità di costi di query supera sia RAG vettoriale che GraphRAG.
- qualità di risposta paragonabile alla ricerca globale di GraphRAG ma con un costo di query di oltre 700 volte inferiore.

Caratteristica	GraphRAG	LazyGraphRAG
Costruzione dell'indice	Utilizza un LLM per: a) estrarre e descrivere entità e le loro relazioni, b) riassumere tutte le osservazioni di ogni entità e relazione, c) usare statistiche del grafo per ottimizzare il grafo delle entità ed estrarre la struttura gerarchica delle comunità.	Utilizza l'estrazione di frasi nominali tramite NLP per estrarre concetti e le loro co-occorrenze, b) usa statistiche del grafo per ottimizzare il grafo dei concetti ed estrarre la struttura gerarchica delle comunità.
Riassunto dell'indice	Utilizza un LLM per riassumere entità e relazioni in ogni comunità.	Nessuno l'approccio "lazy" rimanda tutto l'uso del LLM al momento della query.
Raffinamento della query	Nessuno la query originale viene usata per tutto il processo.	Utilizza un LLM per: a) identificare sottoquery rilevanti e ricombinarle in un'unica query espansa, b) affinare le sottoquery con concetti corrispondenti dal grafo dei concetti.
Corrispondenza della query	Nessuno tutte le query vengono risolte utilizzando tutti i riassunti delle comunità (ricerca in ampiezza).	Per ogni sottoquery $q[3,5]$: a) Classifica i chunk di testo usando embedding e relazioni chunk-comunità, prima in base alla similarità con la query e poi classificando le comunità in base ai chunk più rilevanti. b) Fa una valutazione sui chunk non ancora testati, valutandone la rilevanza (ricerca in ampiezza). c) Quando le ultime z comunità analizzate non hanno prodotto chunk di testo rilevanti inizia a fare ricerca di dati nelle sotto-comunità. d) Questa ricerca iterativa termina quando non ci sono più comunità di nessun livello da analizzare o quando si raggiunge il limite di budget impostato.
Mappatura delle risposte	Utilizza un LLM per rispondere alla query originale su batch casuali di riassunti delle comunità in parallelo.	Per ogni sottoquery $q[3,5]$: a) Costruisce un sottografo di concetti dai chunk di testo rilevanti. b) Utilizza un LLM per estrarre affermazioni rilevanti dalle sottoquery analizzando gruppi di chunk correlati. c) Classifica e filtra le affermazioni estratte per adattarle alla finestra di contesto predefinita.
Riduzione delle risposte	Utilizza un LLM per rispondere alla query originale utilizzando le risposte mappate.	Utilizza un LLM per rispondere alla query espansa utilizzando le affermazioni estratte dalla mappatura.

Table 3.2: Confronto tra GraphRAG e LazyGraphRAG. La tabella evidenzia le differenze nei processi di costruzione dell'indice, recupero delle informazioni e generazione delle risposte.

4

Sviluppo di un Sistema Multiagente per l'Analisi delle Distorsioni Cognitive

4.1 INTRODUZIONE

In questo capitolo viene presentato un esempio pratico di un sistema per la fase iniziale di **rilevazione di entità-relazioni**, come componente di un potenziale sistema *graphRAG*. A differenza della semplice comprensione di un testo generico, come quella offerta dalla soluzione presentata nel capitolo precedente (3), il lavoro qui descritto opera su di un **dominio specifico** ed estremamente complesso: il rilevamento delle distorsioni cognitive secondo la CBT. Questa specializzazione comporta sfide particolari per un sistema di intelligenza artificiale, in quanto riconoscere e interpretare le distorsioni cognitive a partire da un testo non è un compito banale per un LLM.

4.1.1 DISTORSIONI COGNITIVE IN AMBITO CBT

Un breve excursus sul **principio della CBT** può essere utile per inquadrare il contesto: in ambito psicologico, la CBT si basa sull'idea che le situazioni vissute da un individuo influenzino direttamente i suoi pensieri, i quali a loro volta possono generare distorsioni cognitive. Durante una seduta, lo psicologo ascolta attentamente il racconto del paziente, cercando di identificare quali eventi

4.2. IMPLEMENTAZIONE

abbiano innescato **pensieri disfunzionali**. Questi schemi cognitivi distorti, una volta resi consapevoli, possono essere corretti attraverso interventi mirati, favorendo un miglior adattamento emotivo e comportamentale.

