



Università Politecnica Delle Marche

Facoltà di Ingegneria

CORSO DI LAUREA MAGISTRALE IN INGEGNERIA MECCANICA,
SISTEMI MECCANICI E DELL'AUTOMAZIONE

**Sviluppo e Testing di interfacce per l'analisi avanzata di dati
strutturati e non strutturati attraverso prompt engineering
nell'ambito dell'ingegneria meccanica**

**Development and Testing of interfaces for advanced analysis of
structured and unstructured data through prompt engineering
in the field of mechanical engineering**

Relatore:

Prof. Adriano Mancini

Candidato: 1100650

Andrea Evangelista

ANNO ACCADEMICO 2022 / 2023

Sommario

Questa tesi esplora lo sviluppo e l'applicazione di tecniche avanzate per l'analisi di documenti strutturati e non, sottolineando l'importanza del prompt engineering nell'era del Big Data e dell'intelligenza artificiale nell'ambito dell'ingegneria meccanica. I nuovi approcci ai dati possono essere implementati nei contesti aziendali esistenti attraverso l'utilizzo di interfacce dedicate, le quali rappresentano il mezzo principale per consentire la comunicazione e l'interazione tra i dati e i team di analisi dedicati. Con l'adozione di queste tecniche, si fornisce ai team un progresso notevole, consentendo un aumento di semplicità, velocità e efficienza di analisi. Tali benefici sono da ricondurre alla combinazione dell'intelligenza artificiale e con la scienza del prompt engineering, poiché consente di evitare errori dovuti alla componente umana e accelerare allo stesso tempo le fasi di studio di BigData.

La tesi fornisce una dettagliata descrizione dello stato dell'arte e delle possibilità concrete che possono essere colte con l'adozione di tale metodologia. Durante la presentazione verranno illustrati, oltre alla parte teorica, vari test effettuati per illustrare le possibilità di utilizzo, analizzare le potenzialità e i limiti di questi approcci. L'elaborato inizia con una panoramica dello stato dell'arte del *prompt engineering*, evidenziando le sfide e le opportunità legate al trattamento di informazioni eterogenee sotto forma strutturata e non. Successivamente, si illustrano gli strumenti e i metodi innovativi utilizzati nelle analisi per poi passare allo studio pratico dei dati in entrambe le forme con specifici casi di studio e applicazioni, dove le tecniche

di *prompt engineering* sono state impiegate per migliorare l'efficacia e l'efficienza nell'analisi di documenti. Si esaminano dettagliatamente le strategie di progettazione dei prompt, i metodi per valutare la loro efficacia e le sfide associate alla loro implementazione in ambienti reali. Infine, la tesi si conclude con lo studio dei risultati ottenuti e una riflessione critica sugli impatti etici e sociali dell'uso di queste tecnologie, proponendo una serie di linee guida e migliori pratiche per il loro sviluppo e utilizzo responsabile. L'obiettivo è fornire una risorsa comprensiva che non solo presenti soluzioni tecniche avanzate, ma che stimoli anche una discussione più ampia sull'importanza del design consapevole e centrato sull'utente nell'era dell'informazione.

Indice

1	Introduzione	7
1.1	Industria 4.0	7
1.1.1	Dati strutturati	8
1.1.2	Dati non strutturati	9
1.2	Presentazione del progetto di tesi	10
2	Stato dell'arte	12
2.1	Smart factories	12
2.2	Intelligenza artificiale	14
2.3	Analisi di dati	15
2.4	Prompt engineering	18
2.4.1	Impostazioni del Large Language Model	19
2.4.2	I fondamentali	20
2.4.3	Elementi di un prompt	21
2.4.4	Suggerimenti per un prompt efficace	21
2.4.5	Possibili utilizzi	23
2.4.6	Tecniche di prompting	25
3	Strumenti e Metodi	35
3.1	Descrizione del progetto	35
3.2	Analisi dei dati non strutturati	36
3.2.1	AskYourPDF	36
3.2.2	ChatDOC	37
3.2.3	ChatPDF	38

3.2.4	LightPDF	38
3.2.5	ChatGPT 3.5	39
3.2.6	ChatGPT 4	39
3.3	Analisi dei dati strutturati	40
3.3.1	PandasAI	41
3.3.2	ChatGPT 4	42
4	Approccio e analisi dei dati	44
4.1	Selezione dei dati non strutturati da analizzare	44
4.2	Preparazione dei dati per l'analisi	45
4.3	Selezione dei dati strutturati da analizzare	46
4.4	Preparazione dei dati per l'analisi	47
4.4.1	Presentazione progetto	47
4.4.2	Database progetto	48
4.4.3	Preparazione del database per l'analisi con PandasAI	48
4.4.4	Preparazione del database per l'analisi con ChatGPT4	51
5	Risultati	55
5.1	Caso di studio 1: Analisi di dati non strutturati	55
5.1.1	Risultati test AskYourPDF	56
5.1.2	Risultati test ChatPDF	64
5.1.3	Risultati test ChatGPT-4, dati non strutturati	72
5.2	Caso di studio 2: Analisi di dati strutturati	78
5.2.1	Risultati test PandasAI	78
5.2.2	Risultati test ChatGPT-4, dati strutturati	83
6	Conclusione	92
6.1	Sintesi dei risultati ottenuti	92
6.2	Sviluppi futuri	93
6.3	Considerazioni generali finali	95
	Bibliografia	102

Capitolo 1

Introduzione

In questo capitolo presenteremo una panoramica approfondita del contesto industriale caratteristico degli anni correnti. In secondo luogo verrà presentato, in maniera sintetica, il progetto di tesi.

1.1 Industria 4.0

L'industria 4.0 rappresenta una fase evolutiva significativa nel settore industriale, caratterizzata dall'integrazione di tecnologie avanzate come l'Internet delle Cose (IoT), la robotica, la realtà virtuale e aumentata, e ovviamente, l'Intelligenza Artificiale (AI). Questo paradigma si basa fortemente sulla raccolta, l'analisi e l'uso efficiente dei dati per ottimizzare i processi, migliorare la produttività e sostenere la presa di decisioni basate su evidenze concrete. In questo contesto, la gestione efficace dei dati è fondamentale. I big data sono volumi enormi di dati che possono essere analizzati computazionalmente per rivelare pattern, tendenze e associazioni. Si distinguono principalmente due categorie diverse di dati: *strutturati* e *non strutturati*.

- *Dati Strutturati*: Sono dati altamente organizzati e facilmente interpretabili da macchine. Solitamente si trovano in database relazionali.
- *Dati Non-Strutturati*: Sono dati non organizzati o in un formato non predefinito, rendendo più complesse le operazioni di analisi. Esempi

includono dati provenienti da fonti come email, video, dati di social media o, più genericamente, file PDF, word, presentazioni ecc...

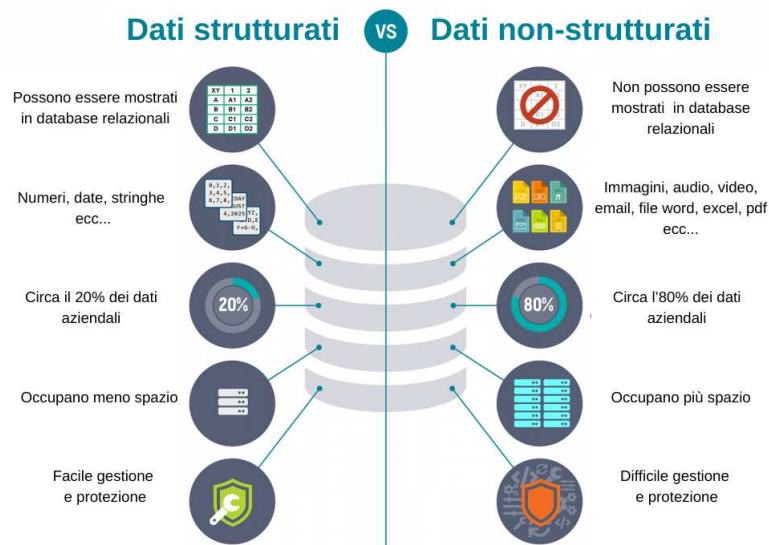


Figura 1.1: Fonte: <https://vitolvecchia.altervista.org/differenza-tra-dati-strutturati-non-strutturati-e-semi-strutturati/>

1.1.1 Dati strutturati

I dati *strutturati* sono dati che hanno un formato standardizzato per un accesso efficiente da parte del software e degli utenti. In genere, si tratta di tabelle con righe e colonne che definiscono chiaramente gli attributi dei dati. Tra le tipologie più diffuse di questa famiglia di dati si individuano, in particolare, i file Excel, i Database SQL, i dati e i sistemi di macchinari. I computer sono in grado di elaborare efficacemente i dati strutturati per ottenere approfondimenti grazie alla loro natura quantitativa. Ad esempio, una tabella strutturata di dati sui clienti di una azienda contenente colonne come nome, indirizzo e numero di telefono può fornire informazioni come il numero totale di clienti e la località con il numero massimo di clienti. Al contrario, i dati non strutturati, come un elenco di post sui social media,

sono più difficili da analizzare. Le caratteristiche principali di questi dati sono:

- Attributi definibili e relazionali: i dati hanno gli stessi attributi per tutti i valori e sono collegati tra loro da chiavi comuni,
- Quantitativi: i dati si prestano bene ad ogni tipo di analisi, specialmente quella matematica,
- Archiviazione: i dati possono essere archiviati in database relazionali e gestiti con un linguaggio SQL (Structured Query Language)

Essendo, per definizione, strutturati consistono allo stesso tempo sia a numerosi vantaggi che svantaggi; la loro struttura predefinita e fissa è senza dubbio utile per la rapida analisi ma anche limitante, permettendo di essere usati solo per il loro scopo principale. C'è, quindi, una mancanza di flessibilità che corrisponde alla sfida principale quando si parla di dati strutturati.

1.1.2 Dati non strutturati

I dati *non strutturati* non hanno un modello predefinito e non possono essere organizzati in righe e colonne, mancano cioè di una definizione. Possono essere memorizzati in molti moduli per poi essere memorizzati come oggetti di vario genere: immagini, audio, video, file di documenti e altri formati di file. L'elemento comune con tutti i dati di questa tipologia è una mancanza di definizione. Data la vasta gamma di formati che comprende questa tipologia di dati, non sorprende che questo tipo di dati costituisca in genere circa l'80% dei dati di un'organizzazione. I dati non strutturati possono, però, avere origini molto diverse: estratti da un linguaggio umano con NLP (Natural Language Processing), acquisiti attraverso sensori, estratti dai social media, acquisiti da database NoSQL. Questa caratteristica ne rende difficile la comprensione e ambigua la collocazione. Inoltre, tendono a essere molto più grandi e occupano volumi molto superiori rispetto ai

dati strutturati arrivando anche a scale del petabyte. Per queste ragioni gli strumenti e le tecniche che si sono dimostrati così efficaci nel trasformare dati strutturati in business non funzionano quando si tratta di dati non strutturati. Sotto questa spinta, l'utilizzo dell'intelligenza artificiale e del prompt engineering potrebbe essere una soluzione valida e un ottimo metodo per riuscire a lavorare efficientemente con questo tipo di dati.

1.2 Presentazione del progetto di tesi

Per le ragioni espresse in precedenza, si è scelto di sviluppare questo progetto, che mira ad analizzare lo stato attuale del prompt engineering nell'ambito dell'ingegneria meccanica e introdurre delle tecniche che permettano di lavorare in modo efficiente con le tipologie di dati illustrati in precedenza. Naturalmente, è fondamentale individuare, prima di iniziarci a lavorare, la tipologia di dato che si ha davanti. Una volta introdotti i concetti teorici, si passerà ad una fase pratica per testare l'effettiva potenzialità e i limiti attuali dei tool disponibili per l'analisi dei dati. L'analisi iniziale del prompt engineering è uno step importante che fa guadagnare la capacità di formulare correttamente le richieste per interagire nel modo più proficuo con i dati con cui si deve lavorare, al fine di non incorrere in rallentamenti e/o intoppi. In questa fase verranno introdotte dapprima nozioni di base, per permettere una comprensione completa dell'ambito di studio, per poi passare a concetti avanzati come tecniche di prompting moderne e innovative. Una volta illustrati e resi chiari i meccanismi e le tecniche per un prompt engineering vincente si passerà al lavoro pratico con i dati, approcciando prima i dati *non strutturati* e infine quelli *strutturati*, affrontando svariate soluzioni presenti attualmente basate sul modello di intelligenza artificiale, sviluppato da OpenAI, ChatGPT nella sua versione gratuita e premium, rispettivamente ChatGPT 3.5 e ChatGPT 4.

Capitolo 2

Stato dell'arte

In questo capitolo sarà illustrato l'attuale stato dell'arte dell'industria 4.0, ponendo un focus sull'intelligenza artificiale, sul *prompt engineering* e sull'analisi dei *BigData*, al fine di introdurre il lettore alle tecnologie di cui si farà uso nei successivi capitoli di approccio agli strumenti ed analisi.

2.1 Smart factories

Al giorno d'oggi, ci troviamo nel mezzo della quarta rivoluzione industriale, rappresentata dalla tendenza crescente di implementazione di tecnologie produttive per il raggiungimento dell'automazione industriale. Questo obiettivo principale, che ha portato allo sviluppo delle smart factories moderne, si raggiunge adottando il paradigma di industria 4.0. Questo si focalizza sull'ottimizzazione dei processi produttivi attraverso l'automazione, l'interconnessione e l'analisi intelligente dei dati. L'obiettivo è creare sistemi di produzione intelligenti (smart factories) capaci di auto-regolarsi, di prendere decisioni autonome e di essere altamente personalizzabili, automatizzati e flessibili.

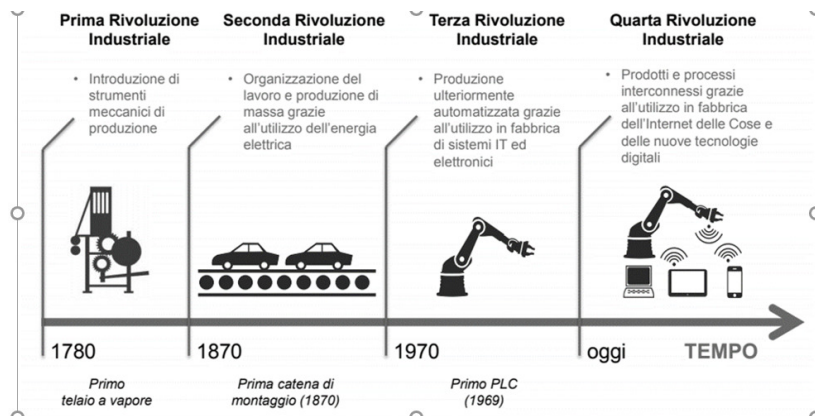


Figura 2.1: Fonte: <https://www.economyup.it/innovazione/cos-e-l-industria-40-e-perche-e-importante-saperla-affrontare/>

Il paradigma di industria 4.0 si basa su nove pilastri, ciascuno rappresentante un aspetto cruciale:

1. **Internet delle Cose (IoT):** Consiste nell'interconnessione di macchinari e dispositivi di produzione tramite internet, consentendo loro di raccogliere, scambiare e analizzare dati in tempo reale per ottimizzare i processi,
2. **Big Data e Analisi dei Dati:** Implica l'uso di tecnologie avanzate per analizzare grandi quantità di dati raccolti dai dispositivi IoT. Questo permette di identificare schemi, prevedere tendenze e prendere decisioni basate su dati concreti,
3. **Sistemi Cyber-Fisici (CPS):** Sono sistemi che integrano meccanismi di elaborazione computazionale con capacità fisiche. Nei CPS, i componenti software e hardware lavorano insieme per monitorare e controllare attività nel mondo fisico,
4. **Automazione e Robotica:** L'uso di robot e sistemi di automazione avanzati, spesso dotati di intelligenza artificiale, per eseguire compiti con precisione, efficienza e sicurezza superiore rispetto agli standard umani,

5. **Manifattura Additiva:** Permette la produzione di parti e componenti attraverso la sovrapposizione di materiali, layer per layer. Questo favorisce una produzione personalizzata, riduce gli sprechi e accorcia i tempi di produzione,
6. **Cloud Computing:** Fornisce le risorse necessarie per elaborare e immagazzinare enormi quantità di dati, permettendo anche l'utilizzo di software e applicazioni attraverso la rete, riducendo così i costi operativi,
7. **Realtà Aumentata (AR) e Realtà Virtuale (VR):** Queste tecnologie supportano l'interazione avanzata tra l'uomo e la macchina, fornendo simulazioni immersive e supporto visivo per la formazione, la manutenzione e il monitoraggio dei processi,
8. **Cybersecurity:** Con l'aumento dell'interconnessione e della dipendenza dai dati, diventa essenziale proteggere le infrastrutture critiche e i dati sensibili da attacchi informatici e altre minacce alla sicurezza.

Questi pilastri lavorano in sinergia per trasformare l'industria tradizionale in una più flessibile, reattiva e efficiente. L'Industria 4.0 non è solo un cambiamento tecnologico; rappresenta anche una trasformazione culturale, richiedendo un nuovo approccio alla gestione, alla formazione del personale e alla strategia aziendale.

2.2 Intelligenza artificiale

Come spiegato nel paragrafo precedente, questa era si distingue dalle precedenti per l'integrazione di tecnologie avanzate come l'IoT, il cloud computing, e il Big Data. L'IoT è il fondamento dell'Industria 4.0, che fornisce la connettività e la raccolta dati necessarie; collega macchine, dispositivi, sensori e persone, creando una rete di dispositivi intelligenti che comunicano tra loro e con i sistemi centrali. Questi dispositivi generano una quantità

massiccia di dati, che sono la materia prima per l'analisi e le decisioni. È evidente dunque la necessità di strumenti affidabili e efficienti che permettano la loro analisi. Si può allora introdurre l'intelligenza artificiale come il naturale sviluppo evolutivo dell'IoT. Mentre l'IoT è in grado di raccogliere e trasmettere dati, l'IA è ciò che permette di trasformare questi dati in insight, decisioni e azioni. L'IA analizza i dati raccolti dai dispositivi IoT, apprende da questi dati, e prende decisioni intelligenti. In questo senso, l'IA può essere vista come la "figlia" dell'IoT: nasce dall'esigenza di dare significato e azione ai dati raccolti. Queste capacità apre un mondo di nuove possibilità in ambito industriale. Ad esempio, nella manifattura, sistemi AI possono prevedere guasti prima che accadano (manutenzione predittiva), ottimizzare la catena di produzione in tempo reale, e personalizzare la produzione. Questo non solo aumenta l'efficienza e riduce i costi, ma migliora anche la sicurezza e la sostenibilità.

2.3 Analisi di dati

Un altro tema centrale nell'evoluzione dell'industria moderna è l'analisi dei *BigData*, che comporta un'importanza vitale di una gestione accurata e strategica dei dati. Viviamo in un'era in cui la quantità di dati generati ogni secondo rompe ogni precedente record storico. Questa esplosione informativa, conosciuta comunemente come *BigData*, è la linfa vitale dell'Industria 4.0. La fusione senza precedenti tra tecnologie fisiche, digitali e biologiche, ha portato a una produzione di dati senza precedenti. Le aziende che sanno navigare in questo mare di informazioni sono quelle che non solo sopravvivono ma prosperano nel mercato competitivo di oggi. L'analisi dei Big Data nell'Industria 4.0 non è solo una questione di gestire enormi volumi di dati, ma di trasformare questa abbondanza informativa in intuizioni preziose, in decisioni strategiche e in azioni innovative. Questo processo richiede strumenti sofisticati, competenze avanzate e, soprattutto, una visione strategica. Questi dati, infatti, se opportunamente analizzati

e gestiti, possono tradursi in una comprensione approfondita dei processi aziendali, prevedere le tendenze di mercato, personalizzare l'esperienza del cliente e ottimizzare le operazioni. Tuttavia, la gestione non ottimale di questi dati può non solo far perdere opportunità vitali ma può anche esporre le aziende a rischi significativi. Ci sono, infatti, degli aspetti che devono necessariamente caratterizzare i dati. Innanzitutto, la qualità deve essere elevata: è essenziale assicurarsi che i dati siano accurati, completi e pertinenti. Poi, la sicurezza deve essere solida: in un'epoca in cui le violazioni dei dati sono all'ordine del giorno, proteggere le informazioni sensibili è fondamentale. Inoltre, devono essere integrabili: in un mondo connesso, la capacità di integrare dati provenienti da fonti diverse è chiave per ottenere una visione completa. I *BigData* sono un importante strumento per affrontare una serie di attività aziendali, dalla customer experience agli analytics. Ecco alcuni esempi di queste attività:

- **Sviluppo del prodotto:** aziende come Netflix e Procter & Gamble usano i *BigData* per anticipare la domanda dei clienti. Costruiscono modelli predittivi per nuovi prodotti e servizi classificando attributi chiave di prodotti o servizi passati e attuali. Inoltre, P&G utilizza i dati e gli analytics dei focus group, dei social media, dei test di mercato e dei primi rollout nei negozi per pianificare, produrre e lanciare nuovi prodotti,
- **Manutenzione predittiva:** i fattori che possono prevedere guasti meccanici possono essere nascosti tra i dati strutturati, come l'anno, la marca e il modello dell'attrezzatura, oppure nei dati non strutturati che coprono milioni di voci di registro, dati dei sensori, messaggi di errore e temperatura del motore. Analizzando queste indicazioni di potenziali problemi prima che si verifichino, le organizzazioni possono implementare la manutenzione in modo più efficiente in termini di costi e massimizzare i tempi di attività di parti e apparecchiature,

- **Customer Experience:** i *BigData* consentono di raccogliere dati da social media, visite web, registri delle chiamate e altre fonti per migliorare l'esperienza di interazione e massimizzare il valore fornito. Inizia a fornire offerte personalizzate, riduci il tasso di abbandono dei clienti e gestisci i problemi in modo proattivo,
- **Frode e compliance:** i *BigData* aiutano a identificare i modelli nei dati che indicano frodi e ad aggregare grandi volumi di informazioni per rendere i rapporti normativi molto più veloci,
- **Machine Learning:** è un argomento molto dibattuto in questo momento. La disponibilità di *BigData* per addestrare modelli di machine learning è una pratica in rapida diffusione,
- **Efficienza operativa:** è un'area in cui i *BigData* stanno avendo il maggior impatto. Con il loro utilizzo, è possibile analizzare e valutare la produzione, il feedback e i resi dei clienti e altri fattori per ridurre le interruzioni e anticipare le richieste future. I *BigData* possono essere utilizzati anche per migliorare il processo decisionale in linea con l'attuale domanda di mercato,
- **Promuovi l'innovazione:** i *BigData* aiutano a innovare studiando le interdipendenze tra esseri umani, istituzioni, entità e processi e quindi determinando nuovi modi per utilizzare tali insight. Si utilizzano gli insight sui dati per migliorare le decisioni su considerazioni finanziarie e di pianificazione; si esaminano le tendenze e cosa vogliono i clienti per fornire nuovi prodotti e servizi. Ci sono infinite possibilità.

