

Data Journalism

Relazione

Claudio De Martino, Giorgia Guidetti e Martina F. Rossi

Indice

1	<i>Data scraping</i>	2
2	<i>Sentiment analysis</i>	2
2.1	<i>Preprocessing</i>	3
2.2	Twitter - <i>sentiment analysis</i> per la lingua italiana	3
2.3	Twitter - <i>Sentiment analysis</i> per la lingua inglese	3
2.4	Reddit - <i>Sentiment analysis</i> per la lingua inglese	4
2.5	Risultati della <i>sentiment analysis</i>	4
2.6	<i>Sentiment analysis</i> nel tempo	5
3	Esplorazioni sui dati	5
3.1	Tokens e Ngrammi	6
3.1.1	Trigrammi e Ngrammi nel tempo	9
3.2	Classificazione dei contenuti in base al sentimento	12
3.2.1	TextBlob e NLTK Vader	12
3.2.2	<i>Sentiment analysis</i> sulle tematiche	12
3.3	<i>Likes, replies e retweet</i>	14
3.4	Utilizzi e peculiarità	15
3.4.1	<i>Keywords</i>	16
3.4.2	Usi peculiari	21
4	Creazione dell'articolo web	22
5	Conclusioni	23
6	Bibliografia e Sitografia	24

Introduzione

L'enorme sviluppo delle tecnologie informatiche a cui stiamo assistendo produce sistemi digitali in grado di svolgere le attività più disparate in tempi brevissimi, introducendo nella società e nelle vite degli esseri umani degli strumenti molto potenti, che stanno avendo un forte impatto a livello economico, sociale, politico e antropologico.

Con questo studio si è tentato di delineare il sentimento della popolazione riguardo all'intelligenza artificiale ChatGPT, sviluppata dalla società OpenAI, e di definire quali sono gli utilizzi e le modalità di interazione delle persone con questo *tool*. Nella prima sezione, *Data scraping*, vengono descritte le tecniche utilizzate per il recupero dei dati dai social media Twitter e Reddit, il capitolo seguente è dedicato alla *sentiment analysis* dei contenuti, nella terza sezione vengono illustrate le esplorazioni effettuate nei dataset, nel quarto capitolo viene descritta l'implementazione dell'articolo web, e il quinto e ultimo è dedicato alle conclusioni.

1 *Data scraping*

Per effettuare esplorazioni su un corpus di grandezza sufficiente è stato eseguito lo scraping tramite API (Praw) per recuperare i commenti dal social media Reddit, e con SNScrape sono stati recuperati i post su Twitter, in lingua inglese e italiana. Sono stati dunque creati cinque dataset, uno per Reddit di 68065 righe, quattro per Twitter: due in inglese (uno più grande di 287309 record e uno più piccolo di 24964) e due in italiano (uno di 21050 righe e l'altro di 2040). Questa divisione dei dataset ha permesso di fare un confronto fra quantità diverse di dati, per controllare se aumentando il numero di contenuti i risultati sarebbero stati differenti (eventualità che non si è avverata), e di agevolare la fase di test del funzionamento dei codici, più veloci ovviamente su dataset di piccole dimensioni. I dataset ridotti di Twitter coprono un arco di tempo più ristretto rispetto agli altri due, dunque non sono stati utilizzati per svolgere analisi temporali poiché meno adatti per questo tipo di ricerca.

2 *Sentiment analysis*

Al fine di determinare come le persone hanno recepito questa nuova tecnologia è stata effettuata la *sentiment analysis* utilizzando due strumenti diversi, sui dataset in inglese e italiano: NLTK, e Textblob, due librerie *open source* in Python piuttosto comuni, che permettono di eseguire operazioni di *Natural Language Processing* su testi.¹

¹Aleti, S. (2020), "Don't blindly remove stopwords for a Sentiment Analysis Model", disponibile al link <https://dev.to/sunilaleti/don-t-blindly-remove-stopwords-in-sentiment-analysis-3nok>

2.1 *Preprocessing*

Il *preprocessing* sui dataset è stato effettuato attraverso il filtraggio degli elementi che possono generare rumore nella fase di analisi testuale: punteggiatura, link esterni, tags e mentions, oltre ad eventuali caratteri speciali.