4.1.2 IMPLEMENTAZIONE

Il sistema descritto in questo capitolo è strutturato in una architettura **multi-agente** basata su modelli LLM implementati con *langgraph*¹. Il processo si articola in **3 fasi**:

1. L'estrazione delle entità "*situazione*" e "*pensiero*" dal testo, ottenendo una prima rappresentazione strutturata dei dati.
2. La generazione di un grafo di conoscenza a partire da queste informazioni e la generazione di triplette che sintetizzano le connessioni causali tra le situazioni e i pensieri individuati.
3. L'analisi delle triplette per il rilevamento delle distorsioni cognitive.

4.1.3 EVALUATION

Successivamente, il sistema è stato valutato sperimentalmente utilizzando il dataset *llm cbt-bench*[46] (in particolare il file *distortion.json*). L'**evaluation** ha confrontato l'output del sistema con le distorsioni annotate all'interno del dataset. .

4.2 IMPLEMENTAZIONE

Qui di seguito saranno spiegate nel dettaglio le implementazioni dei tre agenti. Per ognuno di essi verrà mostrato l'output relativo all'input presente nel codice sottostante 4.1:

```
1 SITUATION_INPUT = ""I was born with Attention Deficit Hyperactivity Disorder, O.C.D. and I am also transgendered. As a result, I was a very difficult child to raise. There was constant fighting between myself and my parents. At the age of ten, I was sent away
```

¹<https://github.com/langchain-ai/langgraph>

```
to an institution for two years. While there, I was physically,  
emotionally, and sexually abused. I couldnt talk to anyone because  
I had no trust in adults.""  
2  
3 INPUT_THOUGHTS= ""I was born defective and wrong, so thats why my  
parents had such a hard time raising me. I was hard to love  
because I was so different. It was my fault that they had to send  
me away; I needed to be fixed. Then, I was abused while I was away  
but I couldnt talk to anyone about it because I was scared that  
if I said anything, Id be sent even further away.""
```

Code 4.1: esempio di dati di input per il sistema

4.2.1 AGENTE 1: DEFINIZIONE SCHEMA ENTITÀ-RELAZIONE

Il **primo agente** del sistema ha il compito di elaborare un input testuale contenente una lista di **situazioni** e una di **pensieri**, trasformandolo in uno **schema Entità-Relazioni (ER)**. L'estrazione automatica delle entità e delle relazioni a partire dai dati di input viene fatta usando un modello di linguaggio avanzato (*GPT-4o*)². Per fare questo, struttura le informazioni definendo diverse entità, che possono essere una "situazione" o un "pensiero", successivamente identifica eventuali relazioni causa-effetto tra una situazione e un pensiero.

Per definire una rappresentazione chiara e standardizzata segue la classe *EROutput*, definita tramite modello *Pydantic*³, che ha come attributi una lista delle entità e una lista delle relazioni rilevate. Questi due oggetti hanno i seguenti attributi:

Entità:

- **id**: stringa univoca che definisce una specifica entità.
- **type**: può essere *Thought* o *Situation* a seconda del tipo di entità rilevata
- **text**: la parte di testo ricevuta in input relativa a un pensiero o a una situazione.

Relazione:

- **from_id**: id della situazione che causa un certo pensiero

²<https://openai.com/index/hello-gpt-4o/>

³<https://docs.pydantic.dev/latest/concepts/models/>

4.2. IMPLEMENTAZIONE

- `to_id`: id del pensiero causato da una specifica entità.

Un esempio di output ottenuto è mostrato al codice 4.2:

```
1 Output:
2
3 entities=[
4 Entity(id='S1', type='Situation', text='I was born with Attention
   Deficit Hyperactivity Disorder, O.C.D. and I am also transgendered
   .'),
5 Entity(id='S2', type='Situation', text='As a result, I was a very
   difficult child to raise. '), Entity(id='S3', type='Situation',
   text='There was constant fighting between myself and my parents. '),
   Entity(id='S4', type='Situation', text='At the age of ten, I was
   sent away to an institution for two years. '),
6 Entity(id='S5', type='Situation', text='While there, I was physically
   , emotionally, and sexually abused. '),
7 Entity(id='S6', type='Situation', text='I couldnt talk to anyone
   because I had no trust in adults. '), Entity(id='T1', type='Thought
   ', text='I was born defective and wrong, so thats why my parents
   had such a hard time raising me. '),
8 Entity(id='T2', type='Thought', text='I was hard to love because I
   was so different. '),
9 Entity(id='T3', type='Thought', text='It was my fault that they had
   to send me away; I needed to be fixed. '),
10 Entity(id='T4', type='Thought', text='Then, I was abused while I was
   away but I couldnt talk to anyone about it because I was scared
   that if I said anything, Id be sent even further away.')]
11
12 relationships=[
13 Relationship(from_id='S1', to_id='T1'),
14 Relationship(from_id='S2', to_id='T1'), Relationship(from_id='S3',
   to_id='T1'), Relationship(from_id='S4', to_id='T3'), Relationship(
   from_id='S5', to_id='T4'), Relationship(from_id='S6', to_id='T4',]
```

Code 4.2: example of first agent's output

Questo passaggio consente di trasformare un testo non strutturato in una rappresentazione formale, facilitando le successive elaborazioni del sistema multi-agente. Il primo nodo gioca un ruolo fondamentale nel pipeline del sistema, poiché fornisce una base dati strutturata per l'analisi delle distorsioni cognitive e la generazione di risposte più precise nei nodi successivi.