Ora, mentre l'analisi dei Big Data ci apre porte verso possibilità inimmaginabili, ci si di fronte a una nuova sfida: come navigare in questo mare vasto e profondo in modo efficiente ed efficace? La risposta risiede in un concetto rivoluzionario: il *Prompt Engineering*. Nel paragrafo che segue verrà discusso questo nuovo approccio in forte evoluzione.

2.4 Prompt engineering

Il *Prompt Engineering* è una disciplina relativamente nuova che consente di sviluppare e ottimizzare i prompt per utilizzare in modo efficiente i modelli linguistici (LM) per un'ampia varietà di applicazioni e argomenti di ricerca. Le competenze del *Prompt Engineering* aiutano a comprendere meglio le capacità e i limiti dei modelli linguistici di grandi dimensioni (LLM). Chi si occupa di ricerca usa questo approccio per migliorare le capacità dei LLM su un'ampia gamma di compiti comuni e complessi, come la risposta a domande e il ragionamento aritmetico. Chi sviluppa software invece per progettare tecniche di prompt robuste ed efficaci che si interfacciano con i LLM e con altri strumenti. Il *Prompt Engineering* non riguarda solo la progettazione e lo sviluppo di prompt ma comprende anche un'ampia gamma di competenze, tecniche ed utili per interagire e sviluppare con i LLM. È un'abilità importante per interfacciarsi, costruire e comprendere le capacità dei LLM. È possibile utilizzare l'ingegneria dei prompt per migliorare la sicurezza dei LLM e costruire nuove funzionalità, come l'aumento dei LLM con la conoscenza del dominio e gli strumenti esterni. Il *Prompt Engineering*, dunque, non è solo un approccio, è una vera e propria arte. È l'arte di comunicare con le intelligenze artificiali in modo tale da massimizzare la qualità e la pertinenza delle informazioni ottenute. In un'epoca in cui l'intelligenza artificiale svolge un ruolo cruciale nell'analisi dei dati, saper interagire con queste tecnologie in modo avanzato e sofisticato diventa un vantaggio competitivo inestimabile. Attraverso tecniche di *Prompt Engineering*, possiamo affinare le nostre richieste ai sistemi basati su IA, ottenendo risposte più accurate, insight più profondi e, in definitiva, decisioni aziendali più informate. È un ponte tra l'umano e l'artificiale, un modo per armonizzare la nostra curiosità e creatività con la potenza di calcolo e l'efficienza delle macchine. E in questo scenario, il *Prompt Engineering* emerge come uno strumento essenziale, che ci permette di interagire con

le intelligenze artificiali in modo sempre più efficace, trasformando i dati in oro puro. Investire in queste competenze non è solo una scelta, ma una necessità strategica per qualsiasi azienda che aspiri al successo nell'era dell'informazione.

2.4.1 Impostazioni del Large Language Model

Quando si lavora con i prompt, si interagisce con il LLM tramite un'API o direttamente. È possibile configurare alcuni parametri per ottenere risultati diversi per i prompt:

- **Temperature:** In breve, più bassa è la temperatura, più deterministici sono i risultati, nel senso che viene sempre scelto il token successivo più probabile. Un aumento della temperatura potrebbe portare a una maggiore casualità, incoraggiando risultati più diversi o creativi. In sostanza, stiamo aumentando il peso degli altri possibili token. In termini di applicazione, potremmo voler utilizzare un valore di temperatura più basso per compiti come l'AQ basata sui fatti, per incoraggiare risposte più concrete e concise. Per la generazione di poesie o altri compiti creativi, potrebbe essere utile aumentare il valore della temperatura,
- **Top_p:** Allo stesso modo, con top_p, una tecnica di campionamento con temperatura chiamata campionamento del nucleo, si può controllare quanto il modello sia deterministico nel generare una risposta. Se siete alla ricerca di risposte esatte e concrete, tenete questo valore basso. Se si cercano risposte più diversificate, aumentare il valore a un valore più alto.

La raccomandazione generale è di modificarne solo uno dei due valori e non entrambi. In generale, è da tenere presente anche che i risultati ottenuti nelle applicazioni possono variare a seconda della versione di LLM che state utilizzando.

2.4.2 I fondamentali

Si può ottenere molto con semplici prompt, ma la qualità dei risultati dipende da quante informazioni vengono fornite e da quanto sono ben fatte. Un prompt può contenere informazioni come l'istruzione o la domanda che si sta passando al modello e includere altri dettagli come contesto, ingressi o esempi. Si possono usare questi elementi per istruire meglio il modello e di conseguenza ottenere risultati migliori. Di seguito viene illustrato un esempio basilare:

```
Prompt
Il cielo è

Output
blu
Il cielo è blu nelle giornate serene. In una giornata nuvolosa, il cielo può essere grigio o bianco.
```

Figura 2.2: Esempio basilare di un prompt.

In questo caso, il modello linguistico produce una serie di stringhe che hanno senso in base al contesto "Il cielo è". L'output potrebbe essere inaspettato o lontano dal compito che vogliamo svolgere. Con questo esempio si evince la necessità di fornire più informazioni in input con il prompt, come un contesto ampio o dettagli aggiuntivi sull'istruzione da voler far compiere al LLM, per ricevere l'output desiderato. Un esempio di miglioramento è il seguente:

```
Prompt
Completare la frase:
Il cielo è

Output
così bello oggi.
```

Figura 2.3: Esempio basilare di un prompt.

Avendo detto al modello di completare la frase il risultato è molto più corretto perché segue esattamente ciò che gli abbiamo detto di fare, "completare la frase". Questo approccio, che consiste nel progettare prompt ottimali per istruire il modello a eseguire un compito, è definito *prompt engineering* (ingegneria dei prompt). L'esempio precedente è un'illustrazione di base di ciò che è possibile fare oggi con i LLM. Quelli odierni, sono in grado di eseguire tutti i tipi di compiti avanzati, dalla sintesi del testo al ragionamento matematico alla generazione di codice.

2.4.3 Elementi di un prompt

Man mano che si approfondiscono gli esempi e le applicazioni possibili con il *prompt engineering*, si può notare la presenza di elementi a composizione di un prompt. Questo, infatti, può contenere i seguenti componenti:

- **Istruzione:** un compito o un'istruzione specifica che il modello deve eseguire,
- **Contesto:** può coinvolgere informazioni esterne o contesti aggiuntivi che possono indirizzare il modello verso risposte migliori,
- **Dati in Input:** è l'input o la domanda per la quale ci interessa trovare una risposta,
- **Indicatore dell'Output:** indica il tipo o il formato dell'output.

Non tutti i componenti sono necessari per un prompt e il formato dipende dal compito da svolgere. Successivamente, verranno illustrati con più dettaglio.

2.4.4 Suggerimenti per un prompt efficace

Rendere efficace un prompt è lo scopo principale di chi utilizza il *prompt engineering*. Tra gli aspetti più importanti che contribuiscono a raggiungere l'obiettivo si trovano:

- **Inizio semplice:** quando si inizia a progettare i prompt, bisogna tenere presente che si tratta di un processo iterativo che richiede molta sperimentazione per ottenere risultati ottimali. Si può iniziare con prompt semplici e aggiungere altri elementi e contesti man mano che si punta a ottenere risultati migliori. Per questo motivo è fondamentale modificare i suggerimenti durante il percorso. Quando si ha un compito importante che coinvolge molte sotto-attività diverse, si può cercare di suddividere il compito in sotto-attività più semplici e continuare ad aumentare man mano che si ottengono risultati migliori. In questo modo si evita di aggiungere troppa complessità al processo di progettazione del prompt all'inizio,
- **Le istruzioni:** è possibile progettare prompt efficaci per vari compiti semplici, utilizzando comandi per indicare al modello ciò che si desidera ottenere, come "Scrivi", "Classifica", "Riassumi", "Traduci", "Ordina", ecc. È da tenere presente che è importante anche sperimentare molto per capire cosa funziona meglio. Possono essere provate diverse istruzioni con parole chiave, contesti e dati diversi per vedere cosa funziona meglio per il caso d'uso e le attività. Di solito, più il contesto è specifico e rilevante per l'attività che si sta cercando di svolgere, meglio è,
- **Specificità:** c'è la necessità di essere molto specifici riguardo all'istruzione e al compito che si desidera far eseguire al modello. Più la richiesta è descrittiva e dettagliata, migliori saranno i risultati. Questo è particolarmente importante quando si cerca un risultato o uno stile di generazione desiderato. Infatti, non esistono token o parole chiave specifiche che portano a risultati migliori. È più importante avere un buon formato e un prompt descrittivo. Proprio per questo, fornire esempi nel prompt è molto efficace per ottenere i risultati desiderati in formati specifici. Quando si progettano i prompt, si deve anche tenere presente la lunghezza del prompt, poiché ci sono dei li-

miti per quanto riguarda la sua lunghezza. Bisogna pensare a quanto sia necessario essere specifici e dettagliati. Includere troppi dettagli non necessari non è necessariamente un buon approccio. I dettagli devono essere pertinenti e contribuire al compito da svolgere.

Di seguito è riportato un semplice prompt per estrarre informazioni specifiche da un testo.

```
Prompt:

Estrarre il nome dei luoghi nel seguente testo.
Formato desiderato:
Luogo: <lista_dei_nomi_delle_cittade_separata_da_virgola>
Input: "Sebbene questi sviluppi siano incoraggianti per i ricercatori, molto rimane ancora un mistero. "Spesso abbiamo una scatola nera tra il cervello e l'effetto che vediamo in periferia", dice Henrique Veiga-Fernandes, neuroimmunologo presso il Centro Champalimaud per l'ignoto di Lisbona. "Se vogliamo utilizzarla nel contesto terapeutico, dobbiamo effettivamente comprenderne il meccanismo"".

Output:

Luogo: Centro Champalimaud per l'ignoto, Lisbona".
```

Figura 2.4: Esempio sulla specificità di un prompt.

- **Evitare l'imprecisione:** alla luce dei suggerimenti precedenti sull'essere dettagliati e migliorare il formato, è facile cadere nella trappola di voler essere troppo intelligenti con i suggerimenti e creare descrizioni imprecise. Spesso è meglio essere specifici e diretti. L'analogia è molto simile a quella della comunicazione efficace: più è diretta, più il messaggio è efficace,
- **Fare o non fare?:** un altro aspetto da tenere a mente quando si progettano i prompt è quello di evitare di dire cosa non fare, ma di dire invece cosa fare. Questo incoraggia una maggiore specificità e si concentra sui dettagli che portano a buone risposte da parte del modello.

2.4.5 Possibili utilizzi

Le opportunità di utilizzo del *prompt engineering* sono molteplici. In questa sezione si illustrano i diversi compiti che i prompt sono in grado di svolgere.

- **Sintesi del testo:** uno dei compiti standard nella generazione del linguaggio naturale è la sintesi del testo. Il riassunto del testo può includere molti aspetti e domini diversi. In effetti, una delle applicazioni più promettenti dei modelli linguistici è la capacità di riassumere articoli e concetti in sintesi rapide e di facile lettura,
- **Estrazione di informazioni:** sebbene i modelli linguistici siano addestrati per eseguire la generazione del linguaggio naturale e i compiti correlati, sono anche in grado di eseguire la classificazione e una serie di altri compiti di elaborazione del linguaggio naturale (NLP, Natural Language Processing),
- **Rispondere alle domande:** uno dei modi migliori per far sì che il modello risponda a prompt specifici è migliorare il formato del prompt. Come già detto, un prompt può combinare istruzioni, contesto, indicatori di input e output per ottenere risultati migliori. Anche se questi componenti non sono obbligatori, si tratta di una buona pratica, poiché quanto più specifiche sono le istruzioni, tanto migliori saranno i risultati,
- **Classificazione del testo:** per i casi d'uso più difficili, fornire solo istruzioni non è sufficiente. Bisogna infatti pensare di più al contesto e ai diversi elementi che si possono usare in un prompt. Altri elementi che si possono fornire sono i dati di input o gli esempi,
- **Conversazione:** forse una delle cose più interessanti che si possono ottenere con il prompt engineering è istruire il sistema LLM su come comportarsi, sulle sue intenzioni e sulla sua identità. Questo è particolarmente utile quando si costruiscono sistemi conversazionali come i chatbot del servizio clienti,
- **Generazione di codice:** un'applicazione in cui gli LLM sono molto efficaci è la generazione di codice. È possibile eseguire un gran numero di operazioni di generazione di codice con suggerimenti intelligenti,

- **Ragionamento:** forse uno dei compiti più difficili per un LLM oggi è quello che richiede una qualche forma di ragionamento. Il ragionamento è una delle aree più articolate che possono far emergere applicazioni complesse dai LLM. Ci sono stati, ad esempio, alcuni miglioramenti nei compiti che richiedono capacità matematiche. Detto questo, è importante notare che gli attuali LLM faticano a svolgere compiti di ragionamento, quindi questo richiede tecniche di prompt engineering ancora più avanzate.

2.4.6 Tecniche di prompting

È molto importante migliorare i prompt per ottenere risultati migliori su compiti diversi. Questa, infatti, è l'idea alla base del prompt engineering. Nel corso di questi anni sono state sviluppate svariate tecniche di prompting, ognuna con lo scopo di affinare l'interazione tra l'utente e i sistemi di intelligenza artificiale. Questo capitolo si propone di esplorare queste metodologie, analizzando come ciascuna di esse possa influenzare e guidare il comportamento dei modelli AI. Dalle strategie di formulazione di prompt semplici e diretti a quelle più complesse e strutturate, il lettore sarà introdotto a un mondo dove la precisione nella comunicazione si traduce in risposte più accurate e in un'esperienza utente notevolmente migliorata.

1. **Prompt Zero-Shot:** oggi i LLM, addestrati su grandi quantità di dati e regolati per seguire le istruzioni, sono in grado di eseguire compiti a colpo zero, cioè richiesta e risposta in un colpo solo senza step aggiuntivi. La sintonizzazione delle istruzioni ha dimostrato di migliorare l'apprendimento a colpo zero, Wei et al. (2022)[1]. La sintonizzazione delle istruzioni è essenzialmente il concetto di perfezionamento dei modelli su insiemi di dati descritti tramite istruzioni. Inoltre, l'apprendimento per rinforzo dal feedback umano è stato adottato per scalare la sintonizzazione delle istruzioni, in cui il modello viene allineato per adattarsi meglio alle preferenze umane. Questo

recente sviluppo alimenta modelli come ChatGPT. Quando zero-shot non funziona, si raccomanda di fornire dimostrazioni o esempi nel prompt, il che porta al few-shot prompt,

2. **Prompt Few-Shot:** può essere utilizzato come tecnica, quando il prompt zero-shot non basta, per consentire l'apprendimento in contesto, in cui vengono fornite dimostrazioni nel prompt per guidare il modello verso prestazioni migliori. Le dimostrazioni servono come condizionamento per gli esempi successivi in cui vorremmo che il modello generasse una risposta. In base alla complessità della richiesta si può usare un solo esempio o molti di più (1-shot, 2-shot...10-shot ecc. . .). Il prompt standard di few-shot funziona bene per molti compiti, ma non è ancora una tecnica perfetta, soprattutto quando si tratta di compiti di ragionamento più complessi, dove potrebbe essere utile scomporre il problema in fasi e dimostrarlo al modello. Più recentemente, il è stato diffuso per affrontare compiti più complessi di ragionamento aritmetico, di senso comune e simbolico,
3. **Chain-of-Thought (CoT) Prompt:** introdotto da Wei et al. (2022)[1], il prompt a catena di pensieri (CoT) consente di ottenere capacità di ragionamento complesse attraverso fasi di ragionamento intermedie. Si può combinare con il prompt few-shot per ottenere risultati migliori su compiti più complessi che richiedono un ragionamento prima di rispondere.

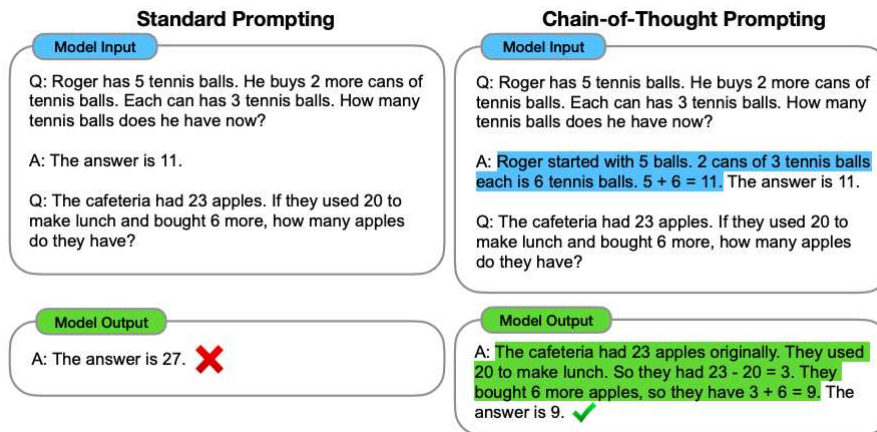


Figura 2.5: Fonte: <https://www.promptingguide.ai/it>

Un'idea emersa più di recente è quella della zero-shot CoT (Kojima et al. 2022)[2] che consiste essenzialmente nell'aggiungere "Pensiamo passo dopo passo" al prompt originale,

4. **Self-Consistency**: una delle tecniche più avanzate per l'ingegneria del prompt è l'autoconsistenza. Proposta da Wang et al. (2022)[3], l'autoconsistenza mira a "sostituire l'ingenua decodifica utilizzata nel prompt Chain of Thought". L'idea è quella di campionare percorsi di ragionamento multipli e diversificati attraverso la CoT Few-Shot e di utilizzare le generazioni per selezionare la risposta più coerente. Questo aiuta a migliorare le prestazioni della richiesta di CoT su compiti che coinvolgono l'aritmetica e il ragionamento di senso comune,
5. **Prompt Generated Knowledge**: una tecnica popolare per migliorare i LLM include la capacità di incorporare conoscenze o informazioni per aiutare il modello a fare previsioni più accurate. Utilizzando un'idea simile, è possibile utilizzare il modello anche per generare conoscenza prima di fare una previsione, come illustrato nell'articolo di Liu et al. (2022)[4]: si può e si deve generare conoscenza da utilizzare come parte del prompt.

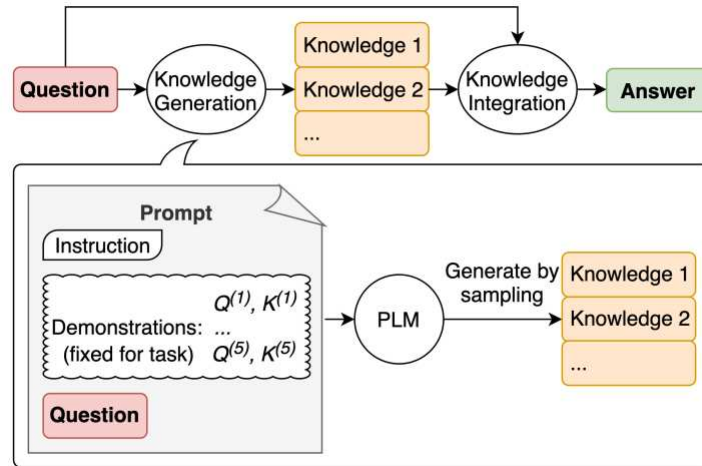


Figura 2.6: Fonte: <https://www.promptingguide.ai/it>

6. **Tree of Thoughts (ToT):** Yao et al. (2023)[5] e Long (2023)[6] hanno proposto Tree of Thoughts (ToT), una generalizzazione della tecnica di chain-of-thought prompting che incoraggia l'esplorazione di più "pensieri" che vengono utilizzati come step intermedi per effettuare problem solving con i Language Model (LM). La tecnica ToT crea un albero di pensieri, dove i pensieri sono sequenze linguistiche che rappresentano i passi per raggiungere la risoluzione del problema. Questo approccio permette ad un LM di valutare i suoi stessi progressi intermedi verso la risoluzione del problema. L'abilità del LM di generare e valutare i "pensieri" viene combinata con algoritmi di ricerca.