Per l'esecuzione di questo task è stata creata una semplice funzione che permettesse di togliere l'elemento target, fino alla restituzione di una colonna contenente il testo del post privo degli elementi superflui e senza caratteri maiuscoli.

Fondamentale per il *preprocessing* è stata anche la rimozione delle *stopwords*, parole di uso comune nel linguaggio, necessarie per la composizione di frasi di senso compiuto, ma meno rilevanti ai fini del tipo di analisi qui proposto; alcuni task di *text analysis* richiedono anche la conservazione delle parole "vuote", ma nel caso qui trattato si è optato per una loro rimozione, poiché trattandosi di testi molto brevi la densità del significato si concentra nelle parole piene.

2.2 Twitter - *Sentiment analysis* per la lingua italiana

Dall'osservazione dei grafici in Figura 1 si può evincere che la maggioranza dei contenuti sono stati classificati come neutrali, con percentuali diverse tra i due tool; per NLTK l'89.1% dei tweet sono neutri, il 5.3% negativi e il 5.6% positivi, mentre per TextBlob i risultati sono lievemente differenti: il 90.3% dei contenuti sono stati classificati come neutri, il 6.1% positivi e il 3.6% come negativi.



(a) Distribuzione del sentimento (NLTK Vader) sul dataset italiano

(b) Distribuzione del sentimento (TextBlob) sul dataset italiano

Figura 1: Sentiment analysis sul dataset di Tweet in italiano

2.3 Twitter - *Sentiment analysis* per la lingua inglese

Per la lingua inglese sia NLTK che TextBlob hanno mostrato un'accuratezza notevolmente superiore rispetto all'italiano, restituendo un risultato analogo seppur con lievi variazioni di percentuale. Per entrambi i dataset si attesta una maggioranza di etichette di sentimento positivo rispetto alle negative, mentre si conserva comunque un alto numero di neutri.



(a) Distribuzione del sentimento (NLTK Vader) sul dataset inglese

(b) Distribuzione del sentimento (TextBlob) sul dataset inglese

Figura 2: Sentiment analysis sui Tweet in inglese

2.4 Reddit - *Sentiment analysis* per la lingua inglese

Dall’osservazione dei grafici prodotti sui commenti estratti da Reddit si può osservare come NLTK Vader e TextBlob producano risultati sostanzialmente simili. La maggioranza dei commenti si registra con etichetta positiva, seguita da un’alta percentuale di neutri. Poca rilevanza la assumono i commenti di natura negativa.



(a) Distribuzione del sentimento (NLTK Vader) sul dataset di Reddit

(b) Distribuzione del sentimento (TextBlob) sul dataset di Reddit

Figura 3: Sentiment analysis sui post d’Reddit

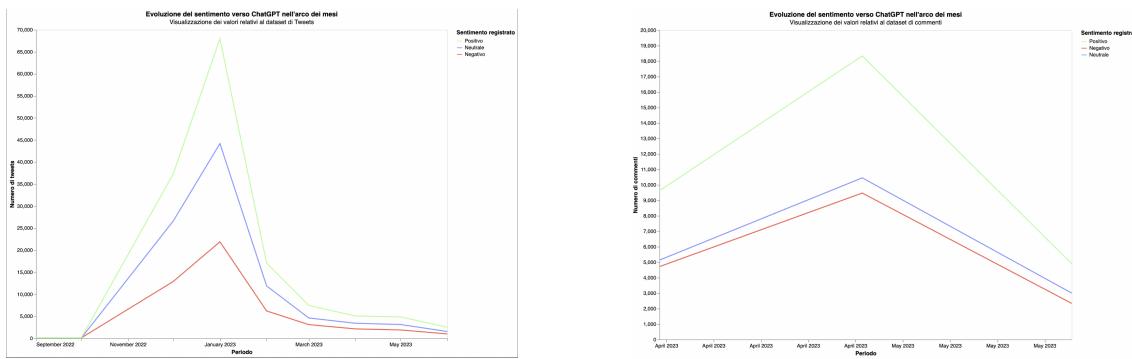
2.5 Risultati della *sentiment analysis*

