



UNIVERSITÀ
DEGLI STUDI
DI TERAMO



Pratiche standard per la generazione di testo

Prof. ssa Romina Eramo
Università degli Studi di Teramo
Dipartimento di Scienze della Comunicazione
rerao@unite.it

Pratiche comuni utilizzate per la generazione di testo

- » Formati di testo
- » Spiegalo come se avessi cinque anni
- » Traduzione universale
- » Chiedere il contesto
- » Unbundling (Separazione dello stile di testo)
- » Riepilogo
- » Chunking (Suddivisione in blocchi del testo)
- » Analisi del sentiment
- » Least to Most (Dal più piccolo al più grande)
- » Richiesta di ruolo
- » Altre tattiche di prompting
- » Classificazione
- » Meta prompting

Formati di testo

- » Tra i formati generabili, gli elenchi semplici vanno bene per la maggior parte delle attività
 - Tuttavia, sono meno strutturati e per alcune attività è vantaggioso ottenere altre strutture di dati
- » Strutture di dati tipiche includono:
 - Dati di testo nidificati (elenchi gerarchici)
 - JSON (formato per lo scambio dati)
 - YAML (formato per la serializzazione di dati)
 - CSV (valori separati da virgola)
 - Mermaid (linguaggio per creare diagrammi)

Formati di testo:

Generazione di elenchi

Input: |

Genera un elenco di personaggi Disney.

Output:

Certo, ecco un elenco di alcuni personaggi Disney popolari:

1. Mickey Mouse

2. Minnie Mouse

...

30. Bagheera (Il libro della giungla)

Formati di testo: Generazione di elenchi

Prompt ottimizzato

Input:

```
Genera un elenco puntato di 5 personaggi Disney maschili.  
Includi solo il nome del carattere per ogni riga.  
Non includere mai il film per ogni personaggio Disney.  
Restituisci solo i personaggi Disney, non includere mai alcun  
commento.
```

```
Di seguito è riportato un elenco di esempio:
```

- * Aladdin
- * Simba
- * Bestia
- * Hercules
- * Tarzan

Output:

- * Woody
- * Buzz Lightyear
- * Stitch
- * Jack Sparrow
- * Principe azzurro

La semplice riformulazione del prompt per *includere esempi* (prompt a pochi colpi) può avere un grande impatto sull'output desiderato.

Formati di testo:

Generazione di elenchi gerarchici

Input:

```
Genera uno schema gerarchico e incredibilmente dettagliato
dell'articolo su:
Quali sono i vantaggi dell'ingegneria dei dati.
Di seguito è riportato un esempio della struttura gerarchica:
Titolo dell'articolo: Quali sono i vantaggi del marketing digitale?
*Introduzione
  a. Spiegazione del marketing digitale
  b. L'importanza del marketing digitale nel mondo degli affari di
    oggi
* Aumento della consapevolezza del marchio
  a. Definizione della brand awareness
  b. In che modo il marketing digitale aiuta ad aumentare la
    consapevolezza del marchio
```

Output:

```
Titolo dell'articolo: Quali sono i vantaggi dell'ingegneria dei dati?
*Introduzione
  a. Spiegazione dell'ingegneria dei dati
  b. L'importanza dell'ingegneria dei dati nel mondo odierno basato
    sui dati

... (10 sezioni dopo)...

*Conclusione
  a. L'importanza dell'ingegneria dei dati nel mondo degli affari
    moderno
  b. Il futuro dell'ingegneria dei dati e il suo impatto
    sull'ecosistema dei dati
```

Formati di testo:

Generazione di grafici Mermaid

Mermaid è un linguaggio di scripting simile a markdown per la generazione di grafici, diagrammi e altre visualizzazioni.

Input:

```
Crea un diagramma di flusso con sintassi Mermaid che mostri una
semplice ordinazione di cibo, con un processo che inizia con "Scegli
il cibo" e termina con "Paga il pasto".
Includi i passaggi per "Aggiungi cibo al carrello" e "Conferma
carrello".
```

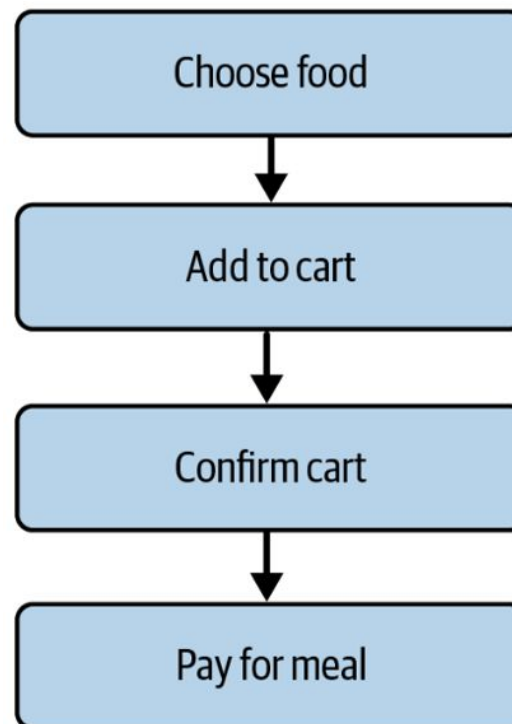
Output:

```
grafico TD
  ChooseFood[Scegli Cibo] --> AddToCart[Aggiungi al Carrello]
  AddToCart --> ConfirmCart[Conferma Carrello]
  ConfirmCart --> PayForMeal[Paga per il pasto]
```

Formati di testo:

Generazione di grafici Mermaid

L'output precedente, quando viene inserito in un renderer di diagrammi Mermaid, creerà questo diagramma di flusso:



Formati di testo:

Generazione di dati CSV

L'esempio generato è un file CSV ben formattato, che può essere caricato direttamente in molti strumenti che leggono i dati CSV

Input:

```
Genera un esempio di dati CSV per cinque studenti con i seguenti campi:  
nome, età, grado.
```

Output:

```
nome,età,grado|  
Giovanni,16,A  
Alex,17,B  
Emma,16,A  
Sofia,16,B  
Liam,17,A
```

Spiegalo come se avessi cinque anni

Input:

Spiega il seguente testo come se avessi 5 anni.

