



AI

- POSTI DI LAVORO
- TOKENIZATION LIMIT
- CHE ORA È?

2024-10-16

OSSERVATORIO ANSSAIF INTELLIGENZA ARTIFICIALE

AGENDA

1. Come procede in Italia e all'estero la perdita dei posti di lavoro?
2. Tokenization Limit
3. Che ora è?
4. Si perderanno veramente tanti posti di lavoro?
5. Il cambiamento

COME PROCEDE LA PERDITA DEI POSTI DI LAVORO?

Il mercato del lavoro in Italia, a ottobre 2024, mostra un segnale positivo di crescita occupazionale. A giugno 2024, il numero di occupati è aumentato di 440.000 unità rispetto all'anno precedente, in particolare grazie ai contratti a tempo indeterminato, che sono cresciuti di 347.000 unità. Tuttavia, ci sono state diminuzioni nelle assunzioni con contratti a termine e una crescita nel lavoro occasionale. Anche il tasso di disoccupazione è in calo, con un'ulteriore riduzione registrata nel secondo trimestre del 2024. Il tasso di disoccupazione si è ridotto dello 0,5% rispetto al 2022, con un aumento complessivo del numero di occupati. Tuttavia, persistono alcune criticità, come il divario di genere nell'occupazione femminile, che resta sotto la media europea. In sintesi, il mercato del lavoro italiano continua a crescere, anche se ci sono settori con dinamiche diverse e sfide da affrontare.

Fonti: QUIFINANZA, FINANZA, REPUBBLICA.IT, PARTITAIVA.IT.
OSSERVATORIO ANSSAIF IA

COME PROCEDE LA PERDITA DEI POSTI DI LAVORO?

Negli Stati Uniti, il mercato del lavoro ha mostrato una crescita significativa di recente. A settembre, sono stati creati 254.000 nuovi posti di lavoro, superando le aspettative degli analisti¹. Il tasso di disoccupazione è sceso al 4,1%, indicando un mercato del lavoro forte e resiliente².

I settori che hanno visto la maggiore crescita includono la ristorazione e i bar, con 69.000 nuove posizioni, e l'assistenza sanitaria, con 45.000 nuovi posti di lavoro¹. Anche l'occupazione nel settore governativo e nell'edilizia ha registrato aumenti significativi¹.

*1 lavocedinyork.com
2 ansa.it*

TOKENIZATION LIMIT

Le "token windows" nei modelli di linguaggio di grandi dimensioni (Large Language Models o LLM) si riferiscono alla capacità del modello di elaborare un determinato numero di token (unità di testo) in una singola finestra di contesto. Un "token" può essere una parola, un carattere o un gruppo di caratteri, a seconda del metodo di tokenizzazione utilizzato.

TOKENIZATION LIMIT

1. Definizione di Token:

Nei modelli di linguaggio, i testi vengono suddivisi in unità più piccole chiamate token. Ad esempio, la frase "Il gatto corre veloce" potrebbe essere suddivisa in token come ["Il", "gatto", "corre", "veloce"].

2. Dimensione della Finestra (Token Window Size):

La dimensione della finestra dei token è il numero massimo di token che un dato modello può considerare contemporaneamente. Ad esempio, un modello con una finestra di 1024 token può prendere in considerazione fino a 1024 token alla volta.

TOKENIZATION LIMIT

3. **Contesto:**

La finestra dei token determina quanto contesto il modello può utilizzare per generare risposte o previsioni. Un modello con una finestra di 2048 token può gestire testi più lunghi e mantenere il contesto su un intervallo di token maggiore rispetto a un modello con una finestra di 512 token.

4. **Implicazioni sulle Prestazioni:**

Una finestra di token più grande permette al modello di avere una migliore comprensione del contesto e delle dipendenze a lungo raggio nel testo, migliorando la coerenza e la qualità delle risposte.

Tuttavia, elaborare una finestra più grande richiede anche più risorse computazionali (memoria e tempo di calcolo).

TOKENIZATION LIMIT

- **5. Limiti della finestra di contesto (input e output)**
 - **GPT-3 (davinci):** può gestire fino a **4.096 token**.
 - **GPT-4:**
 - Alcune versioni di GPT-4 hanno una finestra di contesto fino a **8.192 token**.
 - Versioni più avanzate di GPT-4 possono arrivare fino a **32.768 token**.
 - **BERT:** Solitamente può processare un massimo di **512 token**.
 - **LLaMA:** Dipende dalla versione, ma generalmente gestisce tra **2.048** e **4.096 token**.

TOKENIZATION LIMIT

- **6. Calcolo della dimensione dell'input e output**
 - **Input:** Il testo che si da in input viene tokenizzato, quindi bisogna fare attenzione alla lunghezza del testo. Ad esempio, un testo di 1.000 parole può risultare in circa 1.300-1.500 token (a seconda della lingua e della complessità del testo).
 - **Output:** Anche l'output del modello viene limitato dal numero di token. Se si ha una finestra di contesto di 4.096 token e si da un input di 3.500 token, il modello avrà solo 596 token disponibili per generare una risposta.

TOKENIZATION LIMIT

- **7. Strategie per gestire i limiti del contesto**
 - **Raggruppamento delle conversazioni:** Quando si crea una chatbot o un sistema di conversazione, una strategia comune è quella di tenere in memoria solo le ultime parti di una conversazione, eliminando o riducendo i messaggi più vecchi. Ad esempio, mantenere solo gli ultimi 500 o 1.000 token della conversazione.
 - **Riduzione dell'input:** Se l'input è molto lungo (es. documenti lunghi), è possibile usare tecniche come il **summarization** (riassunto automatico) per ridurre il contenuto prima di passarlo al modello.

