



Università
di Catania

DIPARTIMENTO DI MATEMATICA E INFORMATICA
CORSO DI LAUREA TRIENNALE IN INFORMATICA

Lorenzo Maria Alberto Paoria

Titolo

RELAZIONE PROGETTO FINALE

Relatore: Emiliano Tramontana
Chiar.mo Prof. ---

Correlatori:
Gabriella Verga
Alessandro Midolo

Anno Accademico 2023 – 2024

Indice

Indice.....	3
Introduzione.....	4
Capitolo 1 - Metodologia.	5
1.1 - Titolo.	5
1.1.1 - Titolo.	5
1.1.2 - Titolo.	5
1.2 - Titolo.	5
Capitolo 2 - Risultati.	6
2.1 - Post con codice.	6
2.1.1 - Esempio post con codice.....	7
2.2 - Post senza codice.	9
2.2.1 - Esempio post senza codice.	12
2.3 - Considerazioni sui risultati.	15
2.3.1 - Considerazioni risultati dei post con codice.	15
2.3.2 - Considerazioni risultati dei post senza codice.	16
2.3.3 - Considerazioni generali.....	16
Capitolo 3 - Conclusioni.....	18
Riferimenti.....	19

Introduzione

Questa tesi si propone di confrontare le risposte fornite dagli utenti di StackOverflow con quelle generate da sistemi di intelligenza artificiale avanzati, come ChatGPT, al fine di valutare la capacità di tali tecnologie nel fornire soluzioni tecniche di alta qualità. L'obiettivo principale è analizzare la pertinenza, l'accuratezza e l'efficacia delle risposte prodotte dall'AI rispetto a quelle fornite da esperti umani, per comprendere i limiti e le potenzialità delle moderne tecnologie di linguaggio naturale nel supportare la risoluzione di problemi complessi nel campo dell'informatica.

Nel seguito, verranno illustrati i contenuti dei capitoli della tesi.

- **Metodologia:** Verrà illustrato l'approccio adottato per ottenere i risultati, descrivendo in dettaglio le tecniche e gli strumenti utilizzati nel processo di analisi.
- **Risultati:** Saranno presentati i risultati ottenuti, arricchiti da esempi pratici che ne facilitano la comprensione e dimostrano l'applicazione concreta.
- **Conclusione:** I risultati saranno interpretati ed analizzati per estrapolarne il significato, fornendo una sintesi delle implicazioni emerse e delle possibili prospettive future.

Capitolo 1 - Metodologia.

Questo è il capitolo 1. Questa è una citazione [1]

1.1 - Titolo.

Questo è un sottocapitolo del capitolo 1

1.1.1 - Titolo.

Questo è un sotto-sotto capitolo del capitolo 1.

1.1.2 - Titolo.

Questo è un altro sotto sotto capitolo del capitolo 1

1.2 - Titolo.

Questo è un altro sottocapitolo del capitolo 1.

Capitolo 2 - Risultati.

Una volta analizzati tutti i post e completate le richieste all'IA, sono stati prodotti i risultati per i diversi tipi di catalogazione, principalmente suddivisi in:

- Post con codice.
- Post con un numero di caratteri inferiore e superiore a 700.
- Post con un carattere specifico basato sul punteggio TF-IDF.

I risultati sono stati illustrati attraverso grafici ed esempi per facilitarne l'interpretazione.

2.1 - Post con codice.