Testo: Questa è una recensione riguardante diversi tipi di trattamenti contro il cancro. Abbiamo mirato ad analizzare il microambiente tumorale e le tendenze recenti per le applicazioni terapeutiche e l'efficacia per diversi tipi dei tumori. Tradizionalmente il trattamento del cancro si basava sul cellule neoplastiche. ... (abbreviato per brevità)...

Output:

Il cancro è una malattia che fa crescere e dividere le cellule del tuo corpo, che in un certo senso non è normale. I medici hanno diversi modi per trattare il cancro, come tagliandolo fuori, dandoti medicine o usando raggi speciali per uccidere le cellule tumorali...

Traduzione universale

Input:

Crea del testo semplice.

Output:

Il sole sorge dolcemente, dipingendo il cielo con sfumature di arancione e rosa. Uccelli cinguettano allegramente mentre danno il benvenuto al nuovo giorno. La brezza fresca sussurra attraverso gli alberi, che portano il profumo dell'erba fresca e dei fiori che sbocciano. Vita si risveglia con un senso di tranquillità e promessa. ... (abbreviato per brevità)...

Traduzione universale

Input:

Ora rendi il testo incredibilmente complesso e difficile da leggere.

Output:

Nel regno etero del'asces celeste, in cui il globo splendente del sole inizia la sua graduale asces, si dispiega un fenomeno squisito, generando una sinfonia di brillantezza cromatica sull'ampio baldacchino del firmamento. Gli abitanti degli uccelli, dotati di voci melliflue, pronunciano i loro saluti giubilanti, punteggiando l'arazzo sonoro con note di felicità. Lo zefiro, che si libra teneramente tra le sentinelle arboree, trasporta sulle sue dolci correnti il profumo delle lame verdeggianti e della flora in fiore.

...

Traduzione universale

Input:

Ora trasforma il testo in un formato semplice e di facile lettura in spagnolo, assicurati che sia molto facile da capire e leggere. È necessario assicurarsi che sia il più semplice possibile da comprendere.

- Usa un linguaggio semplice.
- Rendilo conciso.

Output:

En un lugar mágico y celestial, cuando el sol empieza a salir, ocurre algo hermoso en el cielo. Los pájaros cantan alegremente, llenando el aire con sus dulces melodías. Una brisa suave sopla entre los árboles, llevando consigo el olor fresco de la naturaleza. Todo cobra vida y promete momentos felices. Nos invita a disfrutar de las cosas simples y a encontrar la felicidad en los pequeños detalles. En este escenario encantador, encontramos paz y aprendemos que hasta los momentos más simples pueden ser mágicos.

Traduzione universale

Input:

Infine, |traducilo di nuovo in italiano.

Output:

In un luogo magico e celestiale, quando il sole inizia a sorgere, qualcosa di bello accade nel cielo. Gli uccelli cantano allegramente, riempiendo l'aria con le loro dolci melodie. Una leggera brezza soffia tra gli alberi, portando il fresco profumo della natura. Tutto prende vita e promette momenti felici. Ci invita a godere delle cose semplici e a trovare la felicità nei piccoli dettagli. In questo incantevole scenario, troviamo la pace e impariamo ch anche i momenti più semplici possono essere magici.

Chiedere il contesto

E' possibile scrivere un prompt in cui si chiede al modello linguistico di:

1. Restituire un risultato valido a una domanda o a un'affermazione
2. Chiedi più contesto per rispondere in modo appropriato alla domanda

Input:

`Dovrei usare MongoDB o PostgreSQL per il mio progetto?`

Output:

`È difficile fare una raccomandazione senza saperne di più sul tuo requisiti software specifici. Sia MongoDB che PostgreSQL hanno il loro vantaggi, ma la scelta migliore dipende dalle esigenze specifiche del vostro progetto.`

Chiedere il contesto

Input:

Dovrei usare MongoDB o PostgreSQL per il mio progetto?
Fornisci un elenco di requisiti che ti aiuterebbero a fare un e una decisione di alta qualità su quale soluzione di database dovrei scegliere?

Output:

Per prendere una decisione informata tra MongoDB e PostgreSQL per il tuo progetto, considerare i seguenti requisiti:

1. Struttura dei dati: valuta la struttura dei tuoi dati. Se i tuoi dati è altamente destrutturato o privo di schema, il modello di documento flessibile di MongoDB potrebbe essere più adatto. D'altra parte, se i dati hanno uno schema fisso o richiede relazioni complesse, il modello relazionale di PostgreSQL potrebbe essere più adatti.

...

10. Costo: valutare le implicazioni in termini di costi di ciascuna soluzione di database. MongoDB offre sia un'edizione community (gratuita) che un'edizione commerciale con Funzionalità e supporto aggiuntivi. PostgreSQL è open-source e gratuito, con supporto commerciale opzionale disponibile da vari fornitori.

Input:

Dovrei usare MongoDB o PostgreSQL per il mio progetto?

Contesto:

- **Struttura dei dati:** dati altamente non strutturati con diversi formati di documento e nessuno schema fisso. **Scalabilità:** il progetto richiede la capacità di gestire enormi quantità di dati e distribuirli su più server.
- **Conformità ACID:** una forte coerenza e integrità dei dati sono fondamentali per le operazioni finanziarie del progetto.
- **Funzionalità di interrogazione:** Query e aggregazioni geospaziali complesse su dati basati sulla posizione.
- **Comunità ed ecosistema:** comunità attiva con numerose librerie terze parti e strumenti specificamente adattati per l'industria del progetto.
- **Integrità dei dati e vincoli:** applicazione di una chiave esterna rigorosa relazioni e regole complesse di convalida dei dati.
- **Transazioni e concorrenza:** richiede la gestione di alti livelli di operazioni di lettura e scrittura simultanee mantenendo la coerenza dei dati.
- **Maturità e stabilità:** dare priorità a un database consolidato e ampiamente testato per la stabilità a lungo termine.
- **Sviluppo e facilità operativa:** approccio di sviluppo agile con un'esigenza per la prototipazione rapida e la progettazione iterativa di schemi.
- **Costo:** i vincoli di budget richiedono l'utilizzo di una soluzione open source con nessun costo aggiuntivo di licenza o supporto.