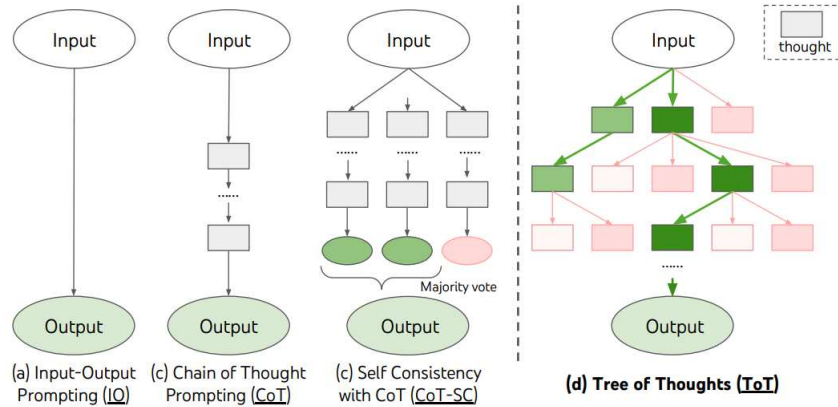


Figura 2.7: Fonte: <https://www.promptingguide.ai/it>

Ad alto livello, le principali idee di Yao et al. (2023)[5] e Long (2023)[6] potenziano entrambe le capacità dei Large Language Model di risolvere compiti complessi utilizzando una ricerca su albero con una conversazione a più giri. Una delle differenze principali sta nelle strategie di ricerca utilizzate: Yao et al. (2023) utilizza algoritmi generici di ricerca come DFS/BFS/beam search, mentre la strategia di ricerca proposta da Long (2023) è controllata da un "ToT Controller", addestrato utilizzando il reinforcement learning (RL). Infine, Hulbert (2023)[7] ha proposto la tecnica di Tree-of-Thought Prompting, che applica il concetto principale della tecnica ToT utilizzando un singolo prompt testuale,

7. Retrieval Augmented Generation (RAG): introdotta da Meta, riguarda i modelli di linguaggio di uso generale, che possono essere perfezionati per svolgere diversi compiti comuni come l'analisi del sentimento e il riconoscimento delle entità nominate. Questi compiti generalmente non richiedono ulteriori conoscenze di base. Per compiti più complessi e che richiedono molte conoscenze, è possibile costruire un sistema basato su un modello di linguaggio che accede a fonti di conoscenza esterne per completare i compiti. Questo consente una maggiore coerenza dei fatti, migliora l'affidabilità delle

risposte generate e aiuta a mitigare il problema dell'"allucinazione". I ricercatori di Meta AI hanno introdotto un metodo chiamato Generazione aumentata dal recupero (RAG) per affrontare tali compiti che richiedono molte conoscenze. Questa tecnica combina un componente di recupero delle informazioni con un modello generatore di testo. Infatti, durante il suo utilizzo il LLM prende un input e recupera un insieme di documenti rilevanti/supportativi da una fonte. I documenti vengono concatenati come contesto con il prompt di input originale e inviati al generatore di testo che produce l'output finale. Questo rende RAG adattivo per situazioni in cui i fatti potrebbero evolvere nel tempo. Questo è molto utile poiché la conoscenza parametrica dei LLM è statica,

8. **Automatic Prompt Engineer (APE)**: Zhou et al., (2022)[8] propone automatic prompt engineer (APE), un framework per la generazione e la selezione automatica delle istruzioni. Il problema della generazione delle istruzioni viene inquadrato come sintesi del linguaggio naturale e affrontato come un problema di ottimizzazione black-box che utilizza gli LLM per generare e ricercare le soluzioni candidate. La prima fase coinvolge un modello linguistico di grandi dimensioni (come modello di inferenza) a cui vengono fornite dimostrazioni in uscita per generare istruzioni candidate per un compito. Queste soluzioni candidate guidano la procedura di ricerca. Le istruzioni vengono eseguite utilizzando un modello di destinazione e poi viene selezionata l'istruzione più appropriata in base ai punteggi di valutazione calcolati. Questa ricerca tratta un argomento importante legato al prompt engineering, ovvero l'idea di ottimizzare automaticamente i prompt,
9. **Active-prompt**: i metodi della catena di pensiero (CoT) si basano su un insieme fisso di esempi annotati manualmente. Il problema è che gli esempi potrebbero non essere i più efficaci per i diversi com-

piti. Per ovviare a questo problema, Diao et al., (2023)[9] hanno recentemente proposto un nuovo approccio di prompting, chiamato Active-Prompt, per adattare i LLM a diversi prompt di esempi specifici per i vari compiti. Il primo passo consiste nell'interrogare il LLM con o senza alcuni esempi di CoT. Vengono generate k possibili risposte per un insieme di domande di addestramento. Sulla base delle k risposte viene calcolata una metrica di incertezza. Le domande più incerte vengono selezionate manualmente per l'annotazione e i nuovi esempi annotati vengono poi utilizzati per dedurre ogni domanda,

10. **Prompt Directional Stimulus:** Li et al., (2023)[10] propone una nuova tecnica di prompting per guidare meglio il LLM nella generazione della sintesi desiderata. Per generare il suggerimento viene addestrato un LM "sintonizzabile". Si vede un maggiore uso di apprendimento rinforzato per ottimizzare i LLM,
11. **ReAct:** Yao et al.,(2022)[11] ha introdotto un framework in cui i LLM vengono utilizzati per generare sia tracce di ragionamento che azioni specifiche dell'attività in modo intercalato. La generazione di tracce di ragionamento consente al modello di indurre, tenere traccia e aggiornare i piani d'azione e persino di gestire le eccezioni. La fase di azione consente di interfacciarsi e raccogliere informazioni da fonti esterne come basi di conoscenza o ambienti. Il framework ReAct può consentire ai LLM di interagire con strumenti esterni per recuperare informazioni aggiuntive che portano a risposte più affidabili e concrete,
12. **Multimodal CoT Prompt:** Zhang et al. (2023)[12] ha recentemente proposto un approccio multimodale di suggerimento a catena di pensiero. Il CoT tradizionale si concentra sulla modalità linguistica. Al contrario, Multimodal CoT incorpora testo e visione in un quadro a due fasi. Il primo passo prevede la generazione di motivazio-

ni basate su informazioni multimodali. Questa è seguita dalla seconda fase, l'inferenza della risposta, che sfrutta le motivazioni informative generate.

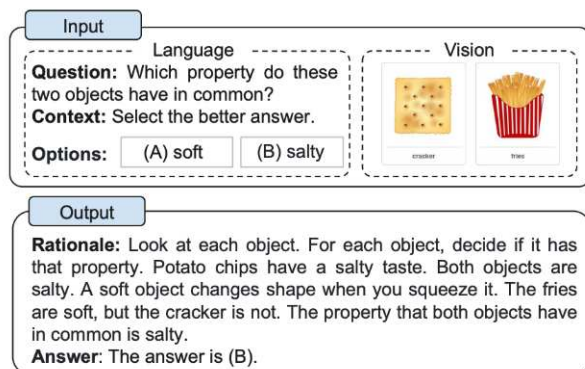


Figura 2.8: Fonte: <https://www.promptingguide.ai/it>

13. **Optimization by PROMpting (OPRO):** Google DeepMind ha recentemente presentato un metodo innovativo per richiedere risposte AI denominato Optimization by PROMpting (OPRO)[13]. Questo metodo utilizza modelli linguistici di grandi dimensioni (LLM) come ottimizzatori in cui i modelli di intelligenza artificiale funzionano tentando diversi prompt finché non trovano quello che si avvicina di più alla risoluzione di un particolare compito. A differenza dei metodi tradizionali che si basano su definizioni matematiche formali, OPRO definisce le attività di ottimizzazione in linguaggio naturale. I ricercatori hanno evidenziato: "Invece di definire formalmente il problema di ottimizzazione... descriviamo il problema di ottimizzazione in linguaggio naturale, quindi istruiamo il LLM a generare in modo iterativo nuove soluzioni basate sulla descrizione del problema e sulle soluzioni trovate in precedenza.". Questa adattabilità consente al LLM di affrontare una vasta gamma di problemi semplicemente modificando la descrizione del problema o aggiungendo istruzioni specifiche.

Tutte queste tecniche, dunque, sono le armi del prompt engineer per riuscire ad ottenere risposte il più efficienti possibili. I capitoli successivi mostrano come il *prompt engineering* può essere usato per raggiungere gli obiettivi desiderati senza incorrere in problemi e/o rallentamenti e mostra anche i limiti attuali di questo approccio, concentrandosi prima sui dati di tipo non strutturato per poi passare a quelli di tipo strutturato.

Capitolo 3

Strumenti e Metodi

In questo capitolo verranno descritti i vari strumenti utilizzati per le prove tecniche svolte con i relativi metodi, in modo da preparare il lettore al capitolo successivo in cui verranno illustrati in modo completo tutti i test svolti.

3.1 Descrizione del progetto

Come accennato nel capitolo introduttivo, l'obiettivo di questa tesi è quello di esplorare lo sviluppo e l'applicazione di tecniche avanzate per l'analisi di documenti strutturati e non, sottolineando l'importanza del prompt engineering nell'era del *BigData* e dell'intelligenza artificiale. Per questo motivo, per entrambe le tipologie di dato, prima di effettuare i vari test, è stata fatta una selezione dei tool presenti allo stato dell'arte in grado di poter svolgere l'analisi di dati. Con questo approccio, è possibile valutare quale strumento svolga il lavoro nel modo migliore e di conseguenza scegliere quale adottare per le proprie esigenze. Di seguito verranno illustrati prima i tool per l'analisi di dati *non strutturati* per poi passare a quelli relativi all'analisi di dati *strutturati*.

3.2 Analisi dei dati non strutturati

Riprendendo quanto anticipato nel capitolo introduttivo 1.1.2, i dati *non strutturati* non hanno un modello predefinito e non possono essere organizzati in righe e colonne, mancano cioè di una definizione. Questa caratteristica ne rende difficile la comprensione. L'utilizzo dell'intelligenza artificiale e del *prompt engineering* potrebbe essere una soluzione valida e un ottimo metodo per riuscire a lavorare efficientemente con questo tipo di dati. In particolare, sono svariate le possibilità quando si parla di strumenti di analisi dati non strutturati, ognuno basato su uno specifico LLM e sfruttato tramite utilizzo di API. Per questo studio sono stati scelti tool che sfruttano il LLM sviluppato dall'azienda statunitense OpenAI, ChatGPT 3.5, essendo uno dei modelli più completi disponibili, nonché il più conosciuto e diffuso al momento insieme alla sua versione più recente ChatGPT 4 utilizzata per l'analisi di dati strutturati. È importante sottolineare che sul mercato esistano altri modelli, anche con un a potenza più elevata rispetto a quella di ChatGPT 4, appartenenti a colossi aziendali, come Gemini di Google o AzureAI di Microsoft e altri, ma al momento il modello di intelligenza artificiale più studiato e affidabile rimane ChatGPT e per questo motivo è stato selezionato per questo progetto di tesi. Anticipando ciò che si trova nel capitolo 4 4.2, per ogni tool si andranno a svolgere dei test per mettere alla prova lo strumento sotto analisi sia in base alle capacità che in base alla lingua utilizzata, per valutare poi l'efficienza di ognuno e le differenze con gli altri. Segue ora una carrellata dei tool utilizzati per l'analisi di dati *non strutturati*.

3.2.1 AskYourPDF

AskYourPDF[14] è uno strumento basato, come detto in precedenza, sul modello di intelligenza artificiale ChatGPT 3.5. Uno dei migliori vantaggi dell'utilizzo di questo tool è la sua capacità di gestire ed elaborare facilmen-

te non solo documenti PDF ma anche molti altri formati. Grazie a questo aspetto è possibile analizzare la maggior parte dei dati non strutturati testuali. Una volta caricato il file, è possibile porre domande e recuperare le informazioni necessarie. AskYourPDF, dunque, agisce come un assistente virtuale, consentendo di caricare senza problemi i vostri PDF e di rispondere alle vostre domande.



Figura 3.1: Logo AskYourPDF.

3.2.2 ChatDOC

ChatDOC[15] è un assistente di lettura molto diffuso che sfrutta l'intelligenza artificiale, può estrarre, individuare e riassumere rapidamente le informazioni dai documenti. Il sito, sfruttando la potenza di ChatGPT 3.5, è in grado di comprendere i testi, le tabelle e i dati all'interno. Lo strumento non si limita a ripetere le informazioni presenti nel documento, ma le interpreta e le sintetizza. Inoltre, è possibile caricare più file in simultanea o addirittura cartelle con documenti e le risposte alle domande sono supportate da citazioni dirette estratte dai file, che si possono controllare cliccando sul link controllare l'accuratezza dell'interpretazione dell'IA.



Figura 3.2: Logo ChatDOC.

3.2.3 ChatPDF

Il principio di funzionamento di questo tool è lo stesso di tutti gli altri ma permette di caricare un solo file per volta e offre una scelta limitata di formati supportati. Tra tutte le opzioni disponibili, ChatPDF[16], risulta essere la più basilare.



Figura 3.3: Logo ChatPDF.

3.2.4 LightPDF

LightPDF[17] offre in realtà una vasta gamma di strumenti in grado di lavorare con PDF, dalla loro conversione alla loro modifica. Uno di questi strumenti prende il nome di "AI for Docs" ed è lo strumento utile per questo test. Permette, come gli altri, di poter lavorare con svariati formati e di ottenere output più elaborati come tabelle o schemi. Inoltre, anche in questo caso è possibile lavorare con più di un documento per volta per poter svolgere uno studio completo.



Figura 3.4: Logo LightPDF.

3.2.5 ChatGPT 3.5

È possibile utilizzare direttamente ChatGPT 3.5[18], essendo la nuova versione a pagamento, per poter lavorare con documenti non strutturati. A differenza degli altri tool, però, non permette di caricare file. Tuttavia, questo limite può essere aggirato grazie al prompt engineering, inserendo un particolare prompt che mette in ascolto il modello e lo prepara a ricevere tutte le informazioni testuali appartenenti al documento. Il rovescio della medaglia è l'ottenimento di risposte superficiali e a volte errate a causa di un'analisi forzata dei dati che non sarebbe possibile fare normalmente.

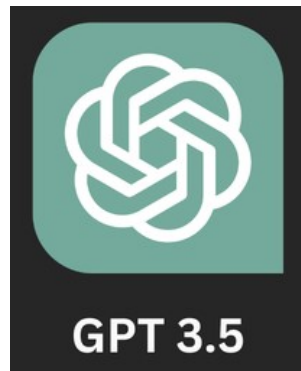


Figura 3.5: Logo ChatGPT-3.5.

3.2.6 ChatGPT 4

Con l'uscita del modello di intelligenza artificiale ChatGPT 4[19] tutti gli altri strumenti sono stati, ovviamente, oscurati. Questo è il linguistico più

avanzato, in grado di fornire risposte più sicure ed efficaci ed è un sistema multimodale in grado di accettare come input sia testi che immagini, mostrando prestazioni a livello umano su una serie di test professionali e accademici. Questa nuova capacità è resa possibile dalla presenza di una componente Vision, da qui ChatGPT-4V[20], che sblocca l'abilità di analisi visiva mai vista prima in un modello. La sua affidabilità, creatività e collaborazione sono state notevolmente perfezionate rispetto ai suoi predecessori, garantendo un'esperienza di interazione senza limiti.

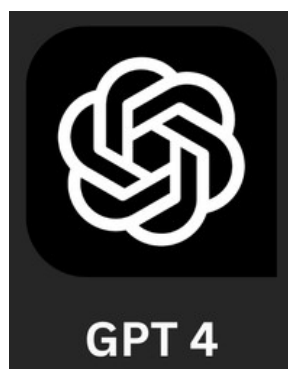


Figura 3.6: Logo ChatGPT-4.

3.3 Analisi dei dati strutturati

Come discusso nel capitolo 1.1.1, i dati *strutturati* sono informazioni con un formato standardizzato per un accesso efficiente da parte del software e degli utenti. Tra le tipologie più diffuse di questa famiglia di dati si individuano, in particolare, i file Excel, i Database SQL, i dati e i sistemi di macchinari. Trattandosi il più delle volte di tabelle con righe e colonne, i computer sono in grado di elaborare queste informazioni in maniera efficiente. Anche in questo caso, come per la controparte rappresentata dai dati non strutturati, l'utilizzo dell'intelligenza artificiale e del *prompt engineering* potrebbe essere una mossa vincente per analizzare efficientemente questi dati. Essendo ben strutturati e definiti, infatti, possono essere studiati non solo in modo rapido ma anche ripetitivo; ad esempio, se si studia

un database in cui cambia il contenuto ma non la forma si può applicare lo stesso tipo di analisi ogni volta, a patto che sia stato progettato in modo corretto. I dati strutturati, con la loro organizzazione precisa e il loro formato facilmente interrogabile, offrono un terreno fertile per l'applicazione del *prompt engineering*. Allo stesso tempo, il *prompt engineering* apre nuove possibilità per accedere e interpretare questi dati, rendendoli più accessibili e utilizzabili in una varietà di contesti. Questa interazione non solo migliora l'efficienza nell'elaborazione e nell'analisi dei dati, ma aumenta anche l'accessibilità delle informazioni, consentendo agli utenti di interagire con i dati in modo più naturale e intuitivo. Invece di navigare attraverso complessi set di dati o eseguire manualmente operazioni di analisi, gli utenti possono utilizzare prompt ben progettati per guidare il sistema nell'esecuzione di analisi sofisticate, come la generazione di sintesi, l'identificazione di tendenze o la creazione di confronti. Questo approccio non solo semplifica il processo di analisi dei dati, ma consente anche una maggiore flessibilità e personalizzazione dell'output. Tutti queste possibilità verranno illustrate nel capitolo dedicato alle prove, 4.4. Per l'analisi di *dati strutturati* sono stati utilizzati due strumenti: la libreria PandasAI e il modello di intelligenza artificiale ChatGPT 4. Nelle seguenti sezioni verranno illustrati nel dettaglio.

3.3.1 PandasAI

PandasAI[21] rappresenta un'evoluzione naturale nell'ambito dell'analisi dati, edificando sulle solide fondamenta offerte da Pandas, una delle librerie più rinomate e utilizzate in Python per la manipolazione e l'analisi di dati strutturati. Pandas, con la sua intuitiva interfaccia e la capacità di gestire facilmente dati tabulari, ha rivoluzionato il modo in cui i dati sono esplorati, puliti e trasformati, diventando uno strumento indispensabile per data scientists e analisti. PandasAI estende questa funzionalità incorporando capacità di apprendimento automatico e intelligenza artificiale direttamen-

te all'interno dell'ecosistema Pandas, mirando a semplificare e rendere più accessibili le tecniche avanzate di data analysis. Attraverso l'integrazione di modelli AI pre-addestrati e algoritmi di machine learning, PandasAI aspira a ridurre la complessità e il tempo necessario per passare dalla preparazione dei dati all'ottenimento di insight significativi, democratizzando l'accesso alle tecnologie avanzate di analisi dati.



Figura 3.7: Logo PandasAI.

Questa innovativa libreria fornisce due principali vantaggi:

- **Facilità d'uso:** è progettata per essere facile da usare, anche se non si ha familiarità con l'intelligenza artificiale generativa o con Pandas. È possibile porre semplicemente delle domande ai dati in linguaggio naturale e PandasAI genererà il codice per rispondere ad esse,
- **Potenza:** può essere utilizzato per eseguire un'ampia varietà di attività, tra cui esplorazione, analisi, visualizzazione, pulizia, imputazione e ingegneria delle funzionalità dei dati.

Sicuramente sarà una libreria che continuerà ad evolvere e ad aumentare gli utenti che la utilizzano ma per ora, essendo nata a metà dell'anno 2023, risulta essere limitata e limitante a causa dei bug che si riscontrano durante il suo utilizzo. Quest'ultimo aspetto sarà più chiaro nella sezione testing relativa al suo impiego, il capitolo 4.4.

3.3.2 ChatGPT 4

Continuando il discorso iniziato nel capitolo 3.2.6, ChatGPT-4[19], come i suoi predecessori, è progettato per capire, interpretare e generare testo

basato su un vasto training set di dati. Questo include l'elaborazione di dati sia strutturati che non strutturati, ma il suo approccio e la sua efficacia possono variare in base alla natura del compito e al tipo di dati. I dati strutturati^{1.1.1}, essendo organizzati in un formato facilmente comprensibile e analizzabile come tabelle con righe e colonne, sono più ottimali quando si parla della loro interpretazione e utilizzo rispetto a quelli non strutturati, essendo più casuali e senza un o schema fisso e richiedendo l'interpretazione del contenuto prima di poterlo analizzare in modo significativo. Quando si parla di analisi di dati strutturati, ChatGPT-4 può essere considerato lo strumento più adeguato come un sistema di database o software specifico per l'analisi dati, essendo ottimizzato per operazioni rapide e complesse su grandi volumi di dati strutturati. Inoltre, ChatGPT può essere molto utile per interpretare i risultati di queste analisi, spiegare le tendenze nei dati, generare narrazioni basate sui dati, o persino per costruire query complesse in linguaggio naturale da eseguire su database.

In conclusione, ChatGPT-4 può essere un potente strumento per lavorare con entrambi i tipi di dati ma soprattutto con quelli di tipo strutturato, offrendo capacità di elaborazione del linguaggio naturale avanzate per interpretare, analizzare e generare testo basato su vasti set di dati.

Capitolo 4

Approccio e analisi dei dati

In questo capitolo verrà trattata la fase di sviluppo del progetto, che segue la fase di selezione dei metodi e degli strumenti trattata nel capitolo precedente. Saranno descritte le scelte effettuate, partendo dai dati presi in analisi fino ad arrivare alle richieste effettuate, ponendo particolare attenzione al legame con gli obiettivi del progetto. L'intento dell'analisi è stato quello di partire dal tipo di documento più generico per poi passare a quello più particolare e definito. Per questo motivo, si inizierà dai dati non strutturati per poi terminare con i dati strutturati.