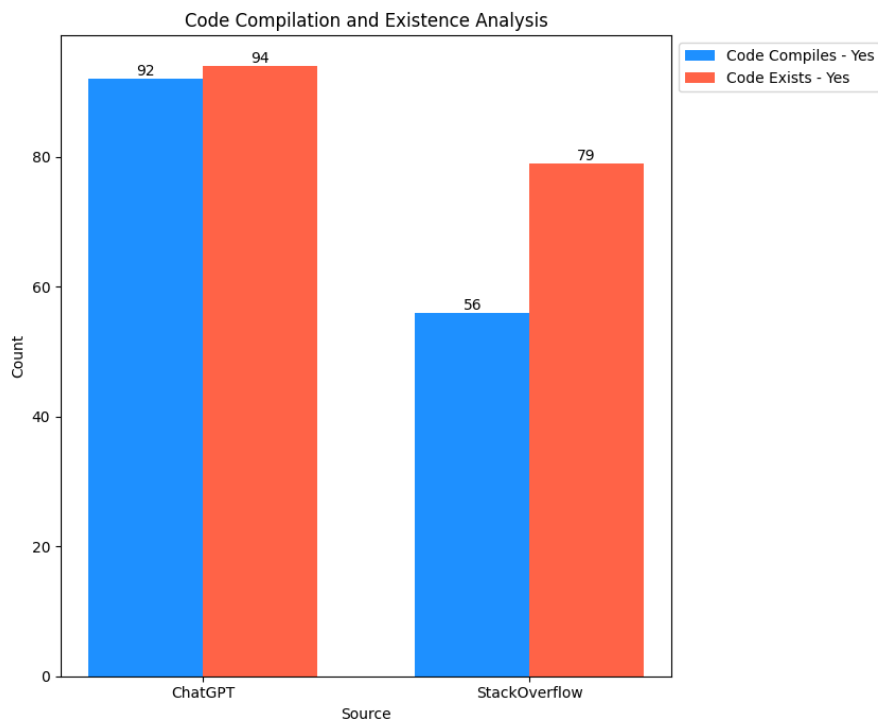
Per i post contenenti codice nelle risposte, è stata valutata l'eseguitività del codice o, nel caso di frammenti decontestualizzati, la correttezza semantica rispetto alla domanda. L'analisi è stata condotta su un dataset di 218 domande. Dai risultati è emerso che i post contenenti codice, già selezionati durante la creazione del dataset, erano:

- ChatGpt: 94 istanze
- StackOverflow: 79 istanze

Di queste:

- ChatGpt: 92 istanze compilano con successo
- StackOverflow: 56 istanze compilano con successo

Tabella 1- Grafico della situazione domande con codice.



Per fornire un esempio testuale dei post analizzati e dei risultati ottenuti, presenteremo un esempio di un post contenente codice e la relativa analisi.

2.1.1 - Esempio post con codice.

Domanda:

```
"ID": "937",  
"Question": "I know that they are implemented  
extremely unsafely in C/C++. Can't they be  
implemented in a safer way? Are the disadvantages  
of macros really bad enough to outweigh the massive  
power they provide? "
```

Risposta StackOverflow:

```
#define TWO 1+1

#define min(x, y) ((x) < (y)) ? (x) : (y))
```

Anche se questo codice è compilabile, può comportarsi in modo inaspettato quando viene utilizzato in espressioni. Pertanto, quando l'IA viene interpellata sulla sua correttezza semantica, lo considera non semanticamente corretto rispetto a quanto richiesto dalla domanda.

Risposta ChatGpt:

```
inline int square(int x) {

    return x * x;

}
```

ChatGPT sconsiglia l'uso delle macro in C++ e raccomanda invece di adottare un approccio basato su funzioni inline. Questo codice si compila correttamente ed è in linea con la domanda, poiché propone una soluzione alternativa ma più sicura rispetto a una che potrebbe comportare dei rischi.

2.2 - Post senza codice.