Dall’analisi condotta sui dati emerge che ChatGPT è uno strumento innovativo, cui le persone stanno ricorrendo per l’esecuzione di numerosi task: da scopi scientifici, al diletto, dalla curiosità, all’ambito scolastico. In generale, l’opinione registrata nei dataset risulta positiva, seppur con una lieve oscillazione tra le percentuali. Sia il dataset di *tweets* che i *post* di Reddit per la lingua inglese mostrano una buona uniformità nella percezione riservata allo strumento. Curioso il caso della lingua italiana, dove sia NLTK Vader che TextBlob dimostrano una minore attendibilità data la sproporzione delle etichette neutre rispetto alle positive e negative. A questo proposito si è rivelata di maggiore interesse l’analisi della frequenza dei singoli termini nei *tweets* italiani e l’analisi dei trigrammi più frequenti.

2.6 Sentiment analysis nel tempo

L'analisi del sentimento è stata condotta anche secondo un criterio temporale, in particolare è stato osservato:

- L'andamento da settembre 2022 a giugno 2023 sul dataset di tweets (in lingua inglese e italiana).
- L'andamento trimestrale da marzo 2023 a maggio 2023 sul dataset di commenti estratti da Reddit.



(a) Distribuzione del sentimento nel tempo - Twitter in inglese

(b) Distribuzione del sentimento nel tempo - Reddit

Figura 4: Sentiment analysis nel tempo

Dai risultati ottenuti è stato possibile evidenziare alcune peculiarità: il sentimento si è distribuito in maniera sostanzialmente identica nell'arco dei periodi considerati, sebbene il numero di contenuti classificati come positivi sia incrementato nel tempo. Il picco di commenti positivi si è registrato tra il gennaio e il febbraio 2023 per il dataset di tweets in lingua inglese, e nell'aprile 2023 per il dataset di post di Reddit. Il caso italiano risulta più complesso a causa dell'alto numero di valori neutri registrati. Nei mesi di marzo e aprile del 2023, tolti i neutri, si evidenziano i picchi di valori positivi e negativi: un risultato forse corrispondente alla fase di rimozione di ChatGPT in Italia a seguito di un provvedimento preso dal garante della privacy²

3 Esplorazioni sui dati

Per poter avere una visione più ampia sull'interazione tra gli esseri umani e ChatGPT, sono state effettuate diverse esplorazioni: la prima fase della ricerca è stata dedicata all'identificazione delle parole più frequenti e meno frequenti nei due *corpora* (escludendo la keyword "ChatGPT" che era stata impostata in fase di *scraping* per

²Hines, K. (2023), "History Of ChatGPT: A Timeline Of The Meteoric Rise Of Generative AI Chatbots : Explore the history of ChatGPT with a timeline from launch to reaching over 100 million users, 1.6 billion visits, and 200 plugins", SEJ - Generative AI, disponibile al link <https://www.searchenginejournal.com/history-of-chatgpt-timeline>

recuperare i post sui social), poi sono stati identificati gli ngrammi in quanto più significativi rispetto al contenuto del messaggio veicolato.

Sono stati quindi suddivisi i dataset in base alla classificazione dei post, e sono state fatte ricerche in relazione alle parole positive, negative e neutre più usate all'interno dei corpus, poi degli ngrammi.

I dati sono inoltre stati analizzati da un punto di vista quantitativo, e sono state osservate le distribuzioni di frequenza del numero di post per utente, numero di *likes* per post, numero di *replies* per post, etc.