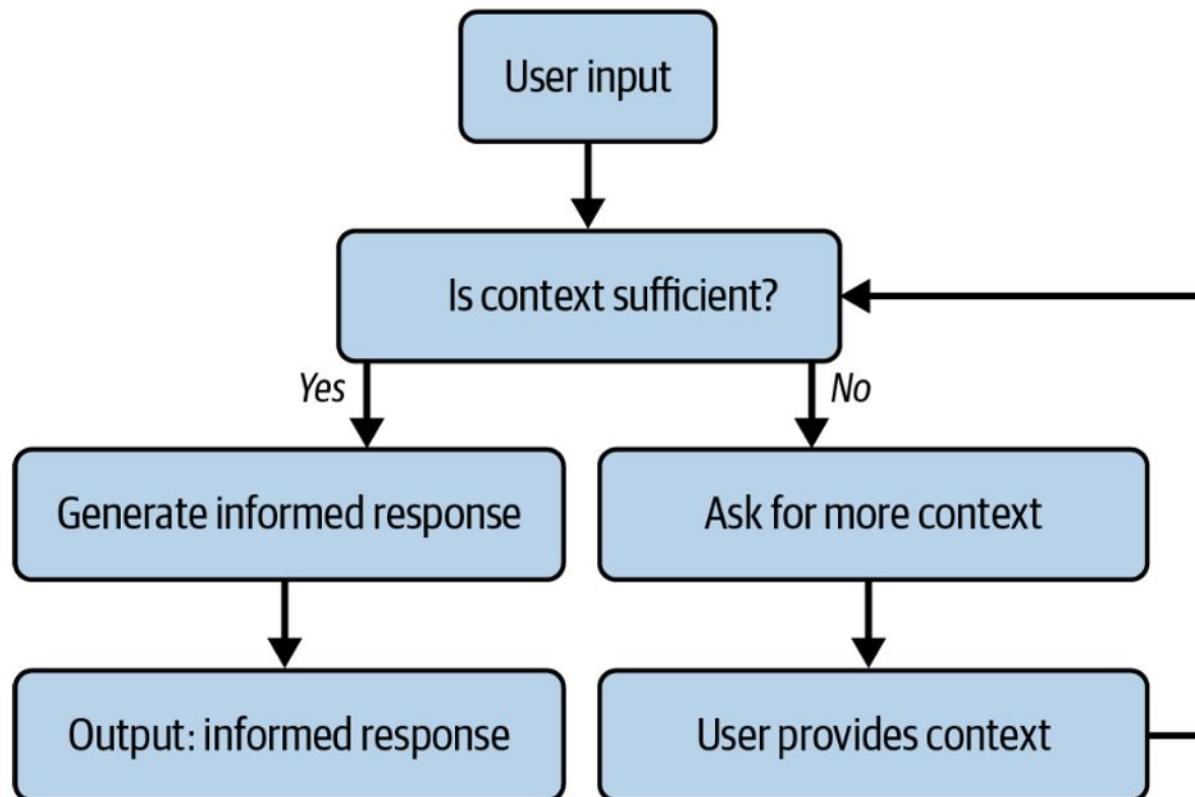
Se hai bisogno di più contesto, specifica cosa ti aiuterebbe a fare una decisione migliore.

Output:

Sulla base di queste considerazioni, sia MongoDB che PostgreSQL hanno il loro punti di forza, ma data l'enfasi posta su una forte coerenza dei dati, query e vincoli di integrità dei dati, PostgreSQL sembra essere un scelta adatta al tuo progetto.

Chiedere il contesto

- » Il processo decisionale di un LLM mentre si chiede il contesto



Separazione dello stile di testo (Unbundling)

- » Potente tecnica di prompt engineering che consente di estrarre e isolare caratteristiche testuali specifiche da un determinato documento
 - es., tono, lunghezza, vocabolario e struttura.
- » Ciò consente di creare nuovi contenuti che condividono caratteristiche simili al documento originale, garantendo coerenza nello stile e nel tono nelle varie forme di comunicazione.
- » I vantaggi di questa tecnica includono:
 - Migliore coerenza del marchio
 - Creazione semplificata di contenuti
 - Adattabilità

Separazione dello stile di testo

Identificazione delle caratteristiche testuali desiderate

- » Per separare correttamente uno stile di testo, è necessario prima identificare le funzioni specifiche che si desidera estrarre dal documento di input.
- » Le caratteristiche testuali comuni da considerare includono:
 - Tono di voce
 - » L'umore o l'atteggiamento generale trasmesso dal testo, ad esempio formale, informale, umoristico o autorevole
 - Lunghezza
 - » Il numero di parole desiderato o la lunghezza generale del contenuto
 - Vocabolario e fraseggio
 - » La scelta di parole ed espressioni, compreso il gergo specifico del settore, i colloquialismi e la complessità

Input:

Analizza il seguente testo e identifica le funzionalità principali che consentiranno ulteriori documenti per imitare lo stesso stile, coerenza e formattazione come nel documento sottostante.

Fornisci tutte le funzionalità come guida alla scrittura di stile che verrà utilizzata per i documenti futuri.

Analizza il seguente testo:

L'intelligenza artificiale sta trasformando rapidamente il mondo che ci circonda, aprendo nuove possibilità e rivoluzionando le industrie. Dall'assistenza sanitaria all'intrattenimento, l'intelligenza artificiale sta rendendo le nostre vite migliori e più comode in innumerevoli modi. Tuttavia, da un grande potere derivano grandi responsabilità. Dobbiamo garantire che le tecnologie di IA siano sviluppate e in modo etico e responsabile per proteggere la nostra privacy e sostenere la nostra valori.

Output:

Guida alla scrittura di stile:

Tono di voce

- Mantieni un tono informativo e ottimista in tutto il documento.
- Enfatizzare l'impatto positivo dell'argomento trattato pur riconoscendo eventuali dubbi o responsabilità.