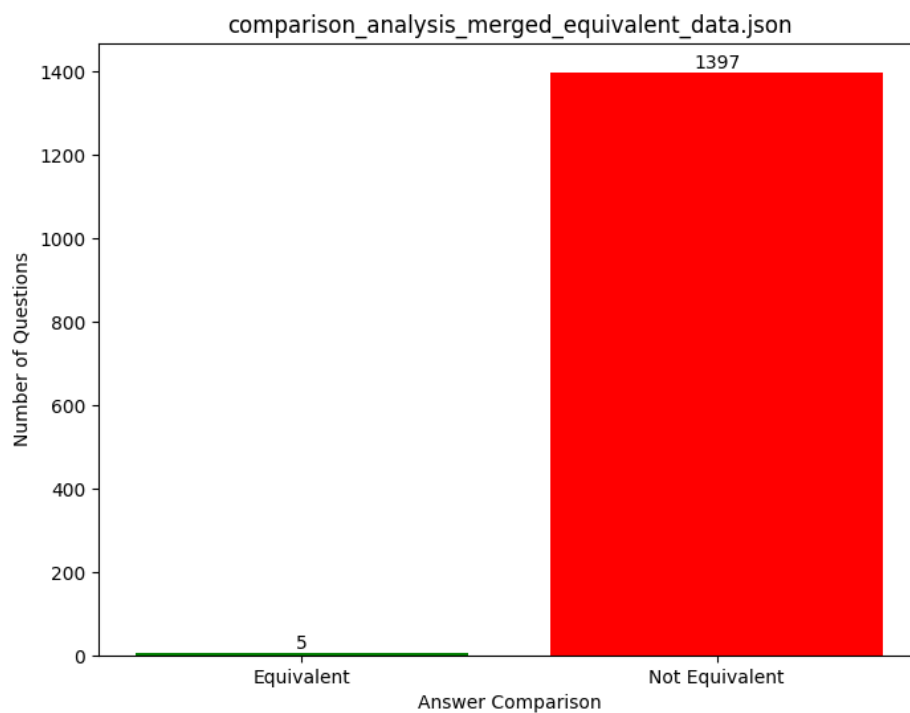
Per i post che non contengono codice, sono state raggruppate diverse categorie, tra cui:

- Post con un numero di caratteri superiore a limitChar.
- Post con un numero di caratteri inferiore a limitChar.
- Post contenenti un termine specifico basato sul punteggio TF-IDF.

Sono state analizzate le equivalenze tra le migliori risposte di StackOverflow e le risposte di ChatGPT per tutti questi tipi di post. L'analisi è stata condotta su un dataset di 1402 domande e dai risultati è emerso che:

- 1397 domande non avevano risposte equivalenti.
- 5 domande avevano risposte equivalenti.

Tabella 2- Grafico della situazione totale per tutte le domande senza codice.



Nel dettaglio possiamo vedere i risultati per singola catalogazione:

Tabella 3- Grafico della situazione per domande inferiori a 700 caratteri.