4.1 Selezione dei dati non strutturati da analizzare

Come introdotto poco fa, il primo test che ha richiamato l'attenzione è stato svolto lavorando con un insieme di dati non strutturati. Come illustrato nel capitolo introduttivo 1.1.2, un dato non strutturato non ha un modello predefinito e può essere memorizzato come oggetto di vario genere: immagini, audio, video, file di documenti e altri formati di file. Per questo motivo, e per rimanere in un contesto di Industria 4.0, è stato selezionato un file testuale in formato PDF contenente una normativa sulla Robotica Collaborativa, la ISO/TS 15066:2016[22].

4.2 Preparazione dei dati per l'analisi

Dopo aver selezionato il materiale con cui lavorare sono stati selezionati gli strumenti, introdotti nel capitolo precedente, ed infine sono state svolte le varie analisi. Per effettuare questo test nel modo più completo possibile, sono state svolte tre prove per ogni tool:

- **testo e domande in lingua inglese:** le normative tecniche sono dei documenti redatti in lingua inglese quindi questo simula un approccio in lingua originale sia per i dati che per la serie di domande;
- **testo in lingua inglese e domande in lingua italiana:** si mantiene il documento in lingua originale ma si utilizza un'altra lingua, in questo caso l'italiano, per vedere come il modello di intelligenza artificiale reagisce dovendo simultaneamente tradurre e ricercare una risposta;
- **testo e domande in lingua italiana:** tramite utilizzo di un traduttore di documenti basato su intelligenza artificiale, chiamato DeepL[23], è stato tradotto il documento in un'altra lingua, in questo caso l'italiano, per poi utilizzarlo con i vari tool insieme a delle domande presentate nella stessa lingua.

Questo tipo di approccio serve a testare i limiti del modello di analisi dato in base alle lingue utilizzate oltre che alle capacità di ragionamento vere e proprie. Essendo un documento ricco di informazioni tecniche, tabelle e immagini, è stata prodotta una lista di dieci domande accuratamente scelte per poter testare le svariate capacità dei modelli alla base del funzionamento degli strumenti selezionati per l'analisi. Queste domande sono riportate di seguito:

1. Di cosa parla il documento?
2. Cos'è un robot collaborativo?

3. Quali sono i requisiti di un sistema di robot collaborativi?
4. Fai un riassunto del topic principale e spiegalo in parole semplici.
5. Parlami della seconda immagine (“Figure 2”).
6. Quante sono e di cosa parlano le immagini nel documento? Fai un elenco ordinato.
7. Quante sono e di cosa parlano le formule nel documento? Fai un elenco ordinato.
8. Parlami del modello di corpo umano utilizzato per le prove per il calcolo dei rischi ("Figure A.1" e "Table A.1").
9. Di cosa parla la tabella “Table A.2”?
10. Sulla base delle informazioni acquisite dal documento, un uomo che ostacola il lavoro di un robot è in pericolo?

Una volta somministrata la lista in tutte e tre le tipologie di prove illustrate precedentemente, si passa alla valutazione delle capacità di analisi dati del tool in analisi per poi passare al successivo. Terminata questa fase è stato possibile individuare lo strumento migliore, ad oggi, per l’analisi dati di tipo non strutturato. Nel capitolo successivo, 5.1, verranno illustrati i risultati ottenuti con le opportune considerazioni.

4.3 Selezione dei dati strutturati da analizzare

Il secondo test che è stato sviluppato ha riguardato l’analisi di un insieme di dati strutturati. Come illustrato nel capitolo introduttivo 1.1.1, un dato strutturato ha un formato standardizzato per un accesso efficiente da parte del software e degli utenti. Grazie alla loro natura quantitativa, i computer

sono lo strumento ideale per elaborare efficacemente questi dati per ottenere un'analisi efficiente e completa.

4.4 Preparazione dei dati per l'analisi

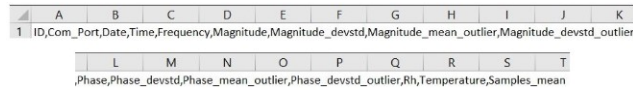
Essendo i file Excel e i Database tra le tipologie più diffuse di questa famiglia di dati, è stato scelto per la seguente prova un database contenente misurazioni eseguite su svariati provini realizzati in calcestruzzo utilizzati in un progetto in corso di sviluppo all'interno dell'Università Politecnica delle Marche e che comprende un team di lavoro capitanato dal professor Gian Marco Revel e dalla ricercatrice Cosoli Gloria con cui il professor Mancini Adriano, nonché relatore di questo elaborato, sta collaborando. Tali dati sono ottenuti nell'ambito del progetto ReCity.

4.4.1 Presentazione progetto

Il progetto in questione, nominato ReCity, riguarda lo sviluppo di una rete di sensori basata su materiali self-sensing per il monitoraggio di strutture ed infrastrutture critiche in ambito sismico. Lo scopo di questa rete di sensori è quello di monitorare strutture/infrastrutture realizzate in materiali cementizi al fine di identificare prontamente eventuali danni, già al momento della loro insorgenza. L'acquisizione continua dei dati che si ha con il monitoraggio, in combinazione con i modelli di intelligenza artificiale, permette di sfruttare l'elevata mole di informazioni per scopi predittivi. Con questi ultimi è possibile dare un aiuto concreto alla definizione di strategie di decision making che vadano a preservare la città e la relativa cittadinanza, mettendo l'uomo al centro dell'ecosistema, preservandone la sicurezza e cercando di migliorarne lo stato di benessere complessivo.

4.4.2 Database progetto

I dati contenuti nel database, utilizzato in questi test per l'analisi di dati strutturati, si riferiscono a time-series acquisite su diversi provini attraverso strumenti impiegati per misurazioni di impedenza elettrica. Queste misurazioni sono state effettuate tramite ausilio di un chip AD5940 (Analog Devices) controllato dalla scheda EVAL-AD5940BIOZ (Analog Devices). Le misurazioni effettuate per ogni provino sono periodicamente inserite in un apposito database, utilizzato poi in questo progetto di tesi con il fine di svolgere su di esso un'analisi completa dei dati e valutare come le tecniche di *prompt engineering* insieme alla potenza dei LLM sono uno strumento fondamentale per gestirle in modo efficace. Di seguito si trova un estratto della struttura del Database in questione.



	A	B	C	D	E	F	G	H	I	J	K
1	ID,Com	Port,Date,Time,Frequency,Magnitude,Magnitude_devstd,Magnitude_mean_outlier,Magnitude_devstd_outlier									
	L	M	N	O	P	Q	R	S	T		
	Phase,Phase_devstd,Phase_mean_outlier,Phase_devstd_outlier,Rh,Temperature,Samples_mean										

Figura 4.1: Struttura Database ReCity.

I dati presenti sono organizzati considerando il tempo e l'ID del provino, memorizzando modulo (colonna magnitude) e fase dell'impedenza elettrica, insieme ad altre grandezze associate. Ai fini del progetto di tesi, per svolgere analisi dei dati e previsioni, è rilevante solo il valore del modulo di impedenza (colonna magnitude), in quanto il valore della fase risulta superflua per lo studio in questione.

4.4.3 Preparazione del database per l'analisi con PandasAI

Come detto all'inizio di questo paragrafo, il dato strutturato selezionato per l'analisi è un database contenente svariate misurazioni. Il primo step per impostare l'analisi riguarda la conversione del file nel formato desiderato. Infatti, trattandosi di un database si presenta con un'estensione *.sql*,

cioè un file SQL (Structured Query Language), un linguaggio di programmazione utilizzato per l'archiviazione e l'elaborazione di informazioni in un database relazionale. Un database relazionale memorizza le informazioni in forma tabulare, con righe e colonne che rappresentano diversi attributi di dati e le varie relazioni tra i valori dei dati. Attraverso l'utilizzo della libreria SQLite e la programmazione in linea di comando, questo file è stato convertito prima in un formato *.db* e poi esportato in un file *.csv* (Comma-Separated Values), un formato che rappresenta uno dei modi più semplici per rappresentare dati in forma tabellare all'interno di un semplice file di testo. Non a caso, questo formato di file è uno dei primi ad essersi diffuso per l'interscambio di dati, e nonostante la presenza di altri formati altrettanto conosciuti ed utilizzati, è ancora oggi diffuso in molte applicazioni. Il file *.csv* ottenuto, contenente il database, si presenta nel seguente modo:

	A	B	C	D	E	F	G	H	I	J	K
1	ID,Com_Port,Date,Time,Frequency,Magnitude,Magnitude_devstd,Magnitude_mean_outlier,Magnitude_devstd_outlier,Phase										
2	T3_TU04L,COM10,02-02-2023,10:44:37,1000,597.26,0.54,597.62,1.5,-69.16,0.08,-69.14,0.12,57.4,13.7,72										
3	T3_TU04L,COM10,02-02-2023,10:44:44,2000,664.24,0.35,664.47,0.82,-104.38,0.09,-104.36,0.13,57.4,13.7,71										
4	T3_TU04L,COM10,02-02-2023,10:44:51,3000,721.19,0.22,721.32,0.48,-123.24,0.08,-123.22,0.11,57.4,13.7,71										
5	T3_TU04L,COM10,02-02-2023,10:44:58,4000,760.07,0.14,760.14,0.26,-134.87,0.07,-134.85,0.11,57.4,13.7,72										
6	T3_TU04L,COM10,02-02-2023,10:45:06,5000,786.05,0.1,786.07,0.13,-142.81,0.07,-142.79,0.1,57.4,13.7,73										
7	T3_TU04L,COM10,02-02-2023,10:45:13,6000,803.67,0.1,803.68,0.11,-148.57,0.06,-148.56,0.09,57.4,13.7,73										
...											
600571	T1_TU03S,COM7,21-09-2023,16:14:06,16000,319.32,0.13,319.42,0.23,-25.12,0.02,-25.09,0.06,64.6,26,64										
600572	T1_TU03S,COM7,21-09-2023,16:14:13,17000,314.57,0.314,61,0.17,-25.65,0,-25.64,0.01,64.6,26,69										
600573	T1_TU03S,COM7,21-09-2023,16:14:20,18000,310.27,0.16,310.26,0.2,-26.07,0.03,-26.07,0.03,64.6,26,77										
600574	T1_TU03S,COM7,21-09-2023,16:14:27,19000,305.21,0.09,305.22,0.1,-26.57,0.02,-26.57,0.02,64.6,26,77										
600575	T1_TU03S,COM7,21-09-2023,16:14:35,20000,300.62,0.5,300.6,0.53,-26.88,0.09,-26.87,0.1,64.6,26,74										
600576											

Figura 4.2: Database utilizzato nella prova di analisi dati strutturati.

A questo punto inizia l'analisi vera e propria del documento strutturato tramite l'utilizzo della libreria PandasAI. Questa permette di avere due tipologie di interazione con il modello di intelligenza artificiale:

- **One-Shot**: domanda e risposta secche, senza tenere traccia di quelle precedenti. In questa modalità si utilizza uno SmartDataFrame.
- **Conversazione**: viene mantenuta una cronologia per simulare una conversazione e senza avere il bisogno di fare un quadro completo

della situazione. In questa versione, a differenza del tipo precedente, c'è bisogno di utilizzare un agente.

A seconda del tipo domande che devono essere fatte per svolgere l'analisi si seleziona una tipologia piuttosto che l'altra. Nel caso del progetto sotto analisi, sono state provate entrambe le tipologie per poi mantenere la versione con l'agente, in modo da far tenere memoria della conversazione al modello. Dopo aver impostato tutto, sono state fatte le seguenti domande:

1. Analizza la magnitudine misurata per ogni tipologia di provino. Ogni tipo è riconosciuto dal relativo ID; diverso ID significa diverso tipo di provino. Crea un grafico per ogni provino dove mostri l'andamento temporale del valore di magnitudine.
2. Fai un forecast dei campioni. Quanto varrà la magnitudine per i prossimi 50 istanti temporali di ogni provino?
3. Fai la stessa previsione usando la libreria Prophet.
4. Fai la stessa previsione usando la libreria NeuralProphet.
5. Trovare il valor medio, max, mix e altri valori statistici relativi alla magnitudine.
6. Quante volte i provini sono entrati nel fenomeno dell'isteresi?
7. Genera grafici con soglie per isteresi.
8. Considera gli ultimi 500 campioni per i provini e studia l'andamento statistico della magnitudine.
9. Genera il codice per automatizzare questa analisi appena svolta sui 500 campioni.
10. Genera il codice per automatizzare il forercasting fatto sui provini.

Una volta somministrata la lista si passa alla valutazione delle capacità di analisi dati della libreria PandasAI in analisi per poi passare allo strumento successivo.

4.4.4 Preparazione del database per l'analisi con ChatGPT4

Dopo aver effettuato il test con uso di PandasAI e, anticipando ciò che verrà illustrato nel prossimo capitolo riguardo i risultati, 5.2, dopo aver scoperto la grande limitazione attuale della libreria è stato ripreso l'utilizzo di ChatGPT 4, data la potenza attualmente raggiunta, che ha generato ulteriori possibilità di utilizzo. Per questi motivi, lo stesso test è stato svolto cambiando il mezzo di analisi per poter svolgere in maniera più precisa e completa lo studio del database. Dopo aver chiesto al modello se fosse in grado di analizzare un file SQLite, e avendo ricevuto una risposta positiva, è stato caricato il file contenente le misurazioni. A questo punto, la nuova prova è pronta per iniziare. In questo caso, prima di fare le domande sui dati, sono state prese ulteriori precauzioni per ridurre ulteriormente le possibilità di errori. Dopo aver chiesto la conversione del file *.sql* in file *.csv* sono state effettuate le seguenti mosse:

- Si richiede di selezionare dal database ottenuto le ultime 10mila righe, in modo da avere solo le misurazioni più recenti per poter studiare le tendenze,
- Caricamento del nuovo file nella chatbox per farlo analizzare ed individuare eventuali anomalie di filtraggio,
- A questo punto si pone la seguente domanda: "Seleziona la magnitudine dei provini dove sia stata utilizzata una frequenza di 10000 Hz. Dopo averli selezionati, raggruppalì per provini e inseriscili in dataframe diversi",
- Da questa fase sono derivati cinque dataframe, che sono stati salvati,

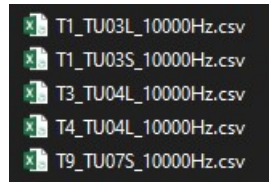


Figura 4.3: Dataframe ottenuti dopo l'analisi del database.

- Upload del primo e, dopo averlo fatto analizzare, somministrazione delle domande.

Le domande utilizzate sono le stesse utilizzate nel caso di studio precedente ma, per eliminare ogni possibilità di incappare nelle stesse limitazioni, sono state riformulate e gestite in modo migliore. Di seguito sono riportate nella nuova versione:

1. Crea un grafico dove mostri l'andamento temporale del valore di magnitude.
2. Fai un forecast del campione. Quanto varrà la magnitude per i prossimi 50 campioni temporali?
3. Fai la stessa previsione usando la libreria Prophet.
4. Fai la stessa previsione usando la libreria Neural Prophet.
5. Genera un codice usando python per automatizzare queste previsioni, per i due approcci. Per ciascuno, genera un codice separato.
6. Trova il valor medio, max, mix e altri valori statistici dei valori raccolto nel dataframe.
7. Considera il valore medio della magnitude. Quando il valore di una misurazione supera il 2% del valore medio considera l'ingresso in isteresi. Quante volte il provino è entrato in isteresi? Quando? Riportami le informazioni relative a ora, data e valore di magnitude delle misurazioni.

8. Genera un grafico con le soglie per isteresi. Tieni pulito l'asse X inserendo solo le misurazioni rilevanti.
9. Considera i campioni temporali per il provino e studia l'andamento statistico.
10. Genera un codice usando python per automatizzare questa analisi statistica.

Terminata questa fase è stato possibile confrontare i due test con i dati strutturati per poi individuare lo strumento migliore, ad oggi, per l'analisi di questo tipo di dati. Nel capitolo successivo, 5.2, verranno illustrati i risultati ottenuti con le opportune considerazioni.

Capitolo 5

Risultati

In questo capitolo, saranno presentati i risultati derivanti da entrambe le tipologie di dati analizzati, suddivisi per ogni strumento utilizzato, illustrati nel capitolo 3. Nella prima parte, saranno esaminati i risultati ottenuti dall'analisi del documento non strutturato e dei diversi tool impiegati, illustrando i vari limiti ed individuando il migliore attualmente disponibile; nella seconda parte, invece, verranno illustrati i risultati derivati dall'analisi del documento strutturato e dei due strumenti adoperati per poi valutare il loro lavoro in termini di affidabilità e efficienza.

5.1 Caso di studio 1: Analisi di dati non strutturati

Come anticipato nell'introduzione di questo capitolo, in questa sezione verranno illustrati e valutati i risultati ottenuti dall'analisi dei dati non strutturati, mettendo a paragone le capacità dei tool utilizzati per poter individuare quale sia il migliore in termini di affidabilità ed efficacia. Per evitare risultati ridondanti, avendo ottenuto risultati simili per la maggior parte dei tool, verrà riportata una selezione. In questo modo, è possibile osservare le differenze di capacità di strumenti che hanno alla base lo stesso modello di intelligenza artificiale. Di seguito, dunque, verranno riportate

tre tipologie di risultati ottenuti: uno insufficiente, uno accettabile e uno quasi perfetto.

5.1.1 Risultati test AskYourPDF

Questo strumento ha ottenuto il punteggio più basso, in termini di qualità e capacità di risposta, comportando la peggior affidabilità riscontrata tra i tool testati. In realtà, tutti i tool usati, ad eccezione dei due che verranno trattati successivamente, hanno riscontrato le stesse limitazioni ma AskYourPDF si aggiudica l'ultima posizione, seppur di poco. Un aspetto negativo importante è la difficoltà riscontrata inizialmente durante il caricamento del documento; infatti, solo dopo svariati tentativi è riuscito ad acquisirlo ed elaborarlo. Inoltre, il fattore comune con questi strumenti è la restituzione di risposte brevi e non molto dettagliate. Di seguito sono riportati i risultati ottenuti, considerando come prima prova quella con documento e domande in lingua inglese, come seconda prova quella con documento in lingua inglese e domande in lingua italiana e come terza ed ultima prova quella con documento e domande in lingua italiana.

5.1.1.1 Prima prova

Durante questa prima prova è stato possibile ottenere delle risposte coerenti e corrette solo nelle prime domande, essendo basilari e senza richiesta di ragionamento. Infatti, le risposte sono corrette e riporta anche le fonti, citando la posizione nel documento in cui ha selezionato le informazioni, ma unisce contenuto derivante dalla sua conoscenza anche quando vietato esplicitamente.

1. What is the document about?

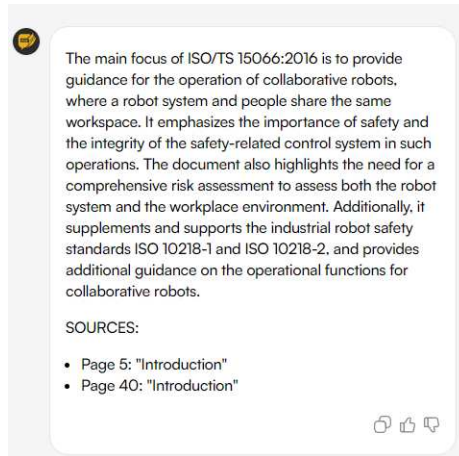


Figura 5.1: Prima domanda AskYourPDF, inglese-inglese.

2. What is a collaborative robot?

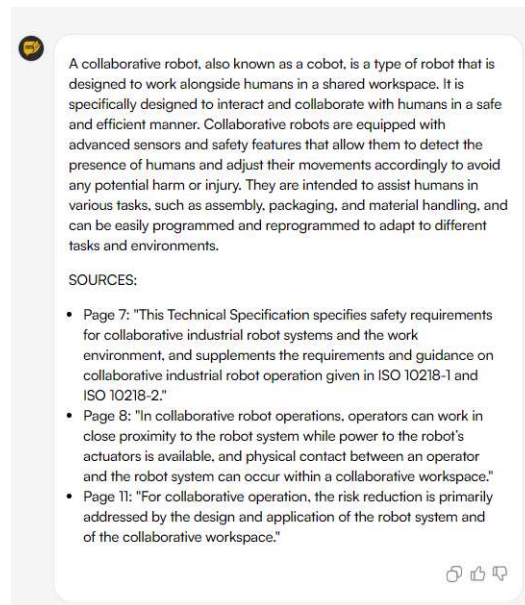


Figura 5.2: Seconda domanda AskYourPDF, inglese-inglese.

3. What are the requirements of a collaborative robot system?

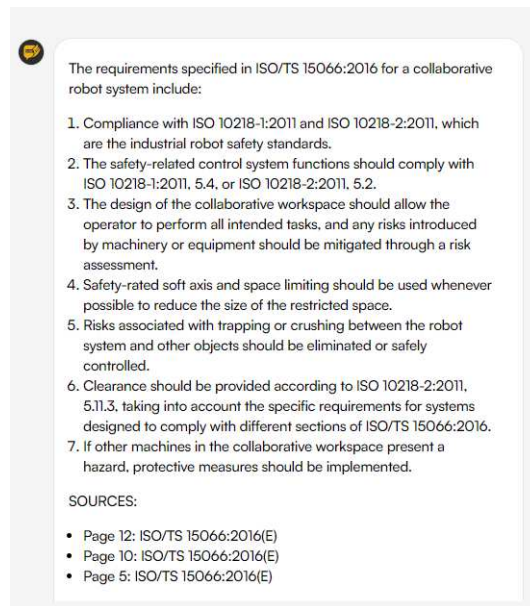


Figura 5.3: Terza domanda AskYourPDF, inglese-inglese.