Length:

...

Vocabolario:

...

Struttura:

...

Contenuto:

- Concentrarsi sulla natura trasformativa e sugli impatti positivi dell'argomento materia.
- Evidenziare settori o aree specifiche in cui l'argomento è stato ha fatto la differenza.
- Riconoscere eventuali responsabilità associate o potenziali rischi, sottolineando la necessità di pratiche etiche e responsabili. Sottolineare l'importanza della tutela della vita privata e del rispetto dei valori in relazione con l'oggetto.

Separazione dello stile di testo

Generazione di nuovi contenuti con le funzioni estratte

- » Dopo aver estratto le funzioni desiderate, è possibile utilizzare queste informazioni per guidare la generazione di contenuti futuri.
- » Per esempio:

```
Scrivi un nuovo post sul blog su [argomento]
utilizzando lo stesso tono di voce, lunghezza,
vocabolario e struttura del testo analizzato in
precedenza.
```
- » Combinando questa tecnica con il *testo di riferimento* (documenti che fungono da fondamento), è possibile produrre contenuti credibili che richiedono revisioni minime.

Separazione dello stile di testo

Estrazione di funzioni testuali specifiche con gli LLM

Input:

Analizza il testo seguente per identificare ed elencare le entità menzionate:

L'intelligenza artificiale sta trasformando rapidamente il mondo che ci circonda, aprendo nuove possibilità e rivoluzionando le industrie. Da dall'assistenza sanitaria all'intrattenimento, l'intelligenza artificiale sta rendendo le nostre vite migliori e più comodo in innumerevoli modi. Tuttavia, da un grande potere derivano grandi responsabilità. Dobbiamo garantire che le tecnologie di intelligenza artificiale siano sviluppate e impiegato in modo etico e responsabile per proteggere la nostra privacy e sostenere i nostri valori.

Output:

Entità:

- Intelligenza artificiale (AI)
- Assistenza sanitaria
- Divertimento
- Privacy
- Valori

Riassumere

Input:

Riassumi il seguente testo:

considerando che la parte della prima parte, in appresso denominata "Acquirente" e la parte della seconda parte, di seguito in qualità di "Venditore", essendo pienamente consapevole delle rispettive obblighe e responsabilità ai sensi della legge applicabile disposizioni e dottrine di common law, abbiano debitamente stipulato un accordo vincolante, di seguito denominato "accordo", per la al fine di effettuare un'operazione che comporta il trasferimento di beni, di seguito denominati "Beni", tra l'Acquirente e il Venditore, soggetti ai termini e alle condizioni qui stabiliti, le Parti Con la presente riconoscono e affermiamo solennemente quanto segue:

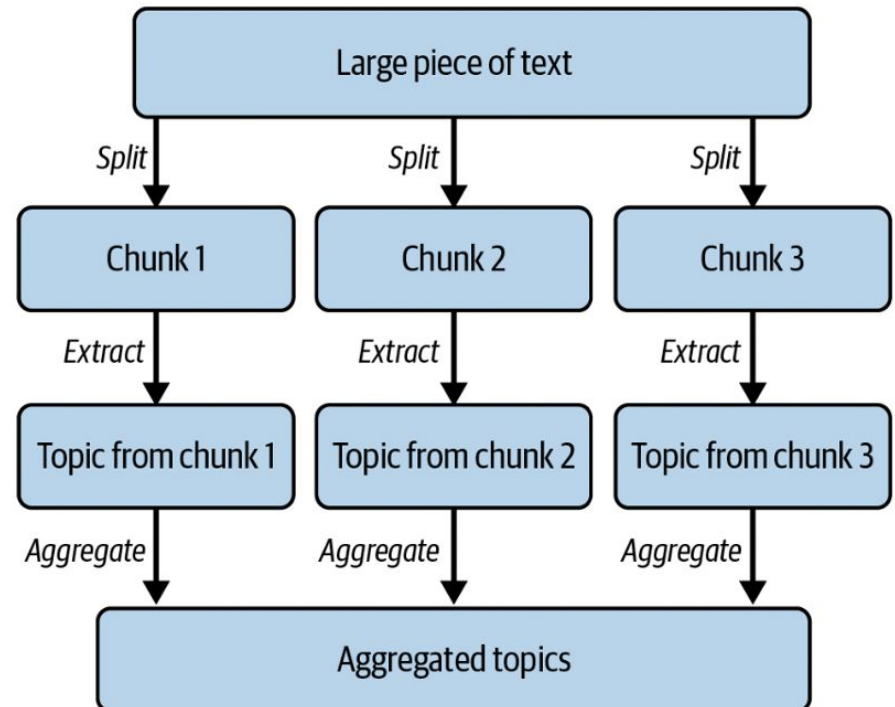
...

Output:

Il presente è un accordo legalmente vincolante tra l'acquirente e il venditore per il trasferimento di merci. L'accordo incorpora i considerando come parte integrante, ma in caso di conflitto, le disposizioni operative del prevale l'accordo. La clausola interpretativa afferma che le parole singolari includono il plurale e viceversa, e le parole che si riferiscono a un genere includono tutti i generi. I riferimenti alle disposizioni di legge comprendono eventuali modifiche o ostituzioni.

Suddivisione in blocchi del testo (Chunking)

- » Si riferisce al processo di suddivisione di grandi parti di testo in unità o blocchi più piccoli e più gestibili.
- » Questi blocchi possono essere basati su vari criteri, come frasi, paragrafi, argomenti, complessità o lunghezza.
- » Dividendo il testo in segmenti più piccoli, i modelli di intelligenza artificiale possono elaborare, analizzare e generare risposte in modo più efficiente.