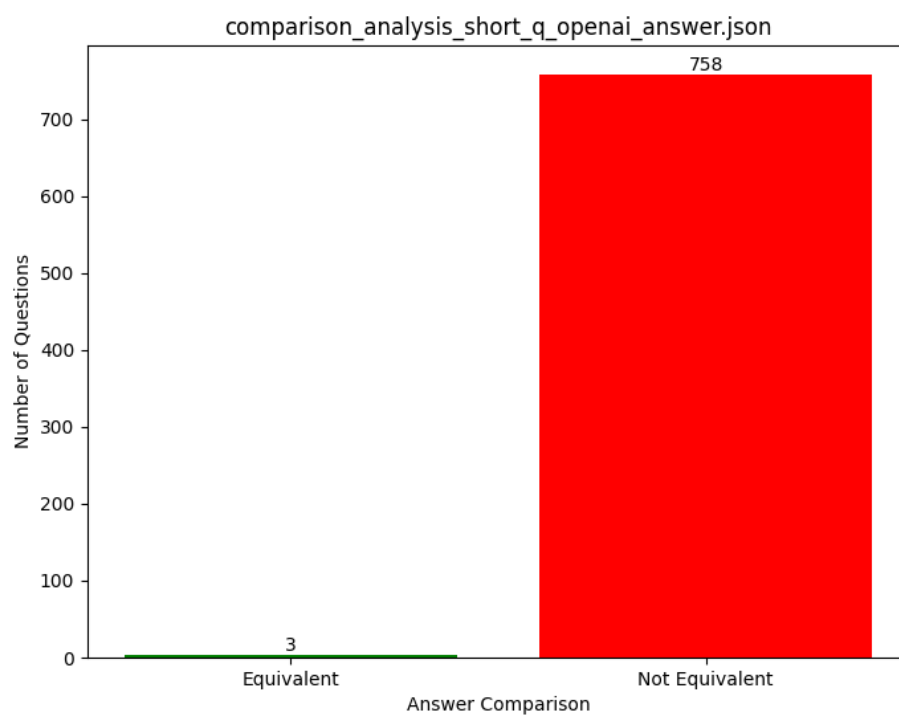


Tabella 4- Grafico della situazione per domande inferiori a 700 caratteri

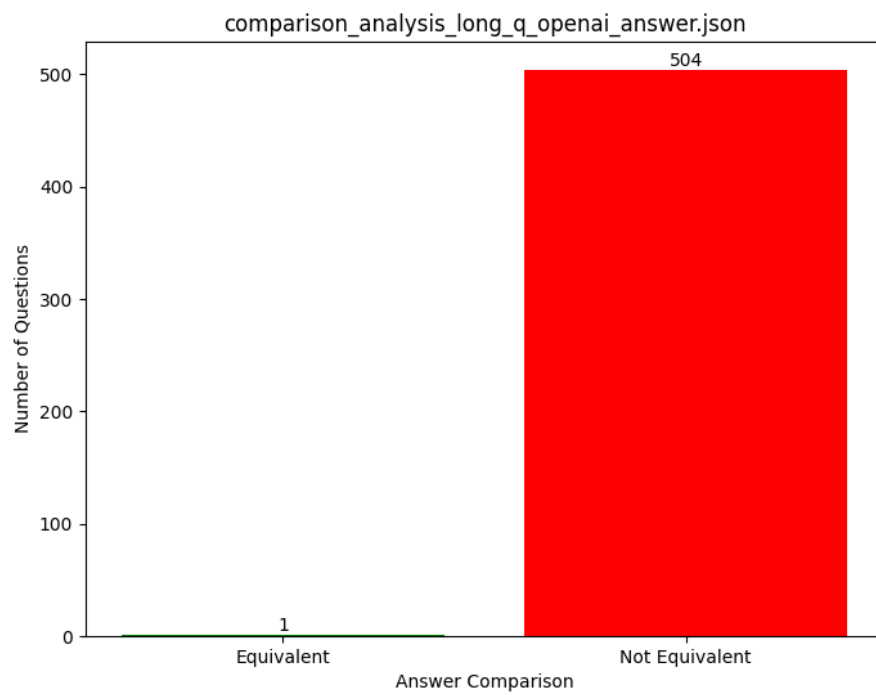
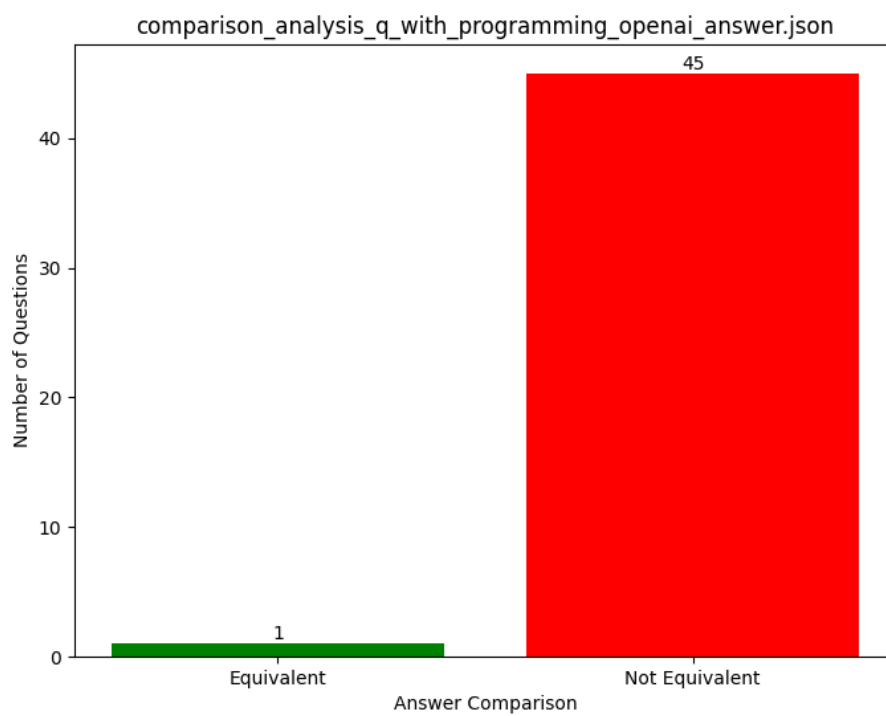