4. Make a summary of the main topic and explain it in simple words.

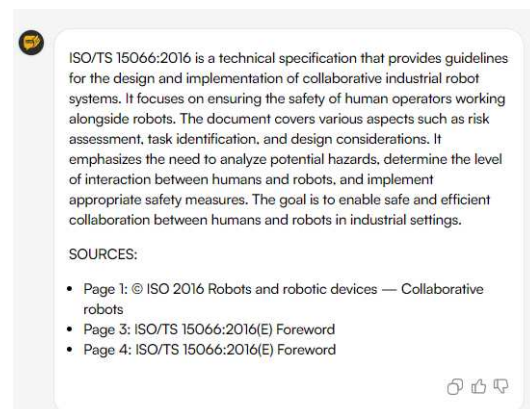


Figura 5.4: Quarta domanda AskYourPDF, inglese-inglese.

Da questo punto in poi sono state riscontrate grandi difficoltà per continuare lo svolgimento della prova, in quanto le domande avrebbero richiesto una capacità di analisi e ragionamento insieme ad una capacità di analisi visiva, aspetto mancante in tutti gli strumenti utilizzati fatta eccezione per l'ultimo che verrà illustrato. Per questo motivo, non è stato possibile ottenere risposte relative ad immagini, tabelle, schemi o formule presenti nel docu-

mento, fino ad arrivare al blocco del tool stesso che ha costretto l'attesa di qualche ora prima di poterlo impiegare per le due prove successive.

5. Tell me about the second image ("Figure 2").

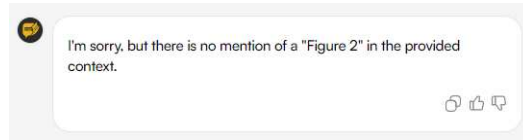


Figura 5.5: Quinta domanda AskYourPDF, primo tipo, inglese-inglese.

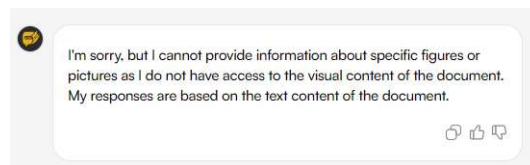


Figura 5.6: Quinta domanda AskYourPDF, secondo tipo, inglese-inglese.

6. How many are there and what are the images in the document about?
Give me an ordered list.

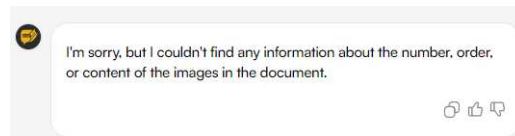


Figura 5.7: Sesta domanda AskYourPDF, inglese-inglese.

7. How many are there and what are the formulas in the document about? Give me an ordered list.

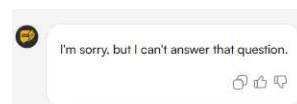


Figura 5.8: Settima domanda AskYourPDF, inglese-inglese.

8. Tell me about the human body model used for testing to calculate risks (Figures A.1 and Table A.1).

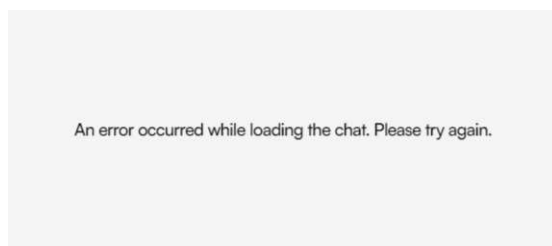


Figura 5.9: Ottava domanda AskYourPDF, inglese-inglese.

5.1.1.2 Seconda prova

In questa prova è emersa una difficoltà di rispondere alle domande in contemporanea ad una traduzione. Infatti, anche se come nel caso precedente le prime risposte sono state corrette e lo strumento non è andato in crash queste risultano essere più vaghe.

1. Di cosa parla il documento?

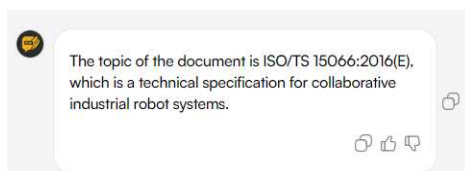


Figura 5.10: Prima domanda AskYourPDF, inglese-italiano.

2. Cos'è un robot collaborativo?

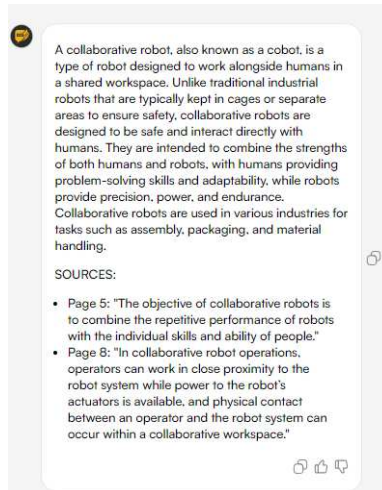


Figura 5.11: Seconda domanda AskYourPDF, inglese-italiano.

3. Quali sono i requisiti di un sistema di robot collaborativi?

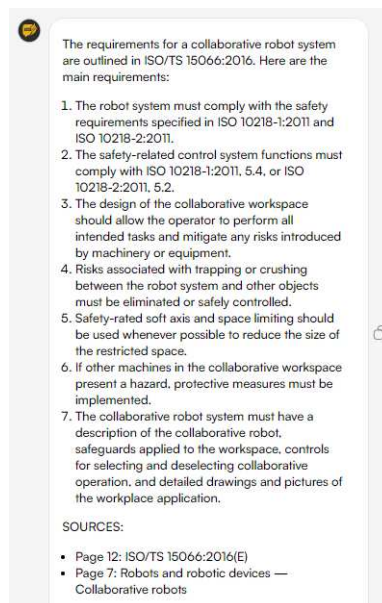


Figura 5.12: Terza domanda AskYourPDF, inglese-italiano.

4. Fai un riassunto del topic principale e spiegalo in parole semplici.

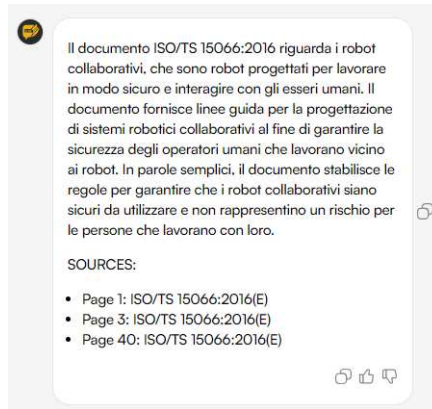


Figura 5.13: Quarta domanda AskYourPDF, inglese-italiano.

Da questo punto in poi sono state riscontrate le difficoltà incontrate anche precedentemente riguardo la capacità di analisi, di ragionamento e di analisi visiva. Questa volta però non è stato riscontrato alcun crash ed è stato quindi possibile porre tutta la lista di domande.

5. Parlatemi della seconda immagine (“Figure 2”).

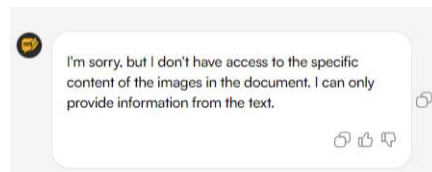


Figura 5.14: Quinta domanda AskYourPDF, inglese-italiano.

6. Quante sono e di cosa parlano le immagini nel documento? Fammi un elenco ordinato.

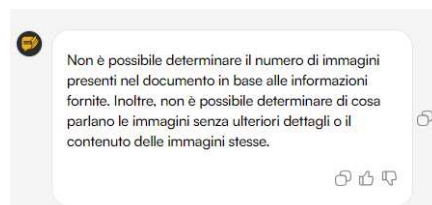


Figura 5.15: Sesta domanda AskYourPDF, inglese-italiano.

7. Quante sono e di cosa parlano le formule nel documento? Fammi un elenco ordinato.

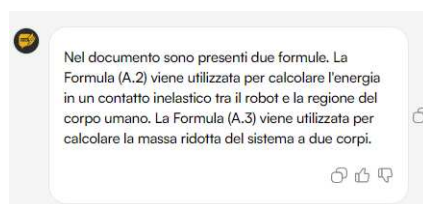


Figura 5.16: Settima domanda AskYourPDF, inglese-italiano.

8. Parlami del modello di corpo umano utilizzato per le prove per il calcolo dei rischi (Figure A.1 e Table A.1).

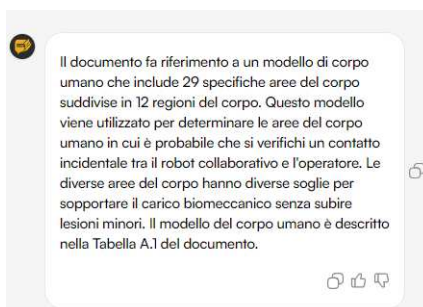


Figura 5.17: Ottava domanda AskYourPDF, inglese-italiano.

9. Di cosa parla la tabella "Table A.2"?



Figura 5.18: Nona domanda AskYourPDF, inglese-italiano.

10. Sulla base delle informazioni acquisite dal documento, un uomo che ostacola il lavoro di un robot è in pericolo?

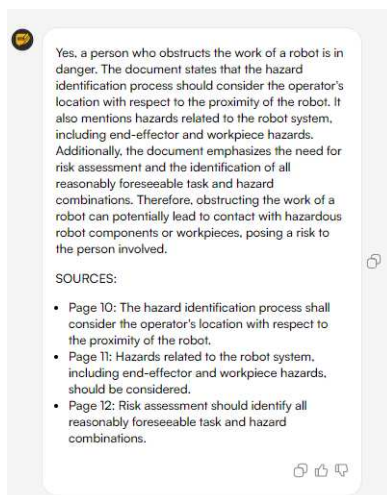


Figura 5.19: Decima domanda AskYourPDF, inglese-italiano.

5.1.1.3 Terza prova

Questa ultima parte è stata caratterizzata da un completo fallimento. Durante la traduzione del documento, effettuata con ausilio di un modello di intelligenza virtuale, il layout del documento è cambiato leggermente. Questa variazione ha causato un'impossibilità di analisi completa da parte dello strumento, che non riusciva ad estrapolare correttamente le informazioni causando risposte per lo più errate. Su dieci domande, infatti, solo tre hanno trovato correttamente una risposta; per le altre sono state ottenute risposte errate o mancanti. Per non inserire informazioni superflue, non vengono aggiunte le foto relative a questa prova.

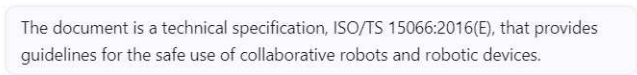
5.1.2 Risultati test ChatPDF

Il risultato intermedio è stato ottenuto con l'utilizzo del tool ChatPDF, che è stato in grado di rispondere a tutte le domande tranne quelle in cui era richiesta una analisi visiva, seppur con risposte brevi e senza citazione delle fonti. Tuttavia, ai fini del test la parte rilevante era rappresentata dall'ottenimento di risposte corrette e pertinenti e non dalla loro lunghezza. Di seguito sono riportati i risultati delle tre prove.

5.1.2.1 Prima prova

Durante questa prima prova è stato possibile ottenere delle risposte corrette in quasi la totalità delle domande. Infatti, solo nel caso delle domande N.6 e N.7, relativa alle immagini e alle formule presenti nel documento, non è stata trovata una risposta; altrimenti, anche per le domande relative l'analisi di tabelle e figure, è stata ottenuta una risposta corretta grazie all'analisi contestuale che è riuscita a colmare la lacuna dell'impossibilità di effettuare analisi visiva.

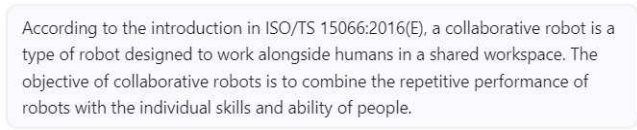
1. What is the document about?



The document is a technical specification, ISO/TS 15066:2016(E), that provides guidelines for the safe use of collaborative robots and robotic devices.

Figura 5.20: Prima domanda ChatPDF, inglese-inglese.

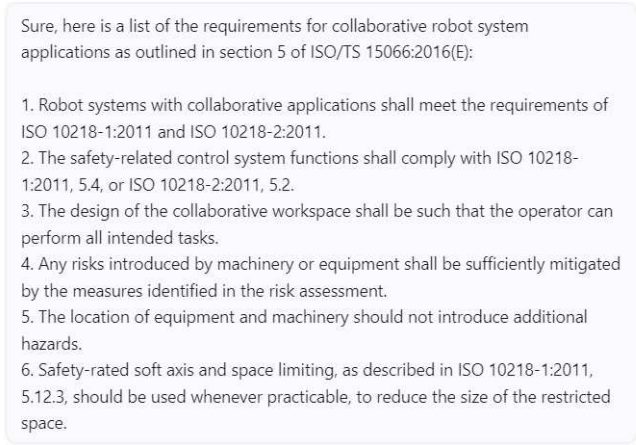
2. What is a collaborative robot?



According to the introduction in ISO/TS 15066:2016(E), a collaborative robot is a type of robot designed to work alongside humans in a shared workspace. The objective of collaborative robots is to combine the repetitive performance of robots with the individual skills and ability of people.

Figura 5.21: Seconda domanda ChatPDF, inglese-inglese.

3. What are the requirements of a collaborative robot system?



Sure, here is a list of the requirements for collaborative robot system applications as outlined in section 5 of ISO/TS 15066:2016(E):

1. Robot systems with collaborative applications shall meet the requirements of ISO 10218-1:2011 and ISO 10218-2:2011.
2. The safety-related control system functions shall comply with ISO 10218-1:2011, 5.4, or ISO 10218-2:2011, 5.2.
3. The design of the collaborative workspace shall be such that the operator can perform all intended tasks.
4. Any risks introduced by machinery or equipment shall be sufficiently mitigated by the measures identified in the risk assessment.
5. The location of equipment and machinery should not introduce additional hazards.
6. Safety-rated soft axis and space limiting, as described in ISO 10218-1:2011, 5.12.3, should be used whenever practicable, to reduce the size of the restricted space.

Figura 5.22: Terza domanda ChatPDF, inglese-inglese.

4. Make a summary of the main topic and explain it in simple words.

The main topic of the PDF is ISO/TS 15066:2016(E), which outlines the requirements for collaborative robot system applications. Collaborative robots are robots designed to work alongside humans in a shared workspace. The objective of collaborative robots is to combine the repetitive performance of robots with the individual skills and ability of people. The PDF provides guidelines for ensuring the safety of collaborative robot systems, including requirements for safety-related control system performance, design of the collaborative workspace, and the use of safety-rated soft axis and space limiting. In simple words, the PDF provides safety guidelines for robots that work alongside humans to ensure that they are safe to use and do not pose a risk to human operators.

Figura 5.23: Quarta domanda ChatPDF, inglese-inglese.

5. Tell me about the second image ("Figure 2").

To answer your question, Figure 2 on page 14 of the PDF shows the truth table for safety-rated monitored stop operations. The table outlines the operations of the safety-rated monitored stop feature, which is a safety feature that allows a robot system to enter a collaborative workspace only when an operator is not present in the workspace. The table shows the different states of the safety-rated monitored stop feature and the corresponding actions that the robot system can take. The figure is intended to help users understand the safety requirements for collaborative robot systems and how the safety-rated monitored stop feature works.

Figura 5.24: Quinta domanda ChatPDF, inglese-inglese.

6. How many are there and what are the images in the document about?
Give me an ordered list.

1. How many images are there in the document and what are they about?
▶ What is the purpose of ISO/TS 15066:2016(E)?
▶ What is the description of the hand guiding method of operation?

To answer your question, there are no images in the PDF file. The document is a technical specification that provides guidelines for the safe use of collaborative robots. It covers topics such as risk assessment, safety requirements, and protective measures for human-robot collaboration. The hand guiding method of operation is described in section 5.5.3 of the document, which explains how an operator can use a hand-operated device to transmit motion commands to the robot system.

Figura 5.25: Sesta domanda ChatPDF, inglese-inglese.

7. How many are there and what are the formulas in the document about? Give me an ordered list.

To answer your first question, there are no formulas in the PDF file. The document provides guidelines and recommendations for the safe use of collaborative robots, but it does not contain any mathematical formulas.

Figura 5.26: Settima domanda ChatPDF, inglese-inglese.

8. Tell me about the human body model used for testing to calculate risks (Figures A.1 and Table A.1).

To answer your question, the human body model used in the tests for risk calculation is described in Annex A of the document. The body model includes 29 specific body areas categorized in 12 body regions, and it is shown in Figure A.1. Table A.1 shows the specific body regions, classified into general body regions, and designated as being located in either the front or the back of the body. The body model is used to determine where on the operator's body contact between the robot and operator is likely to occur, which is critical for determining the risk of injury. The body model assumes that the contact between a robot and a human body part is a fully inelastic two-body collision, and it provides guidance on the force and pressure values that should be used in risk assessments. However, it is important to note that the actual contact conditions, including the geometric shape of the contact area, would need to be compared against the force and pressure values in Table A.2 through measurement or calculations.

Figura 5.27: Ottava domanda ChatPDF, inglese-inglese.

9. What is the table "Table A.2" about?

1. How many formulas are there in the document and what are they about?
▶ What is the purpose of ISO/TS 15066:2016(E)?
▶ What is the description of the hand guiding method of operation?

To answer your question, Table A.2 in Annex A of the document provides data on the maximum permissible biomechanical limits associated with contact between a robot and a human body part. The table lists the specific body areas and the corresponding force and pressure limits that should not be exceeded during contact. The limits are based on the assumption that the contact between a robot and a human body part is a fully inelastic two-body collision. The table also provides data for contact with face, skull, and forehead, but it notes that contact with these areas is not permissible. It is important to note that the actual contact conditions, including the geometric shape of the contact area, would need to be compared against the force and pressure values in Table A.2 through measurement or calculations.

Figura 5.28: Nona domanda ChatPDF, inglese-inglese.

10. Based on the information gained from the paper, is a human obstructing the work of a robot in danger?

To answer your question, the document provides guidance on risk assessment and risk reduction measures for human-robot collaboration. If a human obstructs the work of a robot, there is a potential risk of harm to the human. The document recommends that risk reduction measures be applied to keep contact situations between the robot and operator below threshold limit values. These measures may include designing the robot system and collaborative workspace so that such contact is infrequent and avoidable, and applying protective measures such as barriers or safety sensors to prevent contact between the robot and operator. It is important to note that the specific risk level would depend on the specific circumstances of the situation, and a risk assessment should be conducted to determine the appropriate risk reduction measures.

Figura 5.29: Decima domanda ChatPDF, inglese-inglese.

5.1.2.2 Seconda prova

In questo caso, l'uso di domande in lingua italiana sembra mettere in difficoltà lo strumento. A differenza del caso precedente, infatti, sono state ottenute risposte errate nelle domande 4,5,7 e 9, non solo, come nei casi precedenti, in quelle relative l'analisi visiva. Un aspetto positivo, però, è stata la risposta alla domanda N.6, non ottenuta nel test effettuato in lingua originale. Per non rallentare troppo la lettura dei risultati, verrà riportata solo l'ultima domanda citata.

1. Di cosa parla il documento?

1. Di cosa parla il documento?
Il documento è una specifica tecnica sui robot collaborativi e fornisce informazioni sulla loro sicurezza d'uso. 6

Figura 5.30: Prima domanda ChatPDF, inglese-italiano.

2. Cos'è un robot collaborativo?

Un robot collaborativo, anche noto come cobot, è un tipo di robot progettato per lavorare accanto agli esseri umani in uno spazio di lavoro condiviso. A differenza dei robot tradizionali, che sono tipicamente isolati per impedire l'accesso umano durante il funzionamento, i robot collaborativi sono progettati per essere sicuri per l'interazione umana e possono eseguire compiti in stretta prossimità degli esseri umani. 7

Figura 5.31: Seconda domanda ChatPDF, inglese-italiano.

3. Quali sono i requisiti di un sistema di robot collaborativi?

La sezione 5 del documento fornisce informazioni dettagliate sui requisiti per le applicazioni di sistemi di robot collaborativi, tra cui la performance del sistema di controllo correlato alla sicurezza, la progettazione dello spazio di lavoro collaborativo, le misure di protezione, le funzioni di arresto, i requisiti del dispositivo di abilitazione e le operazioni collaborative. 3

Figura 5.32: Terza domanda ChatPDF, inglese-italiano.

4. Fai un riassunto del topic principale e spiegalo in parole semplici.

Un robot collaborativo, o cobot, è un tipo di robot progettato per lavorare in modo sicuro accanto agli esseri umani in uno spazio di lavoro condiviso. 40

➤ Quali documenti dovrebbero essere mantenuti per un sistema di robot collaborativi?
I documenti che dovrebbero essere mantenuti includono dati di specifica per l'utilizzo del robot collaborativo nell'applicazione, dati di descrizione e specifica per le protezioni applicate allo spazio di lavoro collaborativo e all'intero sistema di lavoro e robot collaborativo, e una descrizione dei controlli per la selezione e la deselegazione dei tipi di operazioni collaborative pertinenti. 25

➤ Quali sono i requisiti per un sistema di robot collaborativi?
Il documento fornisce informazioni dettagliate sui requisiti per le applicazioni di sistemi di robot collaborativi, tra cui la performance del sistema di controllo correlato alla sicurezza, la progettazione dello spazio di lavoro collaborativo, le misure di protezione, le funzioni di arresto, i requisiti del dispositivo di abilitazione e le operazioni collaborative. 40

Figura 5.33: Quarta domanda ChatPDF, inglese-italiano.

5. Parlami della seconda immagine ("Figure 2").

Mi dispiace, ma non posso fornire informazioni sulla "Figure 2" poiché non è stata menzionata nel testo. Potresti fornire maggiori dettagli o fare una domanda specifica sulla figura?

Figura 5.34: Quinta domanda ChatPDF, inglese-italiano.