4.2.2 AGENTE 2: INSERIMENTO DELLO SCHEMA ER NEL GRAFO E DEFINIZIONE DI TRIPLETTE

Il **secondo agente** si occupa di strutturare le informazioni ottenute dall'agente 1 in un **grafo**. Successivamente genera **triplette** di conoscenza basate sulle connessioni causali tra le situazioni e i pensieri individuati. Questo processo serve per permettere agli agenti successivi di analizzare ogni relazione causa-effetto *Situazione - Pensiero*, definendo se questa costituisce una distorsione cognitiva e, in quel caso, di che tipologia di distorsione si tratta.

Dopo aver popolato il grafo, l'agente esegue l'estrazione delle triplete. Una tripletta rappresenta una connessione tra **tre elementi**:

- Entità del tipo **situazione**: l'evento o il contesto da cui scaturisce un pensiero.
- Entità del tipo **pensiero**: la riflessione o convinzione generata dalla situazione.
- La **relazione** presente tra i due

Le triplete vengono memorizzate in una lista di variabili di tipo `Triplet`, classe definita sempre tramite modello *Pydantic*. La struttura delle triplete è definita come segue:

- `thought_id`: id del pensiero che partecipa alla tripletta.
- `triplet`: Il testo della tripletta è costruito concatenando il testo della situazione, il testo della relazione e il testo del pensiero in una frase leggibile in linguaggio naturale.

Un esempio di output ottenuto è mostrato al seguente codice 4.3

```
1 Output 2:
2
3 triplets=[
4 Triplet(thought_id='S1', triplet='"I was born with Attention Deficit
   Hyperactivity Disorder, O.C.D. and I am also transgendered.", "
   make me think that" "I was born defective and wrong, so thats why
   my parents had such a hard time raising me."'),
5 Triplet(thought_id='S2', triplet='"As a result, I was a very
   difficult child to raise.", "make me think that" "I was born
   defective and wrong, so thats why my parents had such a hard time
   raising me."'),
6 Triplet(thought_id='S3', triplet='"There was constant fighting
   between myself and my parents.", "make me think that" "I was
   born defective and wrong, so thats why my parents had such
```

4.2. IMPLEMENTAZIONE

```
7 a hard time raising me."'),
8 Triplet(thought_id='S4', triplet='"At the age of ten, I was sent away
  to an institution for two years.", "make me think that" "It was
  my fault that they had to send me away; I needed to be fixed."'),
9 Triplet(thought_id='S5', triplet='"While there, I was physically,
  emotionally, and sexually abused.", "make me think that" "Then,
  I was abused while I was away but I couldnt talk to anyone about
  it because I was scared that if I said anything, Id be sent even
  further away."'),
10 Triplet(thought_id='S6', triplet='"I couldnt talk to anyone because I
  had no trust in adults.", "make me think that" "Then, I was
  abused while I was away but I couldnt talk to anyone about it
  because I was scared that if I said anything, Id be sent even
  further away."')]
```

Code 4.3: example of second agent's output

4.2.3 AGENTE 3: ANALISI DELLE TRIPLETTE PER RILEVAMENTO DELLE DISTORSIONI

Il **terzo agente** si occupa di analizzare ogni tripletta generata dall'agente 2 al fine di identificare la presenza di **distorsioni cognitive**. Questo passaggio è essenziale per individuare schemi di pensiero distorti che servirà poi come punto di partenza per avere una base di conoscenza più approfondita per l'utente durante le successive conversazioni (tramite graph RAG) e utilizzabile dal sistema per stabilire il percorso di cura.

Tramite modello di linguaggio (*GPT 4o*), ogni tripletta viene utilizzata per classificare le distorsioni cognitive presenti. L'output atteso è una lista di oggetti *Distorsion*, sempre definito come modello *Pydantic*. Ciascuna distorsione ha i seguenti **attributi**:

- **name**: una delle seguenti categorie predefinite: "*all-or-nothing thinking*", "*overgeneralization*", "*mental filter*", "*mind reading*", "*fortune-telling*", "*magnification*", "*emotional reasoning*", "*should statements*", "*labeling*", "*personalization*".
- **explanation**: una breve descrizione, fornita dal modello, del motivo per cui la tripletta è stata classificata con quella specifica distorsione (serve per analizzare i risultati, permettendo di effettuare miglioramenti futuri nella pipeline, oltre che per poter avere la possibilità di capire eventuali errori e correggerli con intervento umano, come previsto dalle normative europee).