Tabella 5- Grafico della situazione per domande con termine basato su TF-IDF (PROGRAMMING)



2.2.1 - Esempio post senza codice.

Per illustrare il tipo di analisi effettuata e i risultati ottenuti, presenteremo un esempio che include una domanda con due risposte equivalenti.

Domanda:

```
"Question": "Where I work, employees use a third-party desktop program for their clients. This program saves data to a flat file. My colleague wants to write a Java program that uploads that flat file to a remote server, opens the desktop program when the flat file is downloaded from a Web site, and checks if the desktop program is running or not by looking at the Windows processes. He keeps calling this helper/utility program a \"wrapper.\" But it doesn't wrap anything! I tried to clear it up with him, but he said, \"Well, I call it a wrapper.\" He now has everyone in the company calling it a \"wrapper.\" What would you call it? I say that it's a helper program or utility program. "
```

Risposta di ChatGpt:

```
"ChatGpt answer": "You're correct in suggesting that the term \"wrapper\" may not be the most accurate description for the program your colleague wants to create. Typically, a \"wrapper\" refers to software that encapsulates or interfaces with another underlying program or library,
```

often providing a simpler or enhanced interface. In the case you've described, the Java program doesn't seem to be wrapping another piece of software but rather performing a variety of tasks\u2014uploading files, launching an application, and checking processes. Given its functionality, you could accurately describe it as a `**\"utility program\"**` or a `**\"helper program.\"**` These terms more clearly convey the program\u2019s role as a tool that assists users in performing specific tasks instead of encapsulating or modifying another program. You might also consider terms like `**\"automation script\"**` or `**\"file management tool,\"**` depending on how your colleague envisions the program functioning and its primary purpose. Whatever you decide to call it, highlighting its utility may help clarify its purpose within the company."

Risposta di StackOverflow:

"Stack Overflow best answer": "I don't think there's an official definition! In my mind, a wrapper program is a program that, in general, starts before the wrapped program starts, and exits no earlier than after the wrapped program exits. A wrapper program would probably also manage the lifetime of the wrapped program (start it for you, stop it for you.)"

```
I recommend against arguing too much over the names  
of things, though. The important thing is to  
design software that works well. Sometimes  
belaboring terminology can aid in this goal;  
sometimes it is a hobgoblin. "
```

La risposta di ChatGPT è più dettagliata e orientata a trovare una terminologia più precisa per descrivere il programma specifico, mentre la risposta di Stack Overflow è più generale e focalizzata sull'aspetto funzionale del software. Le due risposte non sono equivalenti poiché adottano approcci e prospettive differenti.

2.3 - Considerazioni sui risultati.

L'analisi dei dati raccolti ha fornito un quadro chiaro delle differenze tra le risposte fornite da StackOverflow e quelle generate da ChatGPT, evidenziando per ogni fonte punti di debolezza e forza.

2.3.1 - Considerazioni risultati dei post con codice.