6. Quante sono e di cosa parlano le immagini nel documento? Fammi un elenco ordinato.

Certo! Ecco un elenco ordinato delle immagini presenti nel documento:

1. Figura 1 - Spazio di lavoro collaborativo ⁶
2. Figura 2 - Sistema robotico collaborativo ⁶
3. Figura 3 - Sistema robotico collaborativo con arresto monitorato a sicurezza garantita ⁶
4. Figura 4 - Sistema robotico collaborativo con operazione di guida manuale ⁶
5. Figura 5 - Sistema robotico collaborativo con monitoraggio della velocità e della separazione ⁶
6. Figura 6 - Sistema robotico collaborativo con limitazione di potenza e forza ⁶
7. Figura 7 - Sistema robotico collaborativo con limitazione morbida degli assi e dello spazio a sicurezza garantita ⁶
8. Figura 8 - Sistema robotico collaborativo con arresto monitorato a sicurezza garantita e dispositivo di rilevamento della presenza ⁶
9. Figura 9 - Sistema robotico collaborativo con arresto monitorato a sicurezza garantita e scanner di area ⁶
10. Figura 10 - Sistema robotico collaborativo con arresto monitorato a sicurezza garantita e rilevamento di forza/torqu ⁶
11. Figura 11 - Sistema robotico collaborativo con arresto monitorato a sicurezza garantita e sistema di visione ⁶
12. Figura 12 - Sistema robotico collaborativo con arresto monitorato a sicurezza garantita e riconoscimento vocale ⁶
13. Figura 13 - Sistema robotico collaborativo con arresto monitorato a sicurezza garantita e identificazione biometrica ⁶
14. Figura 14 - Sistema robotico collaborativo con arresto monitorato a sicurezza garantita e identificazione RFID ⁶

Figura 5.35: Sesta domanda ChatPDF, inglese-italiano.

7. Quante sono e di cosa parlano le formule nel documento? Fammi un elenco ordinato.

Mi dispiace, ma non ci sono formule nel documento. Potresti fare una domanda su un altro argomento?

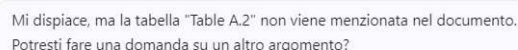
Figura 5.36: Settima domanda ChatPDF, inglese-italiano.

8. Parliami del modello di corpo umano utilizzato per le prove per il calcolo dei rischi (Figure A.1 e Table A.1).

Certamente! Il modello di corpo umano utilizzato nella valutazione del rischio per le applicazioni di robot collaborativi a limitazione di potenza e forza include 29 specifiche aree del corpo suddivise in 12 regioni corporee. Le aree di contatto nel modello del corpo sono mostrate nella Figura A.1, mentre la Tabella A.1 mostra le specifiche regioni corporee, classificate in regioni corporee generali e designate come situate nella parte anteriore o posteriore del corpo. Questo modello del corpo viene utilizzato per determinare dove sul corpo dell'operatore è probabile che si verifichi il contatto tra il robot e l'operatore, il che è critico poiché diverse aree del corpo avranno diverse soglie per sopportare il carico biomeccanico senza subire lesioni minori. ²⁷

Figura 5.37: Ottava domanda ChatPDF, inglese-italiano.

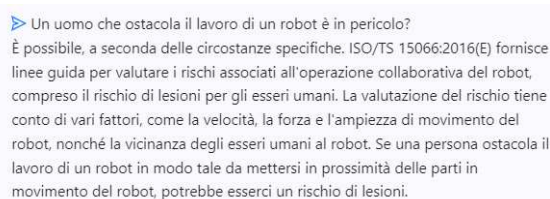
9. Di cosa parla la tabella “Table A.2”?



Mi dispiace, ma la tabella “Table A.2” non viene menzionata nel documento.
Potresti fare una domanda su un altro argomento?

Figura 5.38: Nona domanda ChatPDF, inglese-italiano.

10. Sulla base delle informazioni acquisite dal documento, un uomo che ostacola il lavoro di un robot è in pericolo?



► Un uomo che ostacola il lavoro di un robot è in pericolo?
È possibile, a seconda delle circostanze specifiche. ISO/TS 15066:2016(E) fornisce linee guida per valutare i rischi associati all'operazione collaborativa del robot, compreso il rischio di lesioni per gli esseri umani. La valutazione del rischio tiene conto di vari fattori, come la velocità, la forza e l'ampiezza di movimento del robot, nonché la vicinanza degli esseri umani al robot. Se una persona ostacola il lavoro di un robot in modo tale da mettersi in prossimità delle parti in movimento del robot, potrebbe esserci un rischio di lesioni.

Figura 5.39: Decima domanda ChatPDF, inglese-italiano.

5.1.2.3 Terza prova

L'ultima prova ha ottenuto lo stesso risultato della prima. Non ci sono stati problemi di lingua ma, in alcune domande, è stato necessario specificare di rispondere esclusivamente con le informazioni derivanti dal documento. In questo caso, risponde anche alla domanda N.6 fornendo un elenco di immagini presenti nel documento ma la risposta generata è errata. Per non essere ripetitivi, avendo ricevuto le stesse risposte in lingua diversa, verrà riportato di seguito solo la domanda N.6.

6. Quante sono e di cosa parlano le immagini nel documento? Fammi un elenco ordinato.

Certamente, ecco un elenco ordinato delle immagini presenti nel documento:

1. Figura 1: Rappresentazione schematica di un robot collaborativo
2. Figura 2: Esempio di un sistema robotico collaborativo
3. Figura 3: Esempio di un sistema robotico collaborativo con un'area di lavoro collaborativa
4. Figura 4: Esempio di un sistema robotico collaborativo con un'area di lavoro collaborativa e una zona di pericolo
5. Figura 5: Esempio di un sistema robotico collaborativo con un'area di lavoro collaborativa e una zona di pericolo, con un operatore all'interno dell'area di lavoro collaborativa
6. Figura 6: Esempio di un sistema robotico collaborativo con un'area di lavoro collaborativa e una zona di pericolo, con un operatore all'interno della zona di pericolo
7. Figura 7: Esempio di un sistema robotico collaborativo con un'area di lavoro collaborativa e una zona di pericolo, con un operatore all'esterno dell'area di lavoro collaborativa
8. Figura 8: Esempio di un sistema robotico collaborativo con un'area di lavoro collaborativa e una zona di pericolo, con un operatore all'esterno della zona di pericolo
9. Figura 9: Esempio di un sistema robotico collaborativo con un'area di lavoro collaborativa e una zona di pericolo, con un operatore all'interno dell'area di lavoro collaborativa e un altro operatore all'esterno dell'area di lavoro collaborativa
10. Figura 10: Esempio di un sistema robotico collaborativo con un'area di lavoro collaborativa e una zona di pericolo, con un operatore all'interno dell'area di lavoro collaborativa e un altro operatore all'interno della zona di pericolo
11. Figura 11: Esempio di un sistema robotico collaborativo con un'area di lavoro collaborativa e una zona di pericolo, con due operatori all'interno dell'area di lavoro collaborativa
12. Figura 12: Esempio di un sistema robotico collaborativo con un'area di lavoro collaborativa

Figura 5.40: Sesta domanda ChatPDF, italiano-italiano.

5.1.3 Risultati test ChatGPT-4, dati non strutturati

Il miglior risultato è stato ottenuto, infine, utilizzando direttamente ChatGPT-4 sul sito ufficiale di OpenAI. Grazie alla sua potenza molto elevata e alla componente vision integrata è stato in grado di rispondere a tutte le domande correttamente e con chiarezza, fornendo spiegazioni complete e precise. È stato l'unico caso in cui alla quinta, sesta e settima domanda è stata ottenuta la risposta corretta senza dover trovare escamotage. Grazie alla componente ChatGPT 4-Vision, infatti, ha svolto subito un'analisi del documento e, per la quinta domanda, ha trovato l'immagine richiesta in base al suo nome (presente come didascalia e nel testo), per la sesta, ha associato un nome ad ogni immagine insieme alla sua spiegazione, e, per la settima, ha associato un nome ad ogni formula insieme alla spiegazione dei parametri e del funzionamento, rispondendo in modo completo e spiegando nel dettaglio le informazioni rilevanti. Essendo identiche le risposte e corrette per tutti e tre i test, di seguito vengono riportati solamente i risultati

della prima prova.

5.1.3.1 Prima prova

Come detto poco fa, le tre prove svolte con questo strumento sono state le migliori e le uniche ad aver ottenuto al primo colpo le risposte corrette, spiegate in modo preciso.

1. What is the document about?

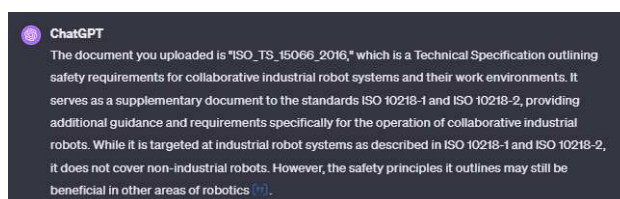


Figura 5.41: Prima domanda ChatGPT-4, inglese-inglese.

2. What is a collaborative robot?

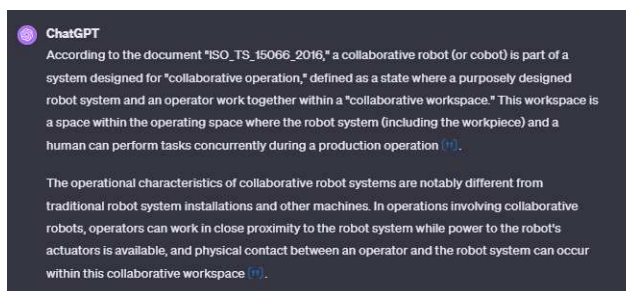


Figura 5.42: Seconda domanda ChatGPT-4, inglese-inglese.

3. What are the requirements of a collaborative robot system?

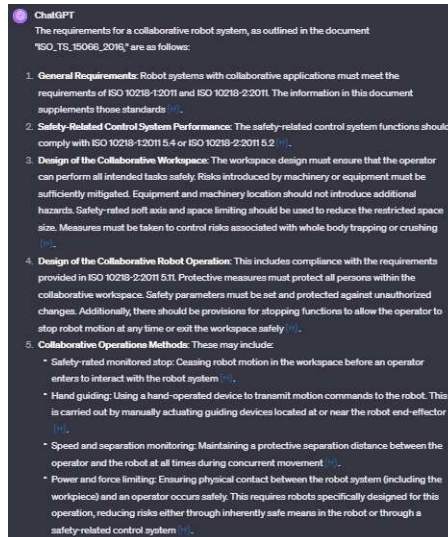


Figura 5.43: Terza domanda ChatGPT-4, inglese-inglese.

4. Make a summary of the main topic and explain it in simple words.

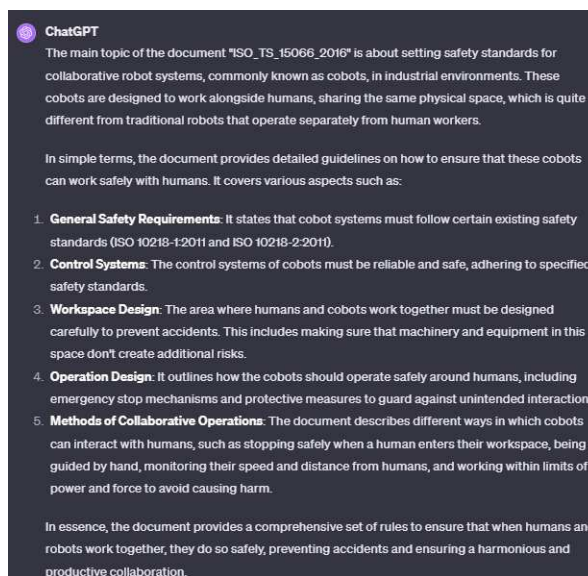


Figura 5.44: Quarta domanda ChatGPT-4, inglese-inglese.

5. Tell me about the second image ("Figure 2").

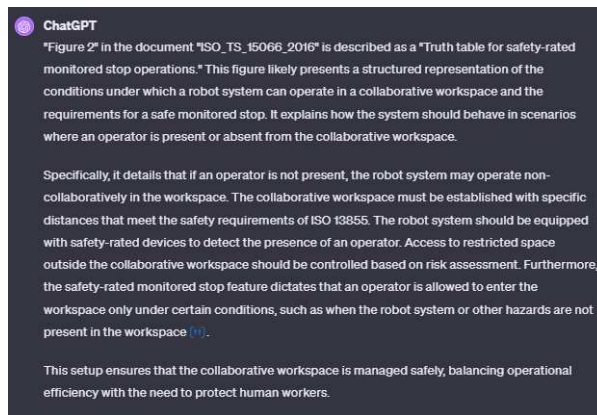


Figura 5.45: Quinta domanda ChatGPT-4, inglese-inglese.

6. How many are there and what are the images in the document about?
Give me an ordered list.

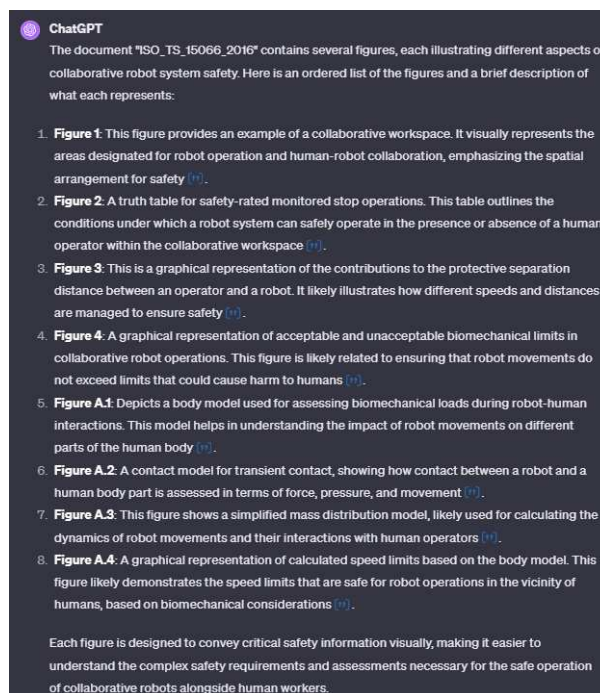


Figura 5.46: Sesta domanda ChatGPT-4, inglese-inglese.

7. How many are there and what are the formulas in the document about? Give me an ordered list.

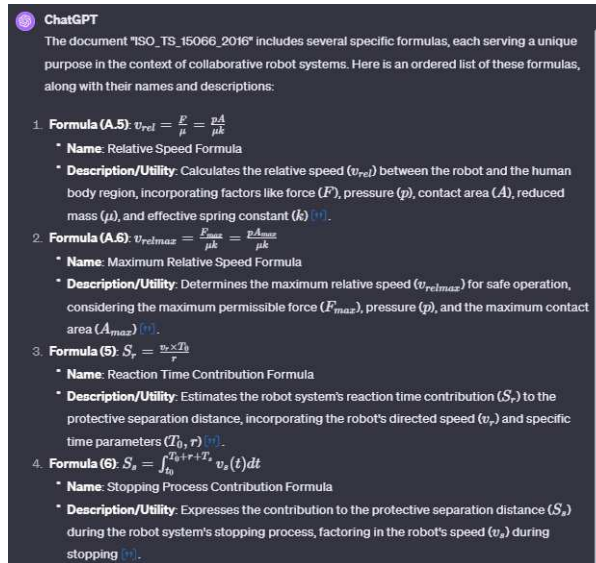


Figura 5.47: Settima domanda ChatGPT-4, prima parte, inglese-inglese.

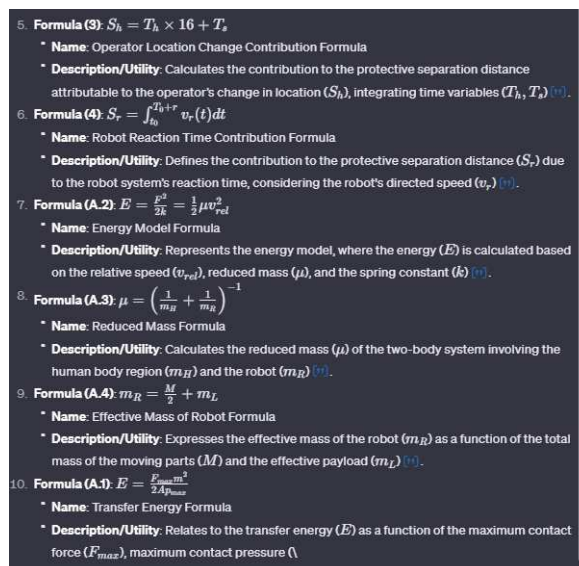


Figura 5.48: Settima domanda ChatGPT-4, seconda parte, inglese-inglese.

8. Tell me about the human body model used for testing to calculate risks (Figures A.1 and Table A.1).

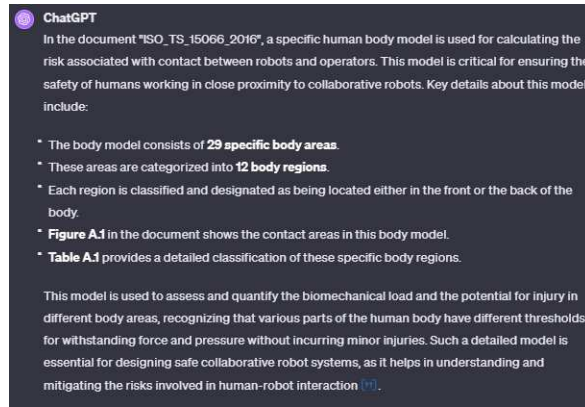


Figura 5.49: Ottava domanda ChatGPT-4, inglese-inglese.

9. What is the table "Table A.2" about?

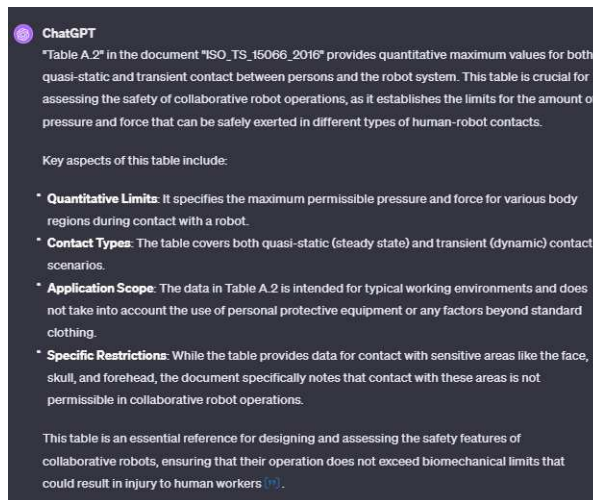


Figura 5.50: Nona domanda ChatGPT-4, inglese-inglese.

10. Based on the information gained from the paper, is a human obstructing the work of a robot in danger?

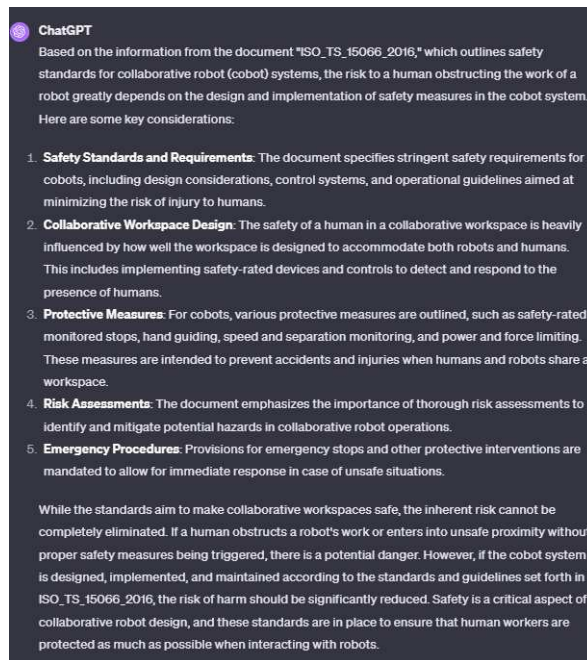


Figura 5.51: Decima domanda ChatGPT-4, inglese-inglese.

5.2 Caso di studio 2: Analisi di dati strutturati

In questa sezione, come anticipato a inizio capitolo, verranno illustrati e valutati i risultati ottenuti dall'analisi dei dati strutturati, mettendo a paragone le capacità dei due tool utilizzati per poter individuare quale sia il migliore in termini di affidabilità ed efficacia.

5.2.1 Risultati test PandasAI

In questo segmento della tesi, ci sarà l'esplorazione dei risultati derivanti dall'applicazione di un avanzato test di analisi su dati strutturati attraverso l'uso di PandasAI, uno strumento all'avanguardia nel campo dell'analisi dei dati. L'obiettivo principale di questa analisi era l'analisi di tendenze e potenziali anomalie all'interno del set di dati, utilizzando le sofisticate capacità di elaborazione e analisi di PandasAI. La scelta di PandasAI come strumento di analisi si è basata sulla sua notevole efficienza nel manipolo-

lare grandi volumi di dati strutturati, sulla sua flessibilità nell'esecuzione di complesse operazioni di dati e sulla sua potente integrazione con l'ecosistema Python per l'analisi dati. Tutte queste caratteristiche sono state ottenute, rispetto alla classica versione di Pandas, grazie all'introduzione di un modello di intelligenza artificiale. Attraverso un approccio metodico e dettagliato, è stato svolto il test con l'uso di una lista di domande, progettate per interrogare il dataset sotto vari aspetti, puntando a una comprensione più profonda delle dinamiche sottostanti. Dopo aver effettuato la conversione del file di partenza contenente il database da analizzare in un file *.csv*, è stato scritto uno script dedicato per effettuare il test, riportato di seguito.

```
database_analisi.py > ...
1 import pandas as pd
2 from pandasai import Agent, SmartDataframe
3 from pandasai.llm.openai import OpenAI
4
5 # Creo il dataframe
6 df = pd.read_csv(r"C:/Users/user/Desktop/TESI/PandasAI/SQL/database.csv")
7
8 # Inserisco la API per il LLM
9 llm = OpenAI(api_token="API Key da usare")
10
11 # Inizializzo un Agente e converso
12 agent = Agent(df, config={"llm": llm, "custom_whitelisted_dependencies":["prophet"]})
13 agent.chat("Domanda da fare.")
14
15 # Non uso l'Agente ma uno SmartDataframe
16 df = SmartDataframe(df, config={"llm": llm, "custom_whitelisted_dependencies":["prophet"]})
17 df.chat("Domanda da fare.")
```

Figura 5.52: Script per analisi dati tramite uso della libreria PandasAI.