Suddivisione in blocchi del testo

Quando suddividere in blocchi

- » **Documenti di grandi dimensioni:** Quando si ha a che fare con documenti estesi che superano il limite massimo di token dell'LLM
- » **Analisi complesse:** In scenari in cui è richiesta un'analisi dettagliata e il documento deve essere suddiviso per una migliore comprensione ed elaborazione
- » **Documenti multitopici:** Quando un documento copre più argomenti ed è utile gestirli singolarmente
- » *Quando non suddividere:*
 - Documenti brevi: Quando il documento è breve e ben entro i limiti del token dell'LLM
 - Analisi semplice: Nei casi in cui l'analisi o l'elaborazione richiesta è semplice e non trae vantaggio dalla suddivisione in blocchi
 - Documenti monotematici: Quando un documento è incentrato su un singolo argomento e la suddivisione in blocchi non aggiunge valore all'elaborazione

Analisi del sentiment

- » Tecnica di PNL (Natural language processing) ampiamente utilizzata che aiuta a identificare, estrarre e comprendere le emozioni, le opinioni o i sentimenti espressi in un testo.
- » L'obiettivo è determinare l'atteggiamento o il tono emotivo trasmesso in un testo, che sia positivo, negativo o neutro.
- » Queste informazioni possono fornire preziose informazioni sulle opinioni dei consumatori su prodotti o servizi, aiutare a monitorare la reputazione del marchio e persino aiutare a prevedere le tendenze del mercato.

Input:

Questo testo è positivo o negativo?

Adoro il design di questo telefono, ma la durata della batteria è abbastanza deludente.

Output:

Il testo ha un tono misto, in quanto contiene sia aspetti positivi che negativi. La parte positiva è "Adoro il design di questo telefono", mentre la parte negativa è "la durata della batteria è piuttosto deludente".

Analisi del sentiment

- » Il prompt può essere migliorato fornendo istruzioni chiare, esempi di offerte e definendo il formato di output

Input:

Utilizzando i seguenti esempi come guida:

Positivo: "Adoro il design di questo telefono!"

Negativo: "La durata della batteria è piuttosto deludente."

Neutro: "Mi è piaciuto il prodotto, ma ha una batteria di breve durata."

Restituisci solo una singola parola di:

- Positivo
- Negativo
- Neutrale

Si prega di classificare il sentiment del seguente testo come positivo, negativo, o neutro: Adoro il design di questo telefono, ma la durata della batteria è piuttosto deludente.

Output:

Neutrale

Analisi del sentiment

Tecniche per migliorare l'analisi del sentiment

- » Rimozione caratteri speciali
 - Emoji, hashtag e punteggiatura possono distorcere il giudizio dell' algoritmo di sentiment o non essere riconosciuti dai modelli
- » Conversione in lettere minuscole
 - I caratteri in minuscolo creano uniformità.
 - Es., Felice e felice vengono trattate in modo diverso dai modelli, causando duplicazioni e imprecisioni.
- » Correzione ortografica
 - Gli errori di ortografia possono causare interpretazioni errate e classificazioni errate

Analisi del sentiment

Limiti e sfide nell'analisi del sentiment

- » Gestire il sarcasmo e l'ironia
 - Un'interpretazione errata delle affermazioni sarcastiche o ironiche può portare a una classificazione imprecisa del sentiment.
- » Identificazione del sentiment specifico del contesto
 - Gli LLM possono avere difficoltà a classificare accuratamente i sentiment senza una guida adeguata o esempi specifici per il dominio.

Least to Most

(Dal più piccolo al più grande)

- » Chiedere al modello di risolvere un problema complesso suddividendolo in passaggi incrementali, partendo dalle domande più semplici per arrivare gradualmente a quelle più complesse.
- » Questo approccio aiuta il modello a "ragionare passo dopo passo" e migliorare l'accuratezza e la coerenza delle risposte.

Least to Most

(Dal più piccolo al più grande)

Problema: Pianificare un viaggio complesso

Obiettivo finale: Organizzare un viaggio di 10 giorni in Europa, con un budget limitato, che includa visite a 3 città principali, attività culturali e tempo libe

Input:

Pianifichiamo un viaggio in Europa utilizzando un approccio graduale. Risolvi il problema passo dopo passo seguendo queste domande:

1. Quali sono tre città principali in Europa che offrono una buona combinazione di cultura, intrattenimento e storia?
2. Qual è il modo più economico per spostarsi tra queste città? Considera opzioni come treni, voli low-cost o autobus.
3. Quali attrazioni principali sono imperdibili in ciascuna delle tre città? Elenca tre attrazioni per città.
4. Come possiamo suddividere il budget giornaliero per coprire alloggio, cibo e attività? Fornisci una stima per ogni categoria.
5. Infine, proponi un itinerario dettagliato di 10 giorni che includa gli spostamenti e le attività principali.

Rispondi a ogni punto uno alla volta e utilizza i risultati delle risposte precedenti per risolvere le domande successive.

Esecuzione Step-by-Step

1. Prima risposta:

- "Tre città principali che combinano cultura, intrattenimento e storia sono Parigi, Roma e Barcellona."

2. Seconda risposta:

- "Il modo più economico per spostarsi tra queste città è utilizzare voli low-cost per lunghi tragitti e treni regionali per distanze più brevi."

3. Terza risposta:

- Parigi: Torre Eiffel, Museo del Louvre, Cattedrale di Notre Dame.
- Roma: Colosseo, Vaticano, Fontana di Trevi.
- Barcellona: Sagrada Família, Parco Güell, Quartiere Gotico.

4. Quarta risposta:

- "Budget giornaliero stimato: 50€ per alloggio, 30€ per cibo, 20€ per attività, totale 100€ al giorno."

5. Quinta risposta:

- "Itinerario dettagliato: 3 giorni a Parigi (attrazioni principali e passeggiate nei quartieri storici), 4 giorni a Roma (visite culturali e tempo libero), 3 giorni a Barcellona (relax e visite). Spostamenti pianificati con voli economici tra Parigi-Roma e Roma-Barcellona."