Un esempio di output ottenuto è mostrato nel codice 4.4

```
1 Output 3:
2
3 distortions=[
4 Distortion(name='labeling', explanation="The thought 'I was born
   defective and wrong' applies a negative label to oneself based on
   perceived shortcomings."), Distortion(name='personalization',
   explanation="The thought 'It was my fault that they had to send me
   away' assumes personal responsibility for being sent
5 to an institution, ignoring other factors."), Distortion(name='
   emotional reasoning', explanation="The belief 'I was born
   defective and wrong' is based on feelings of inadequacy rather
   than objective evidence."),
6 Distortion(name='mind reading', explanation="Assuming that parents
   had a hard time raising them because they were 'defective and
   wrong' without concrete evidence of their thoughts."),
7 Distortion(name='magnification', explanation="Exaggerating the impact
   of being transgender and having ADHD and OCD as being 'defective
   and wrong'.")]
```

Code 4.4: esempio di terzo output

4.2.4 UTILIZZO DEL PROMPT ENGINEERING NELLE FASI DI ESTRAZIONE E ANALISI

Nel sistema multi-agente descritto, il **prompt engineering** gioca un ruolo cruciale nel guidare il comportamento degli LLM durante le fasi di estrazione dello schema Entità-Relazioni (*agente 1*) e del rilevamento delle distorsioni cognitive (*agente 3*). I prompt hanno lo scopo di dare indicazioni al modello su come effettuare l'interpretazione dei dati ricevuti in input e a garantire un output nel formato richiesto ai passaggi successivi.

```
1 PROMPT_DIST_DEFINITIONS= """
2 ...
3 - **all-or-nothing thinking**:
4   *Description*: Seeing things in absolute, black-and-white
   categories, with no middle ground. If your performance falls short
   of perfect, you see yourself as a failure.
5 ...
```

6 """

Code 4.5: esempio di prompt per la definizione di una delle 10 distorsioni

Rilevamento schema ER. In particolare, per la fase di estrazione dello schema ER il prompt è strutturato in modo da:

- **Definire chiaramente l'input:** vengono specificati due campi distinti (*situation* e *thoughts*), dove gli elementi sono divisi tra loro mediante newline o punti.
- **Definizione delle relazioni:** il prompt istruisce l'LLM a creare relazioni di tipo *Triggers* solo quando è individuato un chiaro legame causa-effetto tra una situazione e un pensiero.
- **Imporre requisiti stringenti per l'output:** vengono fornite istruzioni esplicite per generare output in formato strutturato compatibile con quello del modello pydantic della classe *EROutput*

Rilevamento Distorsioni Per la fase di rilevamento delle distorsioni cognitive, il prompt è progettato per:

- **Fornire linee guida dettagliate:** vengono elencate le **10 categorie** di distorsioni accompagnate da brevi descrizioni che fungono da riferimento. Un esempio di queste definizioni è mostrato al codice 4.5
- **Analizzare le triplette:** il prompt istruisce l'LLM ad esaminare ciascuna tripletta, che collega una situazione a un pensiero tramite una relazione, e a confrontarla con le definizioni fornite.
- **Formattare l'output:** il risultato deve essere prodotto in formato strutturato, corrispondente al modello pydantic della classe *TripletsOutput*.

L'approccio complessivo combina una serie di tecniche di prompt engineering, quali:

- L'utilizzo di istruzioni chiare e dettagliate per limitare la variabilità dell'output.
- L'imposizione di vincoli formali nella struttura della risposta, per facilitare l'integrazione automatica nei moduli successivi.
- L'uso di esempi esplicativi, per guidare il modello nella corretta interpretazione del compito.

4.3 RISULTATI SPERIMENTALI

In questa sezione vengono presentati i **risultati sperimentali** ottenuti dal sistema proposto, valutato sul dataset *llm cbt-bench* [46]. In particolare, è stato utilizzato il file `distortion.json`, contenente **146 elementi**, ciascuno con la struttura illustrata nell'Esempio 4.6.

```

1 {
2   "id": "26",
3   "ori_text": "...",
4   "situation": "I have been in a relationship with my best
5     friend for five years...",
6   "thoughts": "He frequently questions my integrity..\\nHe
7     blames himself...",
8   "distortions": [
9     "all-or-nothing thinking",
10    "labeling",
11    "mind reading"
12  ]
13 }
```

Code 4.6: Esempio di struttura di un elemento nel file utilizzato per l'evaluation

Come mostrato nell'esempio, ogni elemento del dataset contiene le seguenti **informazioni**:

- **id**: un identificativo univoco dell'istanza.
- **ori_text**: il testo originale, in cui un paziente descrive i suoi problemi ad uno psicologo.
- **situation**: una lista delle parti del testo che descrivono un contesto o un evento che possono aver causato i problemi del paziente.
- **thoughts**: una lista di pensieri associati alla situazione.
- **distortions**: una lista di distorsioni cognitive, relative al discorso del paziente, utilizzate come ground truth.