Attraverso questo piccolo codice si va a richiamare il database con cui lavorare e lo si pone in una modalità di lettura, si inizializza un LLM, andando ad inserire una propria API (Application Programming Interface), e a seconda del tipo di domanda da fare si procede con o senza Agente; entrambe le tipologie sono riportate nella foto. Per evitare di annoiare il lettore con risultati ripetitivi, di seguito sono riportati i risultati ottenuti da un solo provino dei sei disponibili, in quanto le tipologie di risposta sono state le stesse e cambia solo il contenuto, cambiando il tipo di provino.

1. Analizza la magnitudine misurata per ogni tipologia di provino. Ogni tipo è riconosciuto dal relativo ID; diverso ID significa diverso tipo di

provino. Crea un grafico per ogni provino dove mostri l'andamento temporale del valore di magnitude.

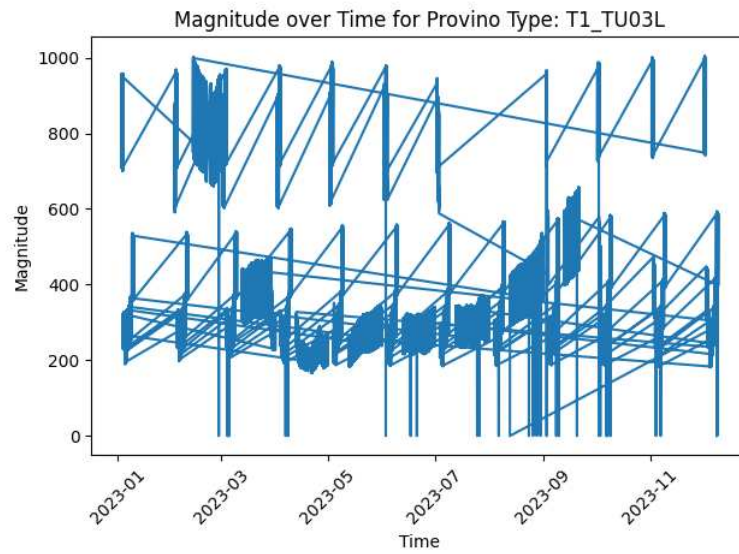


Figura 5.53: Prima domanda PandasAI.

2. Fai un forecast dei campioni. Quanto varrà la magnitude per i prossimi 50 istanti temporali di ogni provino?

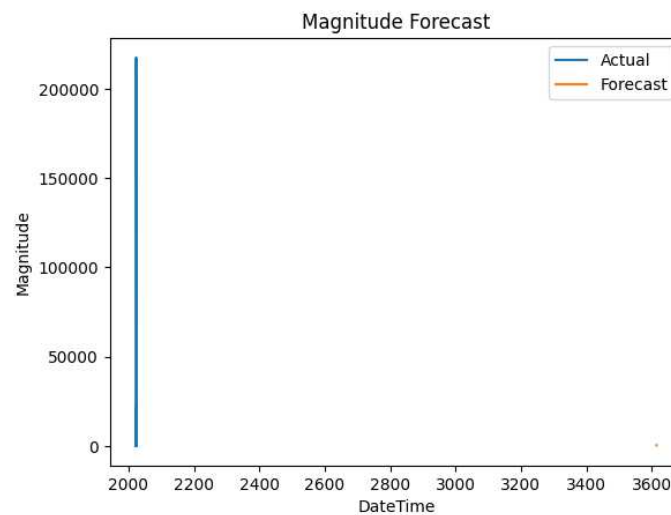


Figura 5.54: Seconda domanda PandasAI.

- Fai la stessa previsione usando la libreria Prophet.

Per questa domanda non è stato possibile ottenere una risposta in quanto l'installazione di nuove librerie è vietato. Non essendo presente di default la libreria Prophet è stato impossibile svolgere questa previsione.

- Fai la stessa previsione usando la libreria NeuralProphet.

Per questa domanda non è stato possibile ottenere una risposta in quanto l'installazione di nuove librerie è vietato. Non essendo presente di default la libreria NeuralProphet è stato impossibile svolgere questa previsione.

- Trovare il valor medio, max, mix e altri valori statistici relativi alla magnitude.

	ID	Magnitude
0	T1_TU03L	358.587266
1	T1_TU03S	269.126286
2	T3_TU04L	363.555402
3	T3_TU04S	316.056451
4	T4_TU04L	718.944822
5	T9_TU07S	866.792865

Figura 5.55: Terza domanda PandasAI, valore medio magnitude.

	ID	Magnitude
0	T1_TU03L	1004.55
1	T1_TU03S	818.94
2	T3_TU04L	986.50
3	T3_TU04S	3519.54
4	T4_TU04L	3838.99
5	T9_TU07S	217449.76

Figura 5.56: Terza domanda PandasAI, valore massimo magnitude.

	ID	Magnitude
0	T1_TU03L	0.0
1	T1_TU03S	0.0
2	T3_TU04L	0.0
3	T3_TU04S	0.0
4	T4_TU04L	0.0
5	T9_TU07S	0.0

Figura 5.57: Terza domanda PandasAI, valore massimo magnitude.

	ID	Magnitude
0	T1_TU03L	172.662078
1	T1_TU03S	131.166107
2	T3_TU04L	219.919439
3	T3_TU04S	388.114360
4	T4_TU04L	690.568872
5	T9_TU07S	4227.487998

Figura 5.58: Terza domanda PandasAI, valore deviazione standard magnitude.

6. Quante volte i provini sono entrati nel fenomeno dell'isteresi?

Per questa domanda non è stato possibile ottenere una risposta. Nonostante l'aggiunta ulteriore di informazioni non è stato possibile completare la richiesta.

7. Genera grafici con soglie per isteresi.

Come per la precedente, ovviamente, non è stato possibile generare un grafico per analizzare l'andamento del fenomeno di isteresi non essendoci stata la risposta alla domanda prima.

8. Considera gli ultimi 500 campioni per i provini e studia l'andamento statistico della magnitude.

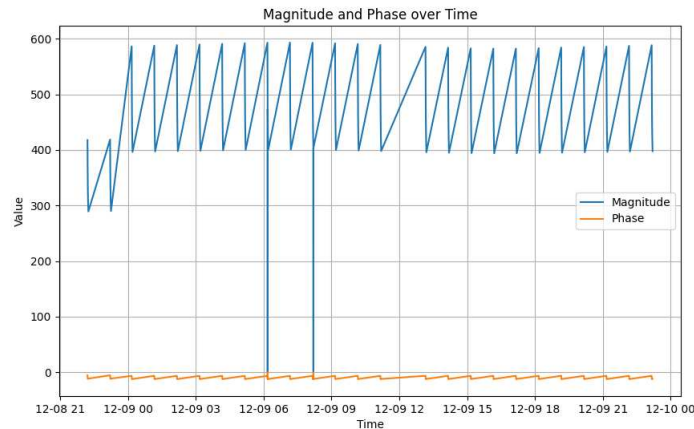


Figura 5.59: Quinta domanda PandasAI, analisi statistica.

9. Genera il codice per automatizzare questa analisi appena svolta sui 500 campioni.
10. Genera il codice per automatizzare il forercasting fatto sui provini.

Per quanto riguarda le ultime due domande, è stato impossibile ottenere una generazione di codice per poter automatizzare l'analisi statistica o la previsione fatta sui provini. Risultano evidenti, dunque, i limiti presenti in questa nuova ed emergente libreria. Sicuramente, nel corso del tempo assisteremo a delle miglorie continue che lo renderanno uno strumento valido per effettuare analisi di dati, sia di tipo strutturato che non.

5.2.2 Risultati test ChatGPT-4, dati strutturati

Nel contesto dell'evoluzione tecnologica e dell'intelligenza artificiale, l'impiego di modelli avanzati di linguaggio come ChatGPT-4 ha aperto nuove frontiere nell'analisi di dati strutturati. Questa sezione della tesi si propone di esplorare e riportare i risultati ottenuti dall'utilizzo di questo strumento per l'analisi del database analizzato nel paragrafo precedente. L'obiettivo è quello di valutare la capacità del modello di generare intuizioni, predizioni e interpretazioni accurate. L'analisi ha inteso sfruttare le potenzialità di

ChatGPT-4 nell'elaborazione del linguaggio naturale e nella comprensione dei contesti, applicando queste capacità all'esame di dataset specifici. Attraverso un'approccio metodologico rigoroso, sono state esaminate le prestazioni di ChatGPT-4 in termini di efficienza, precisione e applicabilità in scenari reali, ponendo particolare attenzione alle sfide, ai limiti e alle potenzialità offerte dal modello. I risultati di questa indagine forniscono un contributo significativo alla letteratura esistente, evidenziando come l'intelligenza artificiale, e in particolare i modelli avanzati di elaborazione del linguaggio naturale, possano essere efficacemente impiegati nell'analisi di dati complessi, aprendo nuove prospettive per la ricerca e lo sviluppo in svariati campi di applicazione. Anche in questo caso, per evitare ripetizioni, verranno riportati solo i risultati relativi al dataframe relativo al primo provino dei cinque disponibili; è importante ricordare che inizialmente i provini erano sei ma dopo le fasi di filtraggio ne sono rimasti cinque utili ai nostri fini.

1. Crea un grafico dove mostri l'andamento temporale del valore di magnitude.

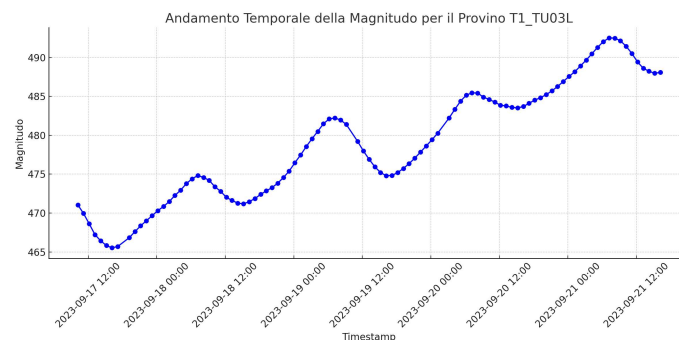


Figura 5.60: Prima domanda ChatGPT-4 dati strutturati, andamento temporale magnitude.

2. Fai un forecast del campione. Quanto varrà la magnitude per i prossimi 50 campioni temporali?

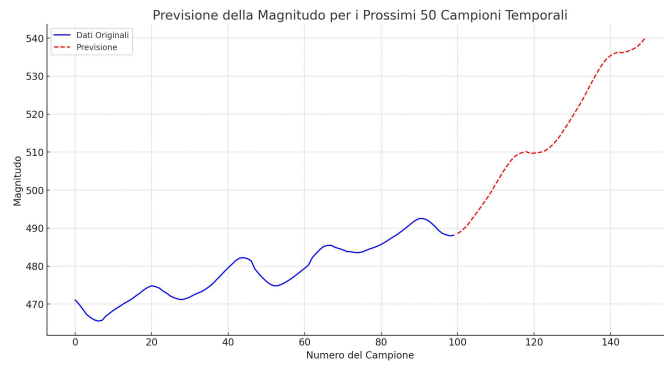


Figura 5.61: Seconda domanda ChatGPT-4 dati strutturati, grafico previsione temporale magnitudo.

Campione	Previsione Magnitudo
0	488.53
1	489.25
2	490.17
3	491.30
4	492.69
5	494.03
6	495.43
7	496.85
8	498.30
9	499.86

Figura 5.62: Seconda domanda ChatGPT-4 dati strutturati, tabella previsione temporale magnitudo.

3. Fai la stessa previsione usando la libreria Prophet.

Per questa domanda non è stato possibile ottenere una risposta in quanto l'installazione di nuove librerie è vietato. Non essendo presente di default la libreria Prophet è stato impossibile svolgere questa previsione. Al contrario della prova precedente però, in modo automatico restituisce come risposta un codice Python da usare nell'ambiente di programmazione in modo da ottenere quanto richiesto.

4. Fai la stessa previsione usando la libreria Neural Prophet.

Per questa domanda non è stato possibile ottenere una risposta in quanto l'installazione di nuove librerie è vietato. Non essendo presente di default la libreria NeuralProphet è stato impossibile svolgere questa previsione. Al contrario della prova precedente però,

in modo automatico restituisce come risposta un codice Python da usare nell'ambiente di programmazione in modo da ottenere quanto richiesto.

5. Genera un codice usando python per automatizzare queste previsioni, per i due approcci. Per ciascuno, genera un codice separato.

```
forecast_prophet_T1_TU03L.py > ...
1 import pandas as pd
2 from prophet import Prophet
3 import matplotlib.pyplot as plt
4
5 # Caricamento dei dati
6 file_path = 'C:/Users/user/Desktop/TESI/ChatGPT 4 post PandasAI/Dataframe filtrati per provino/T1_TU03L_10000Hz.csv'
7 data = pd.read_csv(file_path)
8
9 # Pre-elaborazione dei dati
10 data['ds'] = pd.to_datetime(data['Date'] + ' ' + data['Time'])
11 data = data[['ds', 'Magnitude']]
12 data.rename(columns={'Magnitude': 'y'}, inplace=True)
13
14 # Creazione e addestramento del modello
15 model = Prophet()
16 model.fit(data)
17
18 # Creazione del DataFrame per le previsioni future
19 future = model.make_future_dataframe(periods=50, freq='H')
20
21 # Previsione dei valori futuri
22 forecast = model.predict(future)
23
24 # Estrazione delle previsioni per i prossimi 50 campioni temporali
25 next_50_samples = forecast[['ds', 'yhat']].tail(50)
26
27 # Visualizzazione dei risultati in un grafico
28 fig1 = model.plot(forecast)
29 plt.title('Previsioni Magnitudo')
30 plt.tight_layout() # Ottimizza il layout per evitare il taglio delle parti del grafico
31 plt.savefig('forecast_graph.png', bbox_inches='tight') # Assicurati che tutto il grafico sia incluso
32
33 # Dividere la tabella in due parti se necessario
34 middle_index = len(next_50_samples) // 2
35 tables = [next_50_samples.iloc[:middle_index], next_50_samples.iloc[middle_index:]]
36
37 # Visualizzazione dei risultati in due tabelle separate
38 for i, table in enumerate(tables):
39     fig, ax = plt.subplots(figsize=(10, 6))
40     ax.axis('tight')
41     ax.axis('off')
42     table_data = table.values
43     table_columns = table.columns
44     ax.table(cellText=table_data, colLabels=table_columns, loc='center')
45     plt.savefig(f'forecast_table_part_{i+1}.png')
46
47 # Salvataggio dei risultati delle previsioni in un file CSV
48 next_50_samples.to_csv('forecast.csv', index=False)
```

Figura 5.63: Quinta domanda ChatGPT-4 dati strutturati, codice automazione previsione con Prophet.

```

forecast_neural_prophet_T1_TU03L.py > ...
6 # Caricamento dei dati
7 file_path = 'C:/Users/user/Desktop/TESI/ChatGPT 4 post PandasAI/Dataframe filtrati per provino/T1_TU03L_1000Hz.csv'
8 df = pd.read_csv(file_path)
9
10 # Unione delle colonne Date e Time in una singola colonna datetime e conversione in formato datetime
11 df['ds'] = pd.to_datetime(df['Date'] + ' ' + df['Time'], dayfirst=True)
12
13 # Rinominazione della colonna Magnitude in y per adattarla a NeuralProphet
14 df = df.rename(columns={'Magnitude': 'y'})
15
16 # Mantenere solo le colonne 'ds' e 'y' nel dataframe
17 df = df[['ds', 'y']]
18
19 # Addestramento del modello
20 m = NeuralProphet()
21 metrics = m.fit(df, freq='H') # La frequenza 'H' sta per ora, adatta se necessario
22
23 # Creazione delle future date per il forecasting
24 future = m.make_future_dataframe(df, periods=50)
25 forecast = m.predict(future)
26
27 # Visualizzazione e salvataggio del grafico delle previsioni
28 fig_forecast = m.plot(forecast)
29 pio.write_image(fig_forecast, 'C:/Users/user/Desktop/TESI/ChatGPT 4 post PandasAI/Codice/forecast_plot.png')
30
31 # Visualizzazione e salvataggio del grafico dei componenti
32 fig_components = m.plot_components(forecast)
33 pio.write_image(fig_components, 'C:/Users/user/Desktop/TESI/ChatGPT 4 post PandasAI/Codice/components_plot.png')
34
35 # Salvataggio dei risultati
36 forecast[['ds', 'yhat1']].to_csv('C:/Users/user/Desktop/TESI/ChatGPT 4 post PandasAI/Codice/forecast.csv', index=False)
37
38 """ # Salvataggio della tabella dei risultati come immagine
39 fig, ax = plt.subplots(figsize=(10, 4)) # Crea una nuova figura
40 ax.axis('tight')
41 ax.axis('off')
42 the_table = ax.table(cellText=forecast[['ds', 'yhat1']].values, colLabels=forecast[['ds', 'yhat1']].columns, loc='center')
43 plt.savefig('C:/Users/user/Desktop/TESI/ChatGPT 4 post PandasAI/Codice/results_table.png')
44 plt.close(fig) # Chiude la figura """
45
46 # Impostazioni per la tabella
47 max_rows_per_image = 25 # Numero massimo di righe per immagine
48 total_rows = forecast.shape[0]
49 num_images = (total_rows + max_rows_per_image - 1) // max_rows_per_image # Numero di immagini necessarie
50
51 # Suddivisione e salvataggio della tabella dei risultati come immagini
52 for i in range(num_images):
53     fig, ax = plt.subplots(figsize=(10, min(4 + max_rows_per_image, 4 + (total_rows - i * max_rows_per_image))))
54     ax.axis('tight')
55     ax.axis('off')
56     start_row = i * max_rows_per_image
57     end_row = min(start_row + max_rows_per_image, total_rows)
58     the_table = ax.table(cellText=forecast[['ds', 'yhat1']].values[start_row:end_row],
59                        colLabels=forecast[['ds', 'yhat1']].columns,
60                        loc='center')
61     plt.savefig(f'C:/Users/user/Desktop/TESI/ChatGPT 4 post PandasAI/Codice/results_table_part_{i+1}.png')
62     plt.close(fig)
63
64 # Stampa del percorso dei file salvati
65 print('Files saved:')
66 print('C:/Users/user/Desktop/TESI/ChatGPT 4 post PandasAI/Codice/forecast.csv')
67 print('C:/Users/user/Desktop/TESI/ChatGPT 4 post PandasAI/Codice/forecast_plot.png')
68 print('C:/Users/user/Desktop/TESI/ChatGPT 4 post PandasAI/Codice/components_plot.png')
69 print('C:/Users/user/Desktop/TESI/ChatGPT 4 post PandasAI/Codice/results_table.csv')

```

Figura 5.64: Quinta domanda ChatGPT-4 dati strutturati, codice automazione previsione con NeuralProphet.

6. Trova il valor medio, max, mix e altri valori statistici dei valori raccolto nel dataframe.

Statistica	Valore
Conteggio	100.00
Media	478.92
Minimo	465.53
Massimo	492.51
Mediana	478.27
Deviazione Standard	7.48
25° Percentile	472.84
75° Percentile	484.85

Figura 5.65: Sesta domanda ChatGPT-4 dati strutturati, tabella analisi statistica.

7. Considera il valore medio della magnitudine. Quando il valore di una misurazione supera il 2% del valore medio considera l'ingresso in isteresi. Quante volte il provino è entrato in isteresi? Quando? Riportami le informazioni relative a ora, data e valore di magnitudine delle misurazioni.

ChatGPT
 Il valore medio della magnitudine è circa 478.92. Impostando la soglia per l'ingresso in isteresi al 2% sopra il valore medio, che corrisponde a circa 488.49, abbiamo determinato che il provino è entrato in isteresi in 12 delle misurazioni registrate.

You
 Quando? Riportami le informazioni relative a ora, data e valore di magnitudine delle misurazioni.

ChatGPT
 Il provino è entrato in isteresi nelle seguenti misurazioni:

1. **Data:** 21-09-2023, **Ora:** 02:10:42, **Magnitudine:** 488.90
2. **Data:** 21-09-2023, **Ora:** 03:11:10, **Magnitudine:** 489.66
3. **Data:** 21-09-2023, **Ora:** 04:11:03, **Magnitudine:** 490.47
4. **Data:** 21-09-2023, **Ora:** 05:11:02, **Magnitudine:** 491.30
5. **Data:** 21-09-2023, **Ora:** 06:11:03, **Magnitudine:** 492.03
6. **Data:** 21-09-2023, **Ora:** 07:11:03, **Magnitudine:** 492.51
7. **Data:** 21-09-2023, **Ora:** 08:11:16, **Magnitudine:** 492.49
8. **Data:** 21-09-2023, **Ora:** 09:11:08, **Magnitudine:** 492.14
9. **Data:** 21-09-2023, **Ora:** 10:11:09, **Magnitudine:** 491.43
10. **Data:** 21-09-2023, **Ora:** 11:11:08, **Magnitudine:** 490.52
11. **Data:** 21-09-2023, **Ora:** 12:11:03, **Magnitudine:** 489.42
12. **Data:** 21-09-2023, **Ora:** 13:11:03, **Magnitudine:** 488.61

Questi dati indicano le specifiche istanze in cui è stato registrato l'ingresso in isteresi, basato sulla soglia del 2% sopra il valore medio della magnitudine.