Role Prompting (Richiesta di ruolo)

- » Il Role Prompting consiste nell'assegnare un ruolo specifico all'intelligenza artificiale per guidarne le risposte in modo mirato.
 - Questo approccio consente di simulare personalità, professioni o competenze per ottenere risposte personalizzate, accurate e contestualizzate.
- » Esempio: "Sei un medico esperto. Spiega a un paziente quali sono i sintomi dell'influenza."
- » [Awesome ChatGPT prompts](#) forniscono un ampio elenco di prompt dei ruoli da utilizzare.
 - <https://github.com/f/awesome-chatgpt-prompts/blob/main/prompts.csv>

Input:

Voglio che tu agisca come revisore tecnico. Ti darò il nome di un pezzo della tecnologia e mi fornirai una recensione approfondita - tra cui pro, contro, caratteristiche e confronti con altre tecnologie sul mercato. La mia prima richiesta di suggerimento è: MongoDB.

Output:

MongoDB è un popolare sistema di gestione di database NoSQL, orientato ai documenti, progettato per gestire l'archiviazione e il recupero di grandi volumi di dati.

Tratti somatici:

1. Archiviazione orientata ai documenti: MongoDB memorizza i dati in modo flessibile, simile a JSON, BSON, che semplifica l'archiviazione di strutture di dati complesse come matrici e oggetti nidificati.

...

Pro:

1. Prestazioni elevate: le funzionalità di indicizzazione e memorizzazione nella cache di MongoDB offrono operazioni di lettura e scrittura veloci.

...

Contro:

1. Supporto limitato per le transazioni: mentre MongoDB supporta più documenti ACID, potrebbe non essere la scelta migliore per le applicazioni che richiedono operazioni transazionali complesse.

...

Confronti con altre tecnologie:

1. Couchbase, un altro popolare database NoSQL, offre funzionalità simili a MongoDB, incluso lo schema flessibile e l'archiviazione basata su JSON. Tuttavia Couchbase è più focalizzato sul caching distribuito e potrebbe essere più adatto per applicazioni che richiedono l'accesso ai dati a bassa latenza.

Conclusione:

MongoDB è una soluzione di database NoSQL potente e versatile adatta per varie applicazioni, in particolare quelle che richiedono prestazioni elevate e scalabilità. Il suo modello di dati flessibile, il linguaggio di query avanzato e la forte coerenza lo rende una scelta popolare tra gli sviluppatori.

Role Prompting

Vantaggi del Role Prompting

- » Risultati più mirati:
 - L'IA risponde all'interno del contesto specifico del ruolo assegnato.
 - Migliora la pertinenza delle risposte.
- » Creatività potenziata:
 - Stimola l'IA a "pensare" da prospettive diverse, come un personaggio storico o un esperto di un settore.
- » Personalizzazione dello stile:
 - Le risposte possono essere formali, informali, umoristiche o adatte al contesto desiderato.
- » Esplorazione di prospettive diverse:
 - Aiuta a ottenere risposte da vari punti di vista, utili per analisi o brainstorming.
- » Maggiore coinvolgimento:
 - Le interazioni diventano più accattivanti e immersive.

Role Prompting

Sfide del Role Prompting

- » Pregiudizi o stereotipi:
 - Assegnare ruoli stereotipati può portare a risposte distorte o culturalmente insensibili.
 - Importanza di monitorare e progettare prompt eticamente.
- » Coerenza del ruolo:
 - Durante interazioni prolungate, il modello potrebbe deviare dal ruolo assegnato.
 - Richiede una supervisione costante per mantenere il focus.
- » Mantenimento della qualità:
 - Non sempre le risposte sono in linea con il ruolo, specialmente in contesti complessi.

Role Prompting

Quando Utilizzare il Role Prompting

- » Competenze specifiche: Per domande che richiedono conoscenze approfondite di un settore.
 - Esempio: "Assumi il ruolo di uno scienziato e spiega il processo di fotosintesi."
- » Personalizzazione dello stile di risposta: Adattare il tono al pubblico o al contesto.
 - Esempio: "Rispondi come farebbe un blogger di viaggi entusiasta."
- » Creatività e narrazione: Generare storie, scenari immaginari o dialoghi.
 - Esempio: "Sei un narratore. Descrivi una scena in una città futuristica."
- » Esplorazione di prospettive diverse: Analizzare punti di vista contrastanti su un tema.
 - Esempio: "Sei un politico e poi un attivista. Esprimi le tue opinioni sulla sostenibilità ambientale."
- » Migliorare il coinvolgimento: Creare interazioni più dinamiche, ad esempio, assumendo il ruolo di un personaggio fittizio.
 - Esempio: "Sei un detective in un caso misterioso. Guida l'utente nella risoluzione del crimine."

Evitare le allucinazioni con riferimento

- » Un'allucinazione si verifica quando un modello genera informazioni inesatte, irrilevanti o inventate, non supportate da dati reali.
- » Problema: Le allucinazioni possono compromettere la qualità e l'affidabilità delle risposte fornite dall'intelligenza artificiale.
- » Per ridurre le allucinazioni, si può istruire il modello a:
 - Rispondere esclusivamente utilizzando il contenuto di riferimento fornito.
 - Fornire informazioni pertinenti e chiare in modo che il modello resti nel contesto.

Evitare le allucinazioni con riferimento

Regole per implementare il metodo

- » Definire un contesto chiaro:
 - Racchiudere il testo di riferimento tra virgolette triple ("") per delimitare le informazioni di base.
- » Stabilire i principi guida:
 - Se la risposta non può essere trovata nel testo di riferimento, il modello deve restituire: "Non sono riuscito a trovare una risposta".
 - Formattare le risposte in modo coerente con il riferimento.

Input:

Fare riferimento agli articoli racchiusi tra virgolette triple per rispondere alle query.

È necessario seguire i seguenti principi:

- Nei casi in cui la risposta non viene trovata all'interno di questi articoli, è sufficiente restituire "Non sono riuscito a trovare una risposta".

"""

I clienti B2B tendono ad avere processi decisionali più lunghi e quindi canalizzazioni di vendita più lunghe. Le strategie di costruzione delle relazioni funzionano meglio per questi clienti, mentre i clienti B2C tendono a rispondere meglio alle offerte e ai messaggi a breve termine.

"""

Esempi di risposte:

- Non sono riuscito a trovare una risposta.

- Sì, i clienti B2B tendono ad avere processi decisionali più lunghi e quindi canalizzazioni di vendita più lunghe.