4.3.1 DESCRIZIONE DELL'ESPERIMENTO

Nell' **esperimento** presentato in questo lavoro, per ogni elemento del dataset, sono stati estratti i campi `situation` e `thoughts` come *input* per il sistema multi-agente. Il sistema ha quindi generato, in **output**, un insieme di distorsioni

4.3. RISULTATI SPERIMENTALI

cognitive rilevate ("predictions"). Queste ultime sono state messe a confronto con le distorsioni annotate nel dataset (distortions).

La **valutazione** è stata condotta utilizzando l' **F1 score**[38], metrica ampiamente usata per classificazione multilabel.

4.3.2 CONFRONTO CON LO STATO DELL'ARTE

I **risultati** ottenuti dal nostro sistema, sono stati messi a confronto con quelli riportati nel paper di riferimento del benchmark *llm cbt-bench*[46]. Nella ricerca la valutazione veniva fatta utilizzando un prompt comprensivo dei dati *situation* e *thoughts* e la richiesta di rilevare tramite questi un massimo di tre distorsioni cognitive tra le 10 possibili. Hanno infine valutato l'output ottenuto da diversi modelli LLM utilizzando come metrica l'F1 score.

Nella Tabella 4.1 è riportato un estratto dei punteggi F1 (in %) per alcuni modelli, effettuati nella ricerca. Tra questi interessante è vedere i risultati della versione *GPT-4o* stand-alone, della nostra estensione multi-agente e, come valore massimo di riferimento, la performance umana.

Model	F1 Score (%)
Mistral-v0.3-7B	45.3
Gemma-2-9B	40.1
Llama-3.1-8B	44.1
Llama-3.1-70B	43.9
Llama-3.1-405B	43.3
GPT-4o	43.9
GPT-4o + multiagent system	46.1
<i>Human</i>	49.1

Table 4.1: Confronto dei risultati (F1 Score) tra diversi modelli sul dataset *llm cbt-bench*[46].

Dall'**analisi dei punteggi** emerge che la soluzione multi-agente, basata su *GPT-4o*, ottiene un F1 score pari a 46.1%, superando il semplice *GPT-4o* (43.9%). Benché il *gap* rispetto alla performance umana (49.1%) rimanga, il risultato dimostra che l'approccio multi-agente consente di **migliorare** la capacità di riconoscere le distorsioni cognitive nel testo, grazie alla suddivisione del processo in più fasi (estrazione di entità e relazioni, generazione di triplette, analisi delle distorsioni), all'utilizzo di tecniche di prompt-engineering più specifiche e all'integrazione di conoscenza strutturata.

4.3.3 ANALISI DEI RISULTATI

L'**incremento** di 2.2 punti percentuali rispetto a *GPT-4o* indica che la soluzione proposta migliora la precisione e la recall complessive. Il modulo di estrazione di entità e relazioni fornisce una base semantica più ricca per la classificazione delle distorsioni, mentre la creazione di triplette aiuta a evidenziare i collegamenti causali tra situazioni e pensieri, favorendo un'analisi più approfondita. Nonostante permanga un margine di miglioramento per raggiungere le performance umane, i risultati confermano la validità di un approccio multi-agente in contesti di riconoscimento di distorsioni cognitive.

4.4 POSSIBILI SVILUPPI FUTURI

Nonostante i risultati ottenuti siano promettenti, l'obiettivo finale di un sistema realmente **utilizzabile** nella pratica clinica richiede l'analisi diretta del testo originale (*ori_text*) prodotto dal paziente, anziché una versione già parzialmente analizzata e suddivisa in *situations* e *thoughts*. Tale sfida si rivela notevolmente più complessa, come risortrato in alcuni studi recenti che propongono tecniche di **prompt-engineering** [8] e **approcci multi-agente avanzati** [28] per affrontare l'interpretazione di testi non strutturati.