Figura 5.66: Settima domanda ChatGPT-4 dati strutturati, analisi isteresi.

8. Genera un grafico con le soglie per isteresi. Tieni pulito l'asse X inserendo solo le misurazioni rilevanti.

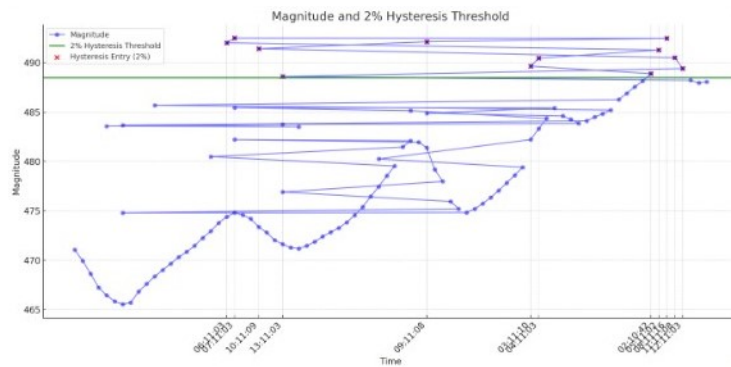


Figura 5.67: Ottava domanda ChatGPT-4 dati strutturati, grafico analisi isteresi.

9. Considera i campioni temporali per il provino e studia l'andamento statistico.

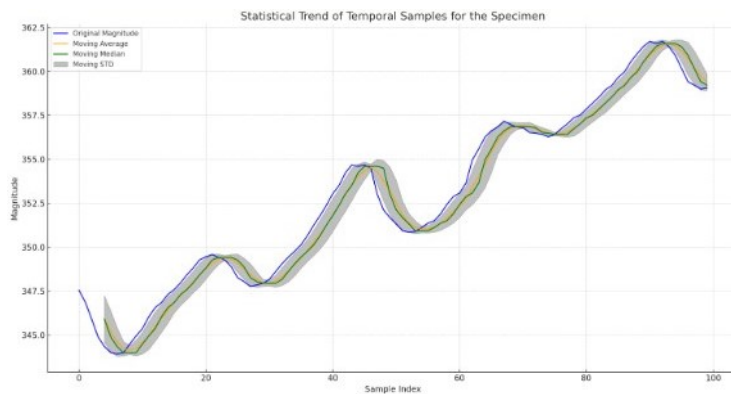


Figura 5.68: Nona domanda ChatGPT-4 dati strutturati, analisi statistica.

10. Genera un codice usando python per automatizzare questa analisi statistica.

```

1 import pandas as pd
2 import seaborn as sns
3 import matplotlib.pyplot as plt
4 #from pandas.plotting import table
5
6 # Caricamento dei dati
7 file_path = 'C:/Users/user/Desktop/TESE/ChatGPT 4 post PandasAI/Dataframe filtrati per provino/T1_TU03L_1000Hz.csv'
8 data = pd.read_csv(file_path)
9
10 # Calcolo delle statistiche descrittive
11 desc_stats = data.describe()
12
13 # Creazione di una figura per visualizzare la tabella
14 fig, ax = plt.subplots(figsize=(12, 4)) # Regola la dimensione della figura se necessario
15 ax.axis('tight')
16 ax.axis('off')
17 the_table = ax.table(cellText=desc_stats.values, colLabels=desc_stats.columns, rowLabels=desc_stats.index, cellloc='center', loc='center')
18
19 # Regolazione delle proprietà delle celle della tabella per migliorare la leggibilità
20 the_table.auto_set_font_size(False)
21 the_table.set_fontsize(5) # Regola la dimensione del font se necessario
22 the_table.auto_set_column_width(col=list(range(len(desc_stats.columns)))) # Regola automaticamente la larghezza delle colonne
23
24 plt.title('Statistiche Descrittive', fontsize=14, pad=10)
25
26 # Salvataggio della tabella come immagine
27 plt.savefig('desc_stats.png', dpi=300, bbox_inches='tight', pad_inches=0.05)
28 plt.show()
29
30 # Creazione dell'istogramma della magnitudo
31 plt.figure(figsize=(10, 6))
32 sns.histplot(data["Magnitudo"], kde=True, color='blue')
33 plt.title("Distribuzione della Magnitudo")
34 plt.xlabel("Magnitudo")
35 plt.ylabel("Frequenza")
36 plt.grid(True)
37 plt.show()
38
39 # Calcolo e visualizzazione della matrice di correlazione
40 correlation_matrix = data.corr()
41 plt.figure(figsize=(10, 8))
42 sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
43 plt.title("Matrice di Correlazione")
44 plt.show()

```

Figura 5.69: Decima domanda ChatGPT-4 dati strutturati, codice per analisi statistica.

È evidente come il miglior risultato per l'analisi di dati di tipo strutturato sia stato ottenuto con l'ausilio di ChatGPT4, essendo uno strumento dotato di elevata potenza che ha permesso di ragionare ed estrapolare i dati nel modo più efficiente possibile, risultando il mezzo più affidabile attualmente per questo tipo di lavoro.

Capitolo 6

Conclusione

In questa sezione conclusiva verrà presentato, nella prima parte, un riassunto dei risultati ottenuti nei test di analisi dati non strutturati e strutturati e, nella seconda, i possibili sviluppi futuri.

6.1 Sintesi dei risultati ottenuti

In conclusione, l'analisi condotta sulla base dei dati raccolti ha permesso di analizzare con precisione le dinamiche inerenti all'analisi di dati, attraverso uno studio incentrato sui dati non strutturati e uno sui dati strutturati. L'indagine sui dati non strutturati ha rivelato pattern e tendenze nascoste, offrendo una comprensione delle capacità e dei limiti attuali. Infatti, nonostante ci sia una vasta gamma di strumenti dedicati all'analisi di questa tipologia di dati, risulta troppo limitante il loro impegno in quanto non è possibile svolgere in modo agevole degli studi senza incorrere in bug, crash o risultati di analisi evidentemente errati. Come dimostrato nella sezione dedicata allo svolgimento delle prove, capitolo 4.2, l'unico tool che è possibile adottare, grazie agli enormi progressi fatti rispetto alle sue versioni precedenti e ai continui aggiornamenti, è ChatGPT-4, che ha permesso di ottenere una analisi completa in modo agevole ed efficiente. Al contrario, tutti gli altri strumenti messi alla prova si sono dimostrati inadatti in più

fasi del test. È evidente che non si possa fare affidamento su questi ultimi se non richiedendo una continua e attenta verifica dei risultati, costringendo anche chi lavora a dover svolgere diverse analisi in modo autonomo e senza ausilio della tecnologia. Parallelamente, l'analisi dei dati strutturati ha fornito una solida base quantitativa per validare ipotesi e misurare impatti con precisione. Nell'analisi comparativa tra PandasAI e ChatGPT-4, infatti, è emerso come PandasAI abbia incontrato significative difficoltà nel gestire complesse richieste di analisi, portando a risultati insoddisfacenti e mancanti. Al contrario, ChatGPT-4 ha dimostrato una sorprendente capacità di navigare e interpretare grandi volumi di dati strutturati con facilità, evidenziando la sua avanzata comprensione dei contesti e delle sfide poste dall'analisi dei big data. Questo successo non solo sottolinea l'efficacia dell'approccio di ChatGPT-4 all'elaborazione dei dati, ma segnala anche una svolta promettente per il futuro dell'analisi dei dati, rendendo il processo più accessibile e gestibile per gli utenti di vari livelli di competenza. I test effettuati, dunque, hanno confermato alcune delle ipotesi iniziali e creato delle aspettative future. Questo lavoro ha dimostrato come, combinando analisi qualitative e quantitative, sia possibile effettuare analisi dati in modo completo ed efficiente riducendo notevolmente la complessità di approccio, che fino a poco tempo fa avrebbe richiesto un'indagine approfondita.

6.2 Sviluppi futuri

Nel corso dello svolgimento delle prove trattate durante questo progetto e dell'analisi dei risultati ottenuti, sono emersi diversi spunti per possibili sviluppi futuri che potrebbero consentire un ulteriore miglioramento in ottica di analisi di dati, sia di tipo strutturato che non. Ad esempio, per quanto riguarda i possibili sviluppi nell'analisi di dati non strutturati si potrebbe pensare di esplorare come l'adattamento fine dei modelli di machine learning possa migliorare significativamente la loro efficacia nell'analisi dei

dati strutturati. Il fine tuning, ovvero l'aggiustamento mirato di un modello pre-addestrato su un nuovo set di dati o per specifiche esigenze, si prospetta come un approccio fondamentale nel futuro dell'analisi dati. Questo processo non solo rende i modelli più adatti al contesto applicativo specifico, ma li rende anche più efficienti nell'identificare pattern e correlazioni significative all'interno di grandi volumi di dati strutturati. Attraverso tecniche avanzate di fine tuning, infatti, è possibile ottimizzare ulteriormente i modelli per migliorare la precisione, la velocità e la pertinenza delle analisi. In futuro, l'evoluzione di queste tecniche, combinata con l'integrazione di algoritmi di apprendimento automatico più sofisticati e la crescente capacità di calcolo, potrebbe portare a sistemi di analisi dati non strutturati incredibilmente precisi, in grado di fornire insight più profondi e azionabili in tempo reale. Questo non solo accelererà il processo decisionale in vari settori, ma aprirà anche nuove frontiere nella personalizzazione dei servizi e nella previsione delle tendenze di mercato, evidenziando l'importanza cruciale del fine tuning nella prossima generazione di strumenti di analisi dati. Per quanto riguarda i possibili sviluppi nell'analisi di dati strutturati, invece, si potrebbe pensare all'adozione di modelli computazionali sempre più sofisticati e personalizzabili, capaci di interpretare la complessità e la variabilità dei dati in maniera efficace. In questo contesto, un'area di particolare interesse riguarda l'integrazione di sistemi di intelligenza artificiale avanzata, in grado di generare dinamicamente codice SQL per interrogazioni complesse, ottimizzando l'accesso e l'analisi delle informazioni contenute nelle basi di dati. Questo approccio non solo permetterà di sfruttare al meglio le capacità di storage e di elaborazione dei moderni database, ma aprirà anche la strada a nuove forme di interazione con i dati, dove gli utenti potranno formulare richieste analitiche in linguaggio naturale che verranno tradotte automaticamente in query SQL efficienti. Parallelamente, l'evoluzione dei modelli predittivi e di classificazione, attraverso l'impiego di tecniche di machine learning e deep learning, contribuirà a rafforzare le capacità di

estrazione di insight rilevanti, migliorando la presa di decisioni basata sui dati. Inoltre, l'attenzione verso la sicurezza e la privacy dei dati guiderà lo sviluppo di soluzioni che garantiscano la protezione delle informazioni sensibili durante l'intero processo di analisi. In sintesi, gli sviluppi futuri nell'analisi di dati strutturati promettono di rivoluzionare il modo in cui accediamo, analizziamo e interpretiamo le grandi quantità di informazioni, guidando verso un'era di maggiore efficienza, precisione e sicurezza nell'uso dei dati.

6.3 Considerazioni generali finali

Il lavoro di tesi presentato si è concentrato sull'esplorazione del potenziale del *prompt engineering* applicato a Large Language Models (LLM) nell'ambito dell'ingegneria, con un focus specifico sull'elaborazione e l'interpretazione di dati sia *strutturati* che *non strutturati*. Attraverso la valutazione di diversi tool disponibili attualmente, è stato possibile dimostrare come tecniche avanzate di *prompt engineering* possano significativamente migliorare la capacità degli LLM di comprendere, elaborare e generare informazioni utili da vasti set di dati eterogenei. Questo ha implicazioni profonde non solo per la ricerca futura ma anche per applicazioni pratiche in svariati campi dell'ingegneria. In primo luogo, i risultati ottenuti confermano l'importanza del *prompt engineering* come strumento per affinare le prestazioni degli LLM, permettendo una maggiore precisione e rilevanza nelle risposte generate. La capacità di modulare l'input fornito agli LLM ha dimostrato di essere cruciale per massimizzare la loro efficacia, specialmente quando si trattano dati complessi e di varia natura. In secondo luogo, l'applicazione di questi modelli ai dati *non strutturati* ha rivelato nuove possibilità per l'estrazione di informazioni e la generazione di insight, superando alcune delle sfide tradizionalmente associate all'analisi di tali dati ed evidenziando le sfide attuali da dover superare. Questo apre la strada a nuove metodologie di ricerca e sviluppo in campi ingegneristici che sono necessarie e

permetteranno di ottenere una gestione efficiente dei dati *non strutturati*. Infine, l'impiego di LLM nei dati strutturati ha evidenziato come questi modelli possano servire non solo per l'interpretazione e l'analisi ma anche per la previsione e la simulazione integrando anche modelli basati su deep learning come Prophet e NeuralProphet. In conclusione, il lavoro ha dimostrato che il *prompt engineering* applicato agli LLM rappresenta un campo promettente e ricco di opportunità per l'avanzamento dell'ingegneria. La capacità di questi modelli di lavorare efficacemente con dati di varia natura apre nuove frontiere per la ricerca e l'applicazione pratica, sottolineando il ruolo centrale che l'intelligenza artificiale continuerà a giocare nell'evoluzione tecnologica e ingegneristica.

Elenco delle figure

1.1	Differenza tra dati strutturati e non strutturati.	8
2.1	Storia dell'industria.	13
2.2	Esempio basilare di un prompt.	20
2.3	Esempio basilare di un prompt.	20
2.4	Esempio sulla specificità di un prompt.	23
2.5	Esempio Chain of Thoughts prompt.	27
2.6	Esempio di Prompt Generated Knowledge.	28
2.7	Esempio di Tree of Thoughts.	29
2.8	Esempio di Prompt Multimodale.	32
3.1	Logo AskYourPDF.	37
3.2	Logo ChatDOC.	38
3.3	Logo ChatPDF.	38
3.4	Logo LightPDF.	39
3.5	Logo ChatGPT-3.5.	39
3.6	Logo ChatGPT-4.	40
3.7	Logo PandasAI.	42
4.1	Struttura Database ReCity.	48
4.2	Database utilizzato nella prova di analisi dati strutturati. . .	49
4.3	Dataframe ottenuti dopo l'analisi del database.	52
5.1	Prima domanda AskYourPDF, inglese-inglese.	57
5.2	Seconda domanda AskYourPDF, inglese-inglese.	57

5.3	Terza domanda AskYourPDF, inglese-inglese.	58
5.4	Quarta domanda AskYourPDF, inglese-inglese.	58
5.5	Quinta domanda AskYourPDF, primo tipo, inglese-inglese. . .	59
5.6	Quinta domanda AskYourPDF, secondo tipo, inglese-inglese.	59
5.7	Sesta domanda AskYourPDF, inglese-inglese.	59
5.8	Settima domanda AskYourPDF, inglese-inglese.	59
5.9	Ottava domanda AskYourPDF, inglese-inglese.	60
5.10	Prima domanda AskYourPDF, inglese-italiano.	60
5.11	Seconda domanda AskYourPDF, inglese-italiano.	61
5.12	Terza domanda AskYourPDF, inglese-italiano.	61
5.13	Quarta domanda AskYourPDF, inglese-italiano.	62
5.14	Quinta domanda AskYourPDF, inglese-italiano.	62
5.15	Sesta domanda AskYourPDF, inglese-italiano.	62
5.16	Settima domanda AskYourPDF, inglese-italiano.	63
5.17	Ottava domanda AskYourPDF, inglese-italiano.	63
5.18	Nona domanda AskYourPDF, inglese-italiano.	63
5.19	Decima domanda AskYourPDF, inglese-italiano.	64
5.20	Prima domanda ChatPDF, inglese-inglese.	65
5.21	Seconda domanda ChatPDF, inglese-inglese.	65
5.22	Terza domanda ChatPDF, inglese-inglese.	65
5.23	Quarta domanda ChatPDF, inglese-inglese.	66
5.24	Quinta domanda ChatPDF, inglese-inglese.	66
5.25	Sesta domanda ChatPDF, inglese-inglese.	66
5.26	Settima domanda ChatPDF, inglese-inglese.	67
5.27	Ottava domanda ChatPDF, inglese-inglese.	67
5.28	Nona domanda ChatPDF, inglese-inglese.	67
5.29	Decima domanda ChatPDF, inglese-inglese.	68
5.30	Prima domanda ChatPDF, inglese-italiano.	68
5.31	Seconda domanda ChatPDF, inglese-italiano.	68
5.32	Terza domanda ChatPDF, inglese-italiano.	69

5.33	Quarta domanda ChatPDF, inglese-italiano.	69
5.34	Quinta domanda ChatPDF, inglese-italiano.	69
5.35	Sesta domanda ChatPDF, inglese-italiano.	70
5.36	Settima domanda ChatPDF, inglese-italiano.	70
5.37	Ottava domanda ChatPDF, inglese-italiano.	70
5.38	Nona domanda ChatPDF, inglese-italiano.	71
5.39	Decima domanda ChatPDF, inglese-italiano.	71
5.40	Sesta domanda ChatPDF, italiano-italiano.	72
5.41	Prima domanda ChatGPT-4, inglese-inglese.	73
5.42	Seconda domanda ChatGPT-4, inglese-inglese.	73
5.43	Terza domanda ChatGPT-4, inglese-inglese.	74
5.44	Quarta domanda ChatGPT-4, inglese-inglese.	74
5.45	Quinta domanda ChatGPT-4, inglese-inglese.	75
5.46	Sesta domanda ChatGPT-4, inglese-inglese.	75
5.47	Settima domanda ChatGPT-4, prima parte, inglese-inglese.	76
5.48	Settima domanda ChatGPT-4, seconda parte, inglese-inglese.	76
5.49	Ottava domanda ChatGPT-4, inglese-inglese.	77
5.50	Nona domanda ChatGPT-4, inglese-inglese.	77
5.51	Decima domanda ChatGPT-4, inglese-inglese.	78
5.52	Script per analisi dati tramite uso della libreria PandasAI.	79
5.53	Prima domanda PandasAI.	80
5.54	Seconda domanda PandasAI.	80
5.55	Terza domanda PandasAI, valore medio magnitude.	81
5.56	Terza domanda PandasAI, valore massimo magnitude.	81
5.57	Terza domanda PandasAI, valore massimo magnitude.	82
5.58	Terza domanda PandasAI, valore deviazione standard magnitude.	82
5.59	Quinta domanda PandasAI, analisi statistica.	83
5.60	Prima domanda ChatGPT-4 dati strutturati, andamento temporale magnitude.	84

5.61	Seconda domanda ChatGPT-4 dati strutturati, grafico previsione temporale magnitude.	85
5.62	Seconda domanda ChatGPT-4 dati strutturati, tabella previsione temporale magnitude.	85
5.63	Quinta domanda ChatGPT-4 dati strutturati, codice automazione previsione con Prophet.	86
5.64	Quinta domanda ChatGPT-4 dati strutturati, codice automazione previsione con NeuralProphet.	87
5.65	Sesta domanda ChatGPT-4 dati strutturati, tabella analisi statistica.	88
5.66	Settima domanda ChatGPT-4 dati strutturati, analisi isteresi.	88
5.67	Ottava domanda ChatGPT-4 dati strutturati, grafico analisi isteresi.	89
5.68	Nona domanda ChatGPT-4 dati strutturati, analisi statistica.	89
5.69	Decima domanda ChatGPT-4 dati strutturati, codice per analisi statistica.	90

Bibliografia

- [1] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models, 2023.
- [2] Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large language models are zero-shot reasoners, 2023.
- [3] Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models, 2023.
- [4] Jiacheng Liu, Alisa Liu, Ximing Lu, Sean Welleck, Peter West, Ronan Le Bras, Yejin Choi, and Hannaneh Hajishirzi. Generated knowledge prompting for commonsense reasoning, 2022.
- [5] Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models, 2023.
- [6] Jieyi Long. Large language model guided tree-of-thought, 2023.
- [7] Dave Hulbert. Using Tree-of-Thought Prompting to boost ChatGPT’s reasoning, May 2023.
- [8] Yongchao Zhou, Andrei Ioan Muresanu, Ziwen Han, Keiran Paster, Silviu Pitis, Harris Chan, and Jimmy Ba. Large language models are human-level prompt engineers, 2023.

- [9] Shizhe Diao, Pengcheng Wang, Yong Lin, and Tong Zhang. Active prompting with chain-of-thought for large language models, 2023.
- [10] Zekun Li, Baolin Peng, Pengcheng He, Michel Galley, Jianfeng Gao, and Xifeng Yan. Guiding large language models via directional stimulus prompting, 2023.
- [11] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models, 2023.
- [12] Zhuosheng Zhang, Aston Zhang, Mu Li, Hai Zhao, George Karypis, and Alex Smola. Multimodal chain-of-thought reasoning in language models, 2023.
- [13] Chengrun Yang, Xuezhi Wang, Yifeng Lu, Hanxiao Liu, Quoc V. Le, Denny Zhou, and Xinyun Chen. Large language models as optimizers, 2023.
- [14] Askyourpdf. <https://askyourpdf.com/>.
- [15] Chatdoc. <https://chatdoc.com/>.
- [16] Chatpdf. <https://www.chatpdf.com/>.
- [17] Lightpdf. <https://lightpdf.com/chatdoc>.
- [18] Chatgpt 3.5. <https://chat.openai.com/?model=text-davinci-002-render-sha>.
- [19] Chatgpt 4. <https://chat.openai.com/?model=gpt-4>.
- [20] OpenAI. Gpt-4v(ision) system card, 2023.
- [21] Pandasai. <https://docs.pandas-ai.com/en/latest/>.
- [22] Iso/ts 15066:2016 robots and robotic devices - collaborative robots, 2016.

[23] DeepL. <https://www.deepl.com/translator/files>.