Output:

Sì, i clienti B2B tendono ad avere processi decisionali più lunghi, il che porta a cicli di vendita più lunghi.

Input:

... Il resto del prompt...

Domanda: Le vendite B2C sono più convenienti?

Output:

Non sono riuscito a trovare una risposta.

Evitare le allucinazioni con riferimento

Dare la direzione e specificare il formato

- » Un approccio avanzato prevede di includere citazioni dirette dal testo per autenticare le risposte.
 - Il modello può essere guidato a includere citazioni nelle sue risposte facendo riferimento a sezioni dei documenti forniti.
- » Questo approccio ha l'ulteriore vantaggio che le citazioni nell'output possono essere autenticate automaticamente abbinando le stringhe all'interno dei documenti dati.

Input:

Ti verrà fornito un documento delimitato da virgolette triple e una domanda. Il tuo compito è quello di rispondere alla domanda utilizzando solo il documento fornito e di citare il passaggio o i passaggi del documento utilizzati per rispondere alla domanda. Se il documento non contiene le informazioni necessarie per rispondere a questa domanda, scrivi semplicemente: "Informazioni insufficienti". Se viene fornita una risposta alla domanda, deve essere annotata con una citazione. Usa il seguente formato per citare i passaggi pertinenti ({"citation": ...}).

"""

Un reattore nucleare è un dispositivo che avvia, controlla e mantiene una reazione nucleare a catena. Il reattore nucleare utilizza l'energia rilasciata dalla fissione nucleare per generare calore, che viene poi spesso utilizzato per produrre vapore per azionare una turbina e produrre elettricità.

- Tratto da https://en.wikipedia.org/wiki/Nuclear_reactor

"""

Domanda: Che cos'è un reattore nucleare?

Output:

Un reattore nucleare è un dispositivo che avvia, controlla e mantiene una reazione nucleare a catena. Utilizza l'energia rilasciata dalla fissione nucleare per generare calore, che viene poi spesso utilizzato per produrre vapore per azionare una turbina e produrre elettricità. {"citation": "Un reattore nucleare è un dispositivo che avvia, controlla e mantiene una reazione nucleare a catena. Il reattore nucleare utilizza l'energia rilasciata dalla fissione nucleare per generare calore, che viene poi spesso utilizzato per produrre vapore per azionare una turbina e produrre elettricità."}

Dai “tempo per pensare”

- » Spesso, guidando esplicitamente un LLM a *ricavare soluzioni dai principi primi*, prima di raggiungere un verdetto, è possibile ottenere risposte più accurate.
- » Fornire a un LLM il tempo di riflessione può spesso portare a risultati migliori.

Input:

Innanzitutto, prova a risolvere il problema da solo. Quindi, vedi se la tua risposta corrisponde alla soluzione proposta. Non giudicare la risposta data finché non hai provato a risolverlo da solo.

Problema:

Jane ha 5 mele, poi John ne prende 2 da Mary. Mary originariamente ne aveva 15 mele. John poi dà le sue mele a Jane.

Quante mele ha Jane?

Soluzione proposta:

1. Jane originariamente aveva 5 mele.
2. John prende 2 mele da Mary, ma questo non influisce sul conteggio di Jane.
3. John poi dà le sue mele a Jane.

Quindi, il conteggio finale delle mele di Jane sono le sue 5 mele originali più le 2 mele che John le regalò, per un totale di 7 mele.

Output:

Analizziamo il problema:

1. Jane ha originariamente 5 mele.
2. John prende 2 mele da Mary. Tuttavia, questa azione non influisce direttamente sul numero di mele che Jane possiede.
3. John dà le sue mele a Jane.

Seguendo questi passaggi, dobbiamo aggiungere le 2 mele che John ha dato a Jane il numero originale di mele di Jane. Pertanto:

5 mele (il numero originale di Jane) + 2 mele (date da John) = 7 mele

Quindi, Jane ora ha 7 mele. La mia soluzione corrisponde alla soluzione proposta.

La tattica del monologo interiore

- » La tattica del monologo interno indica al modello di strutturare parti dell'output che devono essere nascoste all'utente in un formato specifico.
- » In questo modo è facile rimuovere queste parti prima di presentare l'output finale all'utente.

Input:

Passaggio 1: risolvere il problema

Innanzitutto, devi trovare la tua soluzione al problema. Assicurati di non farlo affidandoti alla soluzione dell'utente, in quanto potrebbe non essere corretta. È possibile incapsulare le tue lavorazioni tra virgolette triple ('''') per indicare che la sezione dovrebbe essere nascosta.

Passaggio 2: confrontare le soluzioni

Quindi, confronta la tua soluzione con la soluzione dell'utente. Valutare se la soluzione dell'utente è corretta o meno. Ancora una volta, incapsula questa sezione all'interno di virgolette triple ('''').

Passaggio 3: prepara un suggerimento

Se l'utente ha commesso un errore, determinare un suggerimento appropriato che possa guidarlo verso la risposta corretta, senza rivelarla a titolo definitivo. Questa sezione dovrebbe anch'essa essere racchiusa tra virgolette triple ('''').

Passaggio 4: consegna il suggerimento

Se è stato commesso un errore, presentare all'utente il suggerimento del passaggio precedente. Ricorda di non usare virgolette triple per questo passaggio. Invece di usare un titolo come "Passaggio 4", usa un'intestazione come "Suggerimento:"

Dichiarazione del problema: Un mazzo di carte ha 52 carte: 13 cuori, 13 quadri, 13 fiori e 13 picche. Se due carte vengono pescate dal mazzo senza sostituzione, qual è la probabilità che una delle carte pescate sia un cuore e l'altro un fiore?

Soluzione dell'utente: $(13/52) * (13/51) = 169/2652 = 13/204$

Output:

"""

Passaggio 1 - La mia soluzione:

Per prima cosa, calcola la probabilità di estrarre 1 cuore e 1 fiori da un mazzo di 52 carte senza ricambio.

Nella prima estrazione, la probabilità di estrarre un cuore è $13/52$.