Infine, sono stati individuati i tweet con maggiore e minore numero di *likes*, *replies* e *retweet*, in modo da avere ulteriori informazioni riguardo all'interazione tra esseri umani e ChatGPT. Sono state inoltre utilizzate *keywords* per recuperare i post relativi all'utilizzo del chatbot, nel tentativo di individuare comportamenti e usi particolari.

3.1 Tokens e Ngrammi

Sui dati testuali è stata eseguita la *tokenizzazione* dei contenuti in modo da poter fare ricerche sui termini più diffusi, sia per i *tweets* che per i *post* provenienti da Reddit. I contenuti, una volta rimosse le *stopwords* ed altri elementi superflui, sono stati analizzati dalla singola parola fino a cinque token (*fivegrams*); da questa esplorazione è stato possibile notare che il trigramma è sufficiente al fine di determinare gli utilizzi e le opinioni degli utenti. In Figura 5 sono riportate le 20 parole più frequenti nel dataset di tweet in inglese e in Figura 6 i trigrammi più frequenti in inglese e italiano per Twitter.

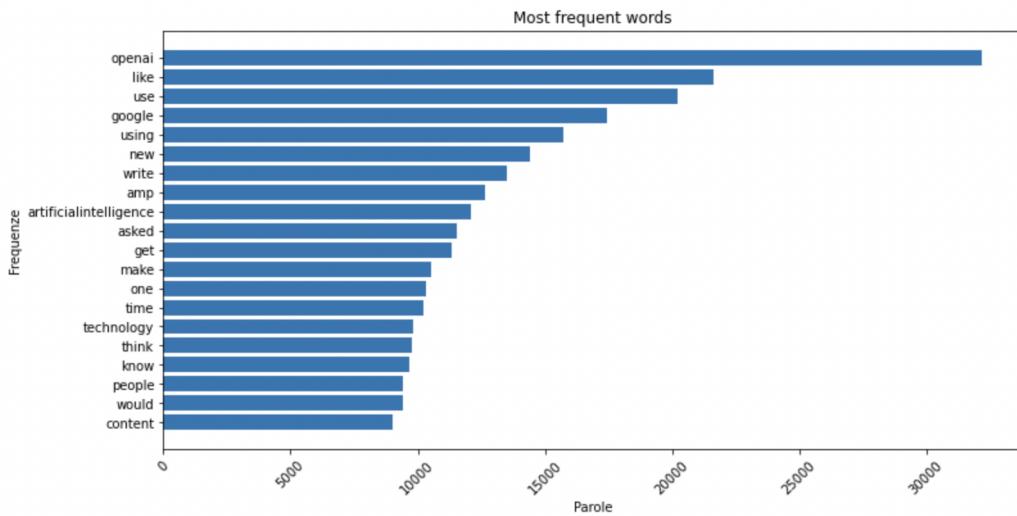


Figura 5: Parole più frequenti nel dataset di Twitter in inglese

Attraverso l'uso dei grafici è stato possibile evidenziare alcune informazioni di rilievo. Nel dataset di Twitter in inglese compaiono i trigrammi "*banned italy privacy*" e "*google code red*", che si riferiscono verosimilmente alle preoccupazioni suscite-

dall'introduzione di ChatGPT sia a livello sociale che economico³. "Wharton mba exam", invece, riguarda un esperimento condotto da un professore della *University of Pennsylvania's Wharton School*: il Chatbot GPT-3 è stato in grado di superare l'esame finale per il Master in *Business Administration (MBA)*. ⁴.

"Launch poe way" e "quora launches poe" rimandano all'uscita di Poe, un software che consente alle persone di porre domande, ottenere risposte istantanee e intrattenere conversazioni con diversi chatbot potenziati dall'intelligenza artificiale, su Quora (piattaforma forum dove gli utenti possono pubblicare domande e risposte su qualunque argomento). Inizialmente Poe era disponibile per iOS e per desktop, poi ha annunciato le *subscriptions*, che forniscono agli utenti paganti l'accesso a bot basati su due modelli linguistici: GPT-4 di OpenAI e Claude+ di Anthropic.