Un primo prototipo di tale metodo, evoluzione del sistema ideato e presentato in questo lavoro di tesi, è stato testato sull'analisi del parametro *ori_text* ottenendo un F1 score pari a 27.99, un risultato ancora distante da valori realmente utili nella pratica clinica. Tuttavia, questo divario è comprensibile alla luce delle numerose difficoltà incontrate nell'interpretazione del testo originale, tra cui:

- **La complessità della segmentazione del testo completo:** l'analisi di *thoughts* e *situation* già identificati e separati risulta decisamente più semplice rispetto all'estrazione di queste informazioni da un testo non pre-processato, in cui la suddivisione in entità significative è meno chiara e più variabile.
- **Ambiguità e riferimenti impliciti:** le persone spesso omettono contesti o utilizzano metafore e sottintesi, rendendo più ardua l'estrazione automatica di entità e relazioni.
- **Possibile presenza di lessico colloquiale o informale:** in un dialogo reale, il paziente potrebbe usare termini non standard, intercalari o frammenti di frasi.

4.4. POSSIBILI SVILUPPI FUTURI

Il risultato ottenuto (27.99) è di un ordine di grandezza simile rispetto alle due ricerche citate in questa sessione, le quali impiegano però dataset diversi e con un numero di esempi differente, rendendo pertanto poco significativi i confronti diretti. Rimane il fatto che l'analisi del testo originale richiede ulteriori strategie di disambiguazione, segmentazione del discorso e modellazione del contesto, aspetti che rappresentano importanti aree di ricerca futura.

Un ulteriore sviluppo riguarda l'estensione del sistema verso un'architettura RAG tramite grafo più completa. In un contesto terapeutico, il grafo costruito dalle entità e dalle relazioni verrebbe sfruttato come *knowledge base* dinamica, interrogabile durante il dialogo con il paziente. In questo modo, l'agente sarebbe in grado di:

- **Recuperare informazioni rilevanti** già emerse in sessioni precedenti, evitando ridondanze e favorendo la continuità del percorso terapeutico.
- **Definire strategie d'intervento personalizzate**, basate su conoscenze strutturate relative ai problemi e alle distorsioni già identificate.
- **Monitorare i progressi nel tempo**, tracciando l'evoluzione dei pensieri e delle situazioni del paziente e confrontandoli con i risultati di sessioni passate.

Questa prospettiva **apre la strada** a un sistema sempre più adatto a integrare il potenziale dei modelli di linguaggio con la robustezza di un grafo di conoscenza, per offrire un supporto avanzato sia al terapeuta sia al paziente.

Bibliografia

- [1] I.M. Alqahtani et al. "Prevalence of depression, anxiety, and stress among the general population during COVID-19 pandemic: A systematic review". In: *J Family Med Prim Care* 12.6 (2023), pp. 1030–1037. doi: 10.4103/jfmpc.jfmpc_1943_22.
- [2] Nasser Saad Salem Alqahtani et al. "Artificial intelligence and social intelligence: preliminary comparison study between AI models and psychologists". In: *Frontiers in Psychology* 15 (2024). doi: 10.3389/fpsyg.2024.1353022. URL: <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2024.1353022/full>.
- [3] L.H. Andrade et al. "Barriers to mental health treatment: results from the WHO World Mental Health surveys". In: *Psychol Med* 44.6 (2014), pp. 1303–1317. doi: 10.1017/S0033291713001943.
- [4] Jinze Bai et al. *Qwen-VL: A Versatile Vision-Language Model for Understanding, Localization, Text Reading, and Beyond*. 2023. arXiv: 2308.12966 [cs.CV]. URL: <https://arxiv.org/abs/2308.12966>.
- [5] AT Beck. "Thinking and depression. I. Idiosyncratic content and cognitive distortions". In: *Arch Gen Psychiatry* 9 (1963), pp. 324–33.
- [6] Tom Brown et al. "Language Models are Few-Shot Learners". In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle et al. Vol. 33. Curran Associates, Inc., 2020, pp. 1877–1901. URL: https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf.
- [7] Ilias Chalkidis et al. "LEGAL-BERT: The Muppets straight out of Law School". In: *Findings of the Association for Computational Linguistics: EMNLP 2020*. Ed. by Trevor Cohn, Yulan He, and Yang Liu. Online: Association for Computational Linguistics, 2020, pp. 2898–2904. doi: 10.18653/v1/