Per la seconda estrazione, dopo aver rimosso 1 carta (supponendo che sia un cuore), il numero totale di carte diventa 51. Quindi, la probabilità di pescare un fiore ora è $13/51$.

Moltiplicando queste due probabilità si ottiene la probabilità di entrambi gli eventi che accade in sequenza, che è $P(\text{Cuore poi Club}) = (13/52) * (13/51)$.

Ripeti il calcolo supponendo che la prima carta estratta sia un fiori e che il La seconda carta pescata è un cuore. Ricorda, l'ordine conta in questo caso come Abbiamo a che fare con eventi separati.

"""

"""

Passaggio 2 - Confronto con la soluzione dell'utente:

Confrontare la soluzione con quella dell'utente. Solo la soluzione dell'utente

Considera la sequenza di estrazione prima di un cuore e poi di un bastone. Essi non hanno tenuto conto della sequenza alternativa (Fiori e Cuori), e pertanto, la soluzione è incompleta.

"""

"""

Passaggio 3 - Preparazione del suggerimento:

Dato che l'utente non ha considerato entrambe le sequenze (cuore poi fiore e fiore poi cuore), un accenno alla considerazione di entrambe le possibili sequenze sarà utile. Questo potrebbe spingere l'utente nel giusto senza rivelare la soluzione a titolo definitivo.

"""

Passaggio 4 - Consegna del suggerimento:

Suggerimento: Ricorda che le due carte possono essere pescate in qualsiasi ordine. Considerare entrambe le possibili sequenze: cuori poi fiori e fiori poi cuori.

Classificazione con LLM

- » La classificazione è il processo di previsione della classe o categoria di un punto dati basandosi su modelli appresi.
 - Esempio: Assegnare etichette come positivo o negativo a recensioni di prodotti.
- » È una delle attività centrali dell'apprendimento automatico, utilizzata in campi come l'elaborazione del linguaggio naturale, la visione artificiale e l'analisi dei dati.
- » Strategie di Classificazione
 - Apprendimento Zero-Shot
 - Apprendimento Few-Shot

Classificazione con LLM

Apprendimento Zero-Shot

- » Il modello classifica i dati senza alcun esempio fornito nel prompt.
 - Si fornisce al modello una domanda chiara e specifica con le possibili classi.
 - Il modello utilizza la sua conoscenza pre-addestrata per fornire una risposta.
- » Vantaggi:
 - Nessun dato di addestramento richiesto.
 - Ideale per prototipi o problemi semplici.
- » Limitazioni:
 - L'accuratezza può essere inferiore rispetto all'apprendimento con esempi.

Classificazione con LLM

Apprendimento Zero-Shot

Input:

Classifica il seguente messaggio come "spam" o "non spam":

"Congratulazioni! Hai vinto un premio. Clicca qui per reclamarlo."

Output:

Spam

Classificazione con LLM

Apprendimento Few-Shot

- » Il modello classifica i dati basandosi su un piccolo numero di esempi forniti nel prompt.
 - Fornisci esempi di input e output nel prompt.
 - Il modello utilizza questi esempi per inferire la struttura e il formato della risposta.
- » Vantaggi:
 - Migliora l'accuratezza rispetto al metodo zero-shot.
 - Utile per problemi più complessi o specifici.
- » Limitazioni:
 - Richiede una maggiore attenzione alla costruzione del prompt.

Classificazione con LLM

Apprendimento Zero-Shot

Input:

Classifica i seguenti messaggi come "spam" o "non spam":

- "Partecipa al nostro concorso e vinci premi straordinari!" -> spam
- "Ecco la tua ricevuta per l'acquisto recente." -> non spam
- "Congratulazioni! Hai vinto un premio. Clicca qui per reclamarlo." -> ?

Output:

Spam

Meta prompting

- » Il meta prompting è una tecnica che prevede la creazione di prompt di testo che, a loro volta, generano altri prompt di testo.
- » Questi prompt di testo vengono quindi utilizzati per generare nuove risorse in molti mezzi, ad esempio immagini, video e altro testo.
- » Perché è utile?
 - Ottimizzazione dei prompt
 - Automazione della progettazione del prompt
 - Flessibilità

Meta prompting

» Come funziona?

- Passaggio 1: Formulare un Meta Prompt per chiedere al modello di creare o modificare un prompt.
- Passaggio 2: Eseguire il prompt generato su sé stesso o su un altro modello.
- Passaggio 3: Valutare e, se necessario, iterare per migliorare l'efficacia.

» Applicazioni:

- Generazione di prompt specializzati
- Adattamento a contesti diversi
- Miglioramento iterativo

Scenario: Creare un prompt efficace per spiegare il concetto di apprendimento automatico a uno studente delle scuole superiori.

Meta Prompt:

Crea un prompt efficace per spiegare il concetto di apprendimento automatico a uno studente delle scuole superiori. Assicurati che il prompt sia semplice, chiaro e utilizzi esempi pratici per facilitare la comprensione.

Risultato Generato (dal modello):

Spiega il concetto di apprendimento automatico come se lo stessi insegnando a uno studente delle scuole superiori. Utilizza esempi pratici, come come i social media raccomandano contenuti o come gli assistenti vocali riconoscono le voci, per rendere il concetto semplice e accessibile.

Utilizzo del Prompt Generato: Una volta generato il prompt, può essere utilizzato per richiedere al modello di completare il compito.

Input:

Spiega il concetto di apprendimento automatico come se lo stessi insegnando a uno studente delle scuole superiori. Utilizza esempi pratici, come come i social media raccomandano contenuti o come gli assistenti vocali riconoscono le voci, per rendere il concetto semplice e accessibile.

Output del Modello:

L'apprendimento automatico è una tecnologia in cui i computer imparano a eseguire compiti senza essere programmati direttamente. Ad esempio, quando usi un social network, l'algoritmo impara dai tuoi "mi piace" per consigliarti post o video che ti potrebbero interessare. Oppure, quando parli a un assistente vocale come Alexa o Siri, il sistema riconosce la tua voce grazie a un modello che è stato addestrato con molti esempi di voci umane.