L'analisi dei post contenenti codice ha mostrato che ChatGPT è in grado di generare codice compilabile con un'alta percentuale di successo, raggiungendo un tasso del 97,87% (92 su 94 istanze). Questo risultato è notevole, considerando che il codice è generato in modo automatico e senza un contesto approfondito rispetto alla domanda. Tuttavia, una critica è necessaria sulla correttezza semantica del codice prodotto. Sebbene il codice di ChatGPT tenda a essere sintatticamente corretto, la sua validità semantica può variare. In alcune istanze, il codice potrebbe compiere operazioni tecnicamente corrette ma non rispondere in modo preciso alla richiesta dell'utente, potenzialmente introducendo bug o comportamenti inattesi. Al contrario, le risposte di StackOverflow, pur avendo un tasso di compilazione inferiore pari al 70,88% (56 istanze su 79), beneficiano del contributo diretto di programmatori esperti che spesso tengono conto del contesto specifico, dei requisiti funzionali e delle best practice di programmazione, producendo risposte più aderenti alle esigenze degli utenti.

Inoltre, i post di StackOverflow tendono a includere spiegazioni dettagliate e discussioni sui pro e contro delle diverse soluzioni, offrendo agli utenti non solo una risposta immediata, ma anche una comprensione più profonda del problema e delle possibili soluzioni. Questo tipo di interazione è cruciale per l'apprendimento e per l'adozione di pratiche di programmazione migliori.

2.3.2 - Considerazioni risultati dei post senza codice.

L'analisi dei post senza codice ha rivelato una differenza significativa tra le risposte generate da ChatGPT e quelle presenti su StackOverflow. Su 1402 domande analizzate, solo 5 hanno mostrato risposte equivalenti, evidenziando una scarsa sovrapposizione tra le due fonti. Questo indica che le risposte di ChatGPT tendono a essere più generiche e meno mirate ai dettagli specifici della domanda, mentre quelle di StackOverflow sono spesso più focalizzate allo scopo ultimo della domanda.

Un aspetto cruciale è lo stile delle risposte di ChatGPT, che tende a essere uniforme, con una struttura discorsiva che abbraccia vari aspetti della domanda. Questo approccio è utile per fornire una panoramica generale, ma può risultare meno efficace quando è necessaria una risposta più precisa e focalizzata. Al contrario, le risposte su StackOverflow sono generalmente più concise e dirette, concentrate sull'aspetto pratico della soluzione. Grazie alla natura collaborativa della piattaforma, le risposte vengono affinate attraverso un processo iterativo di feedback da parte della comunità, portando alla costruzione di soluzioni concrete e ben definite. Questo stile si dimostra particolarmente efficace nel fornire risposte accurate e pratiche, arricchite dall'esperienza diretta di utenti che hanno affrontato problemi simili, garantendo così una maggiore pertinenza e utilità.

2.3.3 - Considerazioni generali

I risultati quindi suggeriscono che, mentre ChatGPT è uno strumento potente per generare rapidamente risposte sintatticamente corrette e fornire spiegazioni articolate, l'affidabilità e l'utilità delle sue risposte possono essere limitate dalla mancanza di contesto e di esperienza pratica che caratterizza invece le risposte degli utenti di StackOverflow. La collaborazione tra l'IA e gli esperti umani potrebbe rappresentare un'area di grande potenziale, combinando la velocità e

l'accessibilità di ChatGPT con la l'affidabilità delle risposte fornite da una comunità di esperti.

Per i post con codice, ChatGPT può servire come punto di partenza rapido per generare soluzioni o per ottenere spiegazioni tecniche di base, mentre StackOverflow può essere utilizzato per approfondire la discussione, valutare diverse soluzioni e scegliere l'approccio più adatto al problema specifico. Per i post senza codice, l'utilizzo di ChatGPT può essere utile per ottenere una risposta iniziale o una comprensione generale del problema, mentre StackOverflow rimane una risorsa indispensabile per ricevere consigli pratici e basati sull'esperienza.

In conclusione, mentre l'intelligenza artificiale ha dimostrato potenzialità promettenti, la sua integrazione con il contributo umano rimane essenziale per garantire soluzioni tecniche di alta qualità e pertinenti alle esigenze degli utenti. L'interazione sinergica tra questi strumenti può migliorare notevolmente l'esperienza degli utenti nella risoluzione di problemi complessi nel campo dell'informatica.

Capitolo 3 - Conclusioni

Riferimenti