BIBLIOGRAFIA

- 2020.findings-emnlp.261. URL: <https://aclanthology.org/2020.findings-emnlp.261/>.
- [8] Zhiyu Chen, Yujie Lu, and William Wang. “Empowering Psychotherapy with Large Language Models: Cognitive Distortion Detection through Diagnosis of Thought Prompting”. In: Jan. 2023, pp. 4295–4304. DOI: 10.18653/v1/2023.findings-emnlp.284.
- [9] Aakanksha Chowdhery et al. *PaLM: Scaling Language Modeling with Pathways*. 2022. arXiv: 2204.02311 [cs.CL]. URL: <https://arxiv.org/abs/2204.02311>.
- [10] S. Clement et al. “What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies”. In: *Psychological Medicine* 45.1 (2015), pp. 11–27. DOI: 10.1017/S0033291714000129.
- [11] Darren Edge et al. *From Local to Global: A Graph RAG Approach to Query-Focused Summarization*. 2024. arXiv: 2404.16130 [cs.CL]. URL: <https://arxiv.org/abs/2404.16130>.
- [12] European Commission. *Artificial Intelligence Act Explorer*. Accessed: 2025-02-14. 2021. URL: <https://artificialintelligenceact.eu/ai-act-explorer/>.
- [13] European Commission. *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on a comprehensive approach to mental health*. 2023. URL: https://www.eumonitor.eu/9353000/1/j4nvhd fcs8bljza_j9vvik7m1c3gyxp/vm3r868xoqzg (visited on 02/18/2025).
- [14] Wensheng Gan et al. “Large Language Models in Education: Vision and Opportunities”. In: *2023 IEEE International Conference on Big Data (Big-Data)* (2023), pp. 4776–4785. URL: <https://api.semanticscholar.org/CorpusID:265352038>.
- [15] Aaron Grattafiori et al. *The Llama 3 Herd of Models*. 2024. arXiv: 2407.21783 [cs.AI]. URL: <https://arxiv.org/abs/2407.21783>.
- [16] Daya Guo et al. *DeepSeek-Coder: When the Large Language Model Meets Programming – The Rise of Code Intelligence*. 2024. arXiv: 2401.14196 [cs.SE]. URL: <https://arxiv.org/abs/2401.14196>.

- [17] Bernal Jiménez Gutiérrez et al. *HippoRAG: Neurobiologically Inspired Long-Term Memory for Large Language Models*. 2025. arXiv: 2405.14831 [cs.CL]. URL: <https://arxiv.org/abs/2405.14831>.
- [18] Aidan Hogan et al. "Knowledge Graphs". In: (2021). URL: <https://doi.org/10.1145/3447772>.
- [19] Yuntong Hu et al. *GRAG: Graph Retrieval-Augmented Generation*. 2024. arXiv: 2405.16506 [cs.LG]. URL: <https://arxiv.org/abs/2405.16506>.
- [20] Minki Kang et al. *Knowledge Graph-Augmented Language Models for Knowledge-Grounded Dialogue Generation*. 2023. arXiv: 2305.18846 [cs.CL]. URL: <https://arxiv.org/abs/2305.18846>.
- [21] Luoma Ke et al. *Exploring the Frontiers of LLMs in Psychological Applications: A Comprehensive Review*. 2024. arXiv: 2401.01519 [cs.LG]. URL: <https://arxiv.org/abs/2401.01519>.
- [22] Mina J. Kian et al. *Can an LLM-Powered Socially Assistive Robot Effectively and Safely Deliver Cognitive Behavioral Therapy? A Study With University Students*. 2024. arXiv: 2402.17937 [cs.R0]. URL: <https://arxiv.org/abs/2402.17937>.
- [23] Jiho Kim et al. *FactKG: Fact Verification via Reasoning on Knowledge Graphs*. 2023. arXiv: 2305.06590 [cs.CL]. URL: <https://arxiv.org/abs/2305.06590>.
- [24] Patrick Lewis et al. "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks". In: *ArXiv abs/2005.11401* (2020). URL: <https://api.semanticscholar.org/CorpusID:218869575>.
- [25] Dawei Li et al. "DALK: Dynamic Co-Augmentation of LLMs and KG to answer Alzheimers Disease Questions with Scientific Literature". In: Jan. 2024, pp. 2187–2205. DOI: 10.18653/v1/2024.findings-emnlp.119.
- [26] Yuhan Li et al. *A Survey of Graph Meets Large Language Model: Progress and Future Directions*. 2024. arXiv: 2311.12399 [cs.LG]. URL: <https://arxiv.org/abs/2311.12399>.
- [27] Lei Liang et al. *KAG: Boosting LLMs in Professional Domains via Knowledge Augmented Generation*. 2024. arXiv: 2409.13731 [cs.CL]. URL: <https://arxiv.org/abs/2409.13731>.