TOKENIZATION LIMIT

- **Conclusione**

Le token windows nei modelli di linguaggio di grandi dimensioni sono una componente chiave che influisce direttamente sulla capacità del modello di gestire e comprendere testi complessi e lunghi. Una finestra più ampia permette una migliore gestione del contesto, ma richiede anche maggiori risorse computazionali.

CHE ORA È?

... È ora di non farsi fregare!

Le chat bot non sanno rispondere, ancora, ad un piccolo quesito come «che ora è?» per farlo le applicazioni AI nel web utilizzano delle tecniche apposite, per ovviare alla problematica.

LeChatMistral OGGI!

 Che ora è?

 Mi dispiace, ma non posso fornire l'ora corrente in tempo reale. Tuttavia, puoi facilmente controllare l'ora sul tuo dispositivo o su un orologio. Se hai altre domande o hai bisogno di assistenza su qualcos'altro, fammi sapere!

ChatGPT IERI!

Mi dici l'ora di oggi?

 Oggi è il 15 ottobre 2024, e l'ora attuale dipende dal tuo fuso orario. Puoi indicarmi la tua posizione, così posso dirti l'ora corretta, oppure puoi controllare direttamente sul tuo dispositivo.

CHE ORA È?

... È ora di non farsi fregare!

 Tu

Che ora è?



COPILLOT OGGI!

 Copilot

Sono le 16:35 qui a Roma. Hai bisogno di sapere l'ora per qualcosa di specifico? 😊



Che ora è?

GEMINI OGGI!

Mostra bozze ▾



In questo momento, sono le **04:39 PM CEST** del 16 ottobre 2024.

CEST sta per **Central European Summer Time**, l'ora legale utilizzata in Italia in questo periodo dell'anno.

Vuoi sapere l'ora in un'altra città o paese? Basta dirmelo e te la fornisco subito.

Oppure, posso aiutarti con:

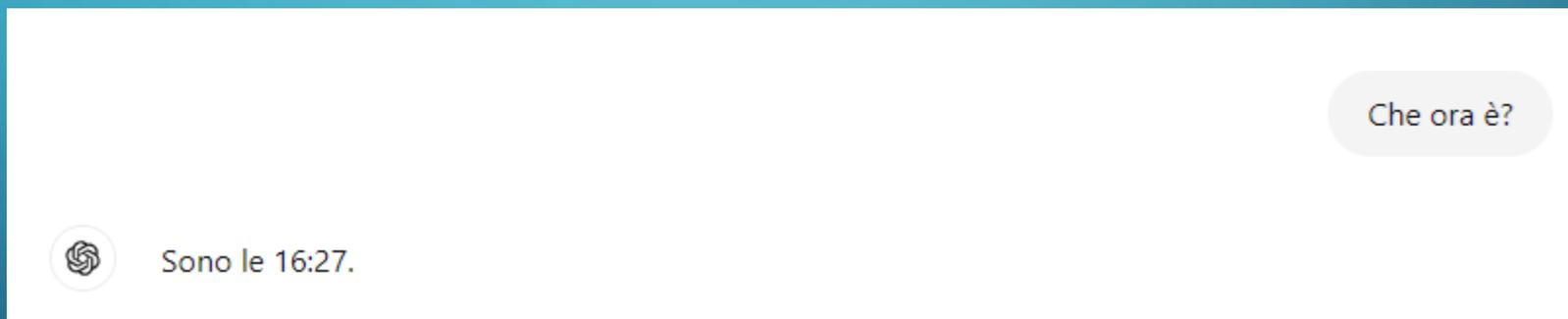
- **La conversione di fusi orari:** Se hai bisogno di sapere l'ora in un luogo con un fuso orario diverso.
- **Informazioni sul cambio dell'ora:** Quando inizierà e finirà l'ora solare in Italia.
- **Altre domande sul tempo:** Come il meteo o le fasi lunari.

Non esitare a chiedere!

CHE ORA È?

... È ora di non farsi fregare!

ChatGPT OGGI, ha finalmente imparato!



CHE ORA È?

... È ora di non farsi fregare!

Tecnica RAG - Retrieval Augmented Generation – per risolvere i problemi precedenti e altri problemi

1. Cosa fa prima del prompt per il Chatbot AI? Ricerca dati nuovi o dati calzanti rispetto al contesto
2. Poi li combina in base a quello che sa
3. Successivamente attiva il Chatbot AI

SI PERDERANNO VERAMENTE TANTI POSTI DI LAVORO?

l'IA sta trasformando il mercato del lavoro, portando sia sfide che opportunità. È importante che i lavoratori si adattino e acquisiscano nuove competenze per rimanere competitivi in questo nuovo scenario.

SUL CAMBIAMENTO

Roy T. Bennett, politico, è ricordato per il suo coraggio e il suo impegno incrollabile per la riforma politica e i diritti umani in Zimbabwe.



SE VI È PIACIUTO CONDIVIDETE CON ALTRI
SE NON VI È PIACIUTO CONDIVIDETE CON ME

GRAZIE PER L'ATTENZIONE E ARRIVEDERCI!

carla.trabuio@gmail.com