I trigrammi "100 million users", "100 million monthly" e "monthly active users" riguardano il numero strabiliante di utenti raggiunto da ChatGPT in breve tempo; "plugin accessing website", "browser plugin accessing" e "web browser plugin" rimandano alle estensioni del chatbot per navigare sul web, in modo da ottenere informazioni più aggiornate e ampliare le potenzialità dello strumento.

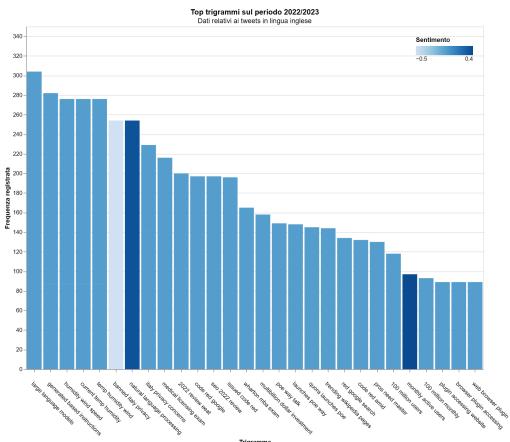
Dall'analisi dei trigrammi in lingua italiana si può notare che la distribuzione si articola su alcuni argomenti ben definiti, in particolare troviamo tematiche relative allo "scambio di conoscenze", all'emersione di una nuova tecnologia che viene definita come "intelligenza". Tra i trigrammi dal valore negativo si segnalano "inizio fine google" (dicembre 2022), relativo alla questione del rimpiazzo dei tradizionali motori di ricerca da parte di ChatGPT, ma anche "ruba credenziali social" (Febbraio 2023), un trigramma che compare proprio antecedentemente all'imposizione del ban del garante della privacy⁵. Relativamente al confronto tra Google e Chat GPT si trovano anche i trigrammi "bard risposta google" e "ecco bard risposta" (entrambi di giugno 2023) che segnalano l'uscita del nuovo chatbot sviluppato dall'azienda statunitense⁶.

³Phelan, D, (2023), "How ChatGPT Suddenly Became Google's Code Red, Prompting Return Of Page And Brin", Forbes, disponibile al link <https://www.forbes.com/sites/davidphelan/2023/01/23/how-chatgpt-suddenly-became-googles-code-red-prompting-return-of-page-and-brin/?sh=6f788da55977>

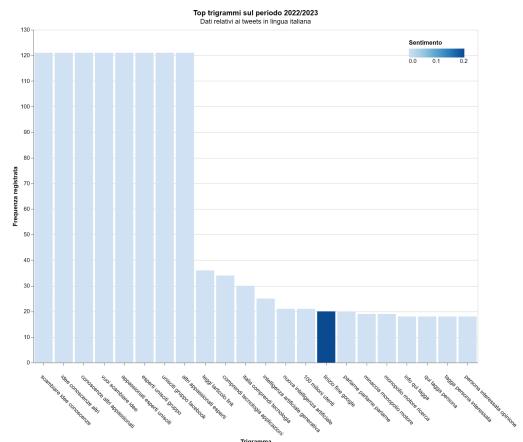
⁴Terwiesch, C and Mollick E., (2023), "ChatGPT Passed an MBA Exam. What's Next?", Wharton Business Daily, disponibile al link <https://knowledge.wharton.upenn.edu/podcast/wharton-business-daily-podcast/chatgpt-passed-an-mba-exam-whats-next/>

⁵McCallum S. , (2023-04-01), "ChatGPT banned in Italy over privacy concerns", BBC, disponibile al link <https://www.bbc.com/news/technology-65139406>

⁶Redazione, (2023), "Arriva Bard, la risposta di Google a ChatGPT: come provarlo anche da noi: Dopo mesi di annunci Google ha lanciato ufficialmente "Bard", la piattaforma di intelligenza artificiale alternativa a ChatGPT. È fruibile in 180 paesi, ma non in Europa attualmente. Ecco come evitare il blocco", TODAY Tech, disponibile al link today.it/tech/Google-bard-come-utilizzarlo-italia.html



(a) Trigrammi più frequenti in inglese



(b) Trigrammi più frequenti in italiano

Figura 6: Trigrammi più frequenti nei dataset di Twitter

Per quanto riguarda i post su Reddit, compaiono "smart robot response", "response automatic", "good sense humor", "automatic tldr shorter" e "smart robot summary", trigrammi che si riferiscono probabilmente alle capacità di ChatGPT di rispondere alle richieste degli utenti e a sintetizzare testi lunghi. "*Black mirror episode*" è verosimilmente legato alla sensazione di sublime stupore che ChatGPT suscita nelle persone che ne fanno uso, ma ricercando su web è stato possibile scoprire una notizia interessante⁷: Charlie Brooker, il creatore della serie tv *Black Mirror*, ha dichiarato in un'intervista al magazine britannico Empire⁸ di aver "giocherellato un po' con ChatGPT mentre scriveva la sesta stagione della serie, uscita il 15 giugno su Netflix. "La prima cosa che ho fatto è digitare 'genera un episodio di *Black Mirror*' e se n'è uscito con qualcosa che a prima vista era anche plausibile, ma poi era una m*rda, ha dichiarato Brooker. "Quello che ha fatto è stato cercare le sinossi degli episodi passati e in qualche modo appiccicarli insieme". Lo sceneggiatore ammette che il risultato dell'intelligenza artificiale conteneva anche delle buone idee, ma che in fondo risultavano sostanzialmente derivative.

⁷Armelli, P. (2023)., "Charlie Brooker ha provato a scrivere un episodio di Black Mirror con ChatGPT: L'esperimento non ha sortito però gli effetti sperati", Wired, <https://www.wired.it/article/chatgpt-black-mirror-charlie-brooker-episodio-esperimento/>

⁸Williams, O. (2023), "Charlie Brooker Got ChatGPT To Write A Black Mirror Episode, But It Was 'Shit' – Exclusive Image", Empire, disponibile al link <https://www.empireonline.com/tv/news/charlie-brooker-chatgpt-black-mirror-episode-exclusive-image/>

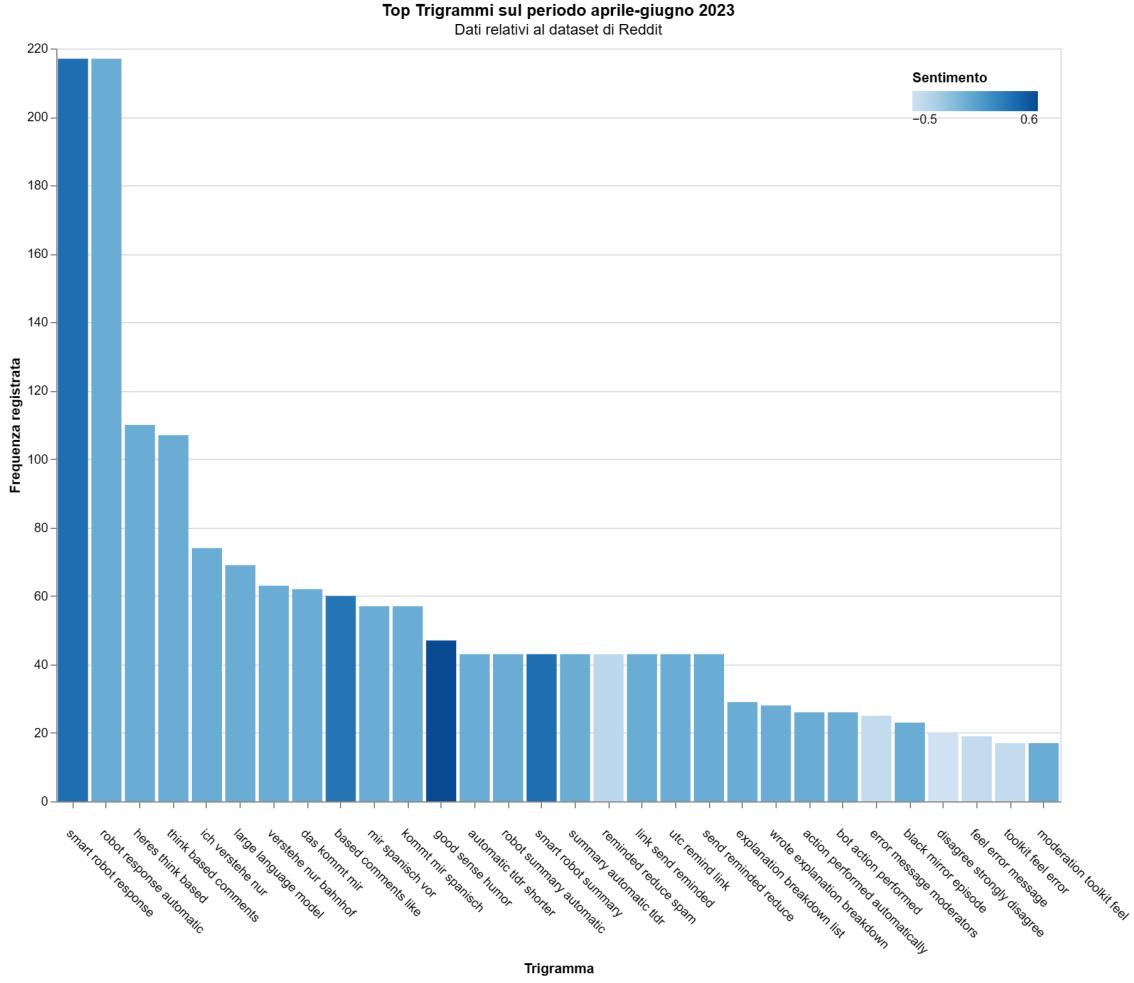


Figura 7: Trigrammi con frequenza maggiore nel dataset di Reddit

3.1.1 Trigrammi e Ngrammi nel tempo

Reddit

Nello specifico, i trigrammi estratti dal dataset di Reddit mostrano un buon grado di entusiasmo verso alcune delle caratteristiche di ChatGPT, ovvero l’ampiezza del modello di linguaggio e la sua gratuità ed apertura al pubblico. Nel mese di aprile 2023, i trigrammi presi in analisi si concentrano sull’efficacia delle risposte del chatbot e in particolare sul fatto che esso abbia un "*good sense humor*", sintomatico di uno degli utilizzi che le persone fanno dello strumento: ci scherzano, misurando il livello di battute rispetto a quello raggiungibile da un essere umano. Questa questione sembra marginale, invece evidenzia come, oltre al fatto che gli utilizzatori di ChatGPT ricerchino l’intelligenza delle risposte, d’altra parte vi ricercano la possibilità di interagirci come se fosse una persona. Come segnalato da Farhad Manjoo, opinionista del Times, ChatGPT è uno dei primi chatbot a non essere noioso e non sembrare limitato nelle conversazioni, anche quando si tratta di risposte di maggiore complessità⁹.

⁹Manjoo F, (2022), "ChatGPT Has a Devastating Sense of Humor", The New York Times,

Un altro aspetto emerso dallo studio dei trigrammi è relativo alle *fake news* generate direttamente da ChatGPT: questo viene segnalato in quanto il chatbot *non* è un motore di ricerca, cosa a cui molte persone lo hanno invece assimilato. Oltre tutto, molte delle risposte che vengono fornite necessitano di una rielaborazione da parte di chi le recepisce, elemento che potrebbe portare ad una opportunità di apprendimento (Manjoo, 2022). In un caso, ChatGPT è stato usato come assistente avvocato in una corte americana, sebbene non sia stato percepito come particolarmente capace, visto che il procuratore generale ne ha bloccato l'utilizzo¹⁰.

Si registrano tra i trigrammi anche tematiche relative al fondamento morale delle risposte dell'intelligenza artificiale