BIBLIOGRAFIA

- [28] Sehee Lim et al. "ERD: A Framework for Improving LLM Reasoning for Cognitive Distortion Classification". In: Jan. 2024, pp. 292–300. DOI: 10.18653/v1/2024.clinicalnlp-1.25.
- [29] Weijie Liu et al. *K-BERT: Enabling Language Representation with Knowledge Graph*. 2019. arXiv: 1909.07606 [cs.CL]. URL: <https://arxiv.org/abs/1909.07606>.
- [30] Shengjie Ma et al. *Think-on-Graph 2.0: Deep and Faithful Large Language Model Reasoning with Knowledge-guided Retrieval Augmented Generation*. 2025. arXiv: 2407.10805 [cs.CL]. URL: <https://arxiv.org/abs/2407.10805>.
- [31] Yingzhuo Ma et al. "Integrating large language models in mental health practice: a qualitative descriptive study based on expert interviews". In: *Frontiers in Public Health* 12 (2024). URL: <https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2024.1475867>.
- [32] Lucian H. Milasan and Daniel Scott-Purdy. "The Future of Artificial Intelligence in Mental Health Nursing Practice: An Integrative Review". In: *International Journal of Mental Health Nursing* 34.1 (2025). e70003 IJMHN-2024-1011.R1, e70003. DOI: <https://doi.org/10.1111/inm.70003>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/inm.70003>.
- [33] OpenAI et al. *GPT-4 Technical Report*. 2024. arXiv: 2303.08774 [cs.CL]. URL: <https://arxiv.org/abs/2303.08774>.
- [34] Tobias A. Opsahl. *Fact or Fiction? Improving Fact Verification with Knowledge Graphs through Simplified Subgraph Retrievals*. 2024. arXiv: 2408.07453 [cs.CL]. URL: <https://arxiv.org/abs/2408.07453>.
- [35] Valerie du Preez et al. "From bias to black boxes: understanding and managing the risks of AI an actuarial perspective". In: *British Actuarial Journal* 29 (2024). DOI: 10.1017/S1357321724000060.
- [36] Microsoft Research. *LazyGraphRAG: Setting a New Standard for Quality and Cost*. <https://www.microsoft.com/en-us/research/blog/lazygraphrag-setting-a-new-standard-for-quality-and-cost/>. 2024.
- [37] Gerard Salton, Anita Wong, and Chung-Shu Yang. "A vector space model for automatic indexing". In: *Communications of the ACM* 18.11 (1975), pp. 613–620.

- [38] Yutaka Sasaki. "The truth of the F-measure". In: *Teach Tutor Mater* (Jan. 2007).
- [39] Jonas Schuett. "Risk Management in the Artificial Intelligence Act". In: *European Journal of Risk Regulation* 15.2 (2024), pp. 367–385. DOI: 10.1017/err.2023.1.
- [40] Hao Shen et al. "Are Large Language Models Possible to Conduct Cognitive Behavioral Therapy?" In: *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 2024, pp. 3695–3700. DOI: 10.1109/BIBM62325.2024.10821773.
- [41] Amit Singhal. *Introducing the Knowledge Graph: things, not strings*. Google Blog. 2012. URL: <https://googleblog.blogspot.com/2012/05/knowledge-graph-things-not-strings.html>.
- [42] V. Traag, L. Waltman, and Nees Jan van Eck. "From Louvain to Leiden: guaranteeing well-connected communities". In: *Scientific Reports* 9 (Mar. 2019), p. 5233. DOI: 10.1038/s41598-019-41695-z.
- [43] Supra Wimbarti, B. H. Ralph Kairupan, and Trina Ekawati Tallei. "Critical review of self-diagnosis of mental health conditions using artificial intelligence". In: *International Journal of Mental Health Nursing* 33.2 (2024), pp. 344–358. DOI: <https://doi.org/10.1111/inm.13303>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/inm.13303>.
- [44] J. Xiong et al. "Impact of COVID-19 pandemic on mental health in the general population: A systematic review". In: *J Affect Disord* 277 (2020), pp. 55–64. DOI: 10.1016/j.jad.2020.08.001.
- [45] Yi Yang, Mark Christopher Siy Uy, and Allen Huang. "FinBERT: A Pre-trained Language Model for Financial Communications". In: *ArXiv abs/2006.08097* (2020). URL: <https://api.semanticscholar.org/CorpusID:219687757>.
- [46] Mian Zhang et al. *CBT-Bench: Evaluating Large Language Models on Assisting Cognitive Behavior Therapy*. 2025. arXiv: 2410.13218 [cs.CL]. URL: <https://arxiv.org/abs/2410.13218>.
- [47] Weijia Zhang et al. "Beyond Relevant Documents: A Knowledge-Intensive Approach for Query-Focused Summarization Using Large Language Models". In: Dec. 2024, pp. 89–104. ISBN: 978-3-031-78494-1. DOI: 10.1007/978-3-031-78495-8_6.

BIBLIOGRAFIA

- [48] Lingfeng Zhong et al. "A Comprehensive Survey on Automatic Knowledge Graph Construction". In: *ACM Computing Surveys* 56 (Sept. 2023). doi: 10.1145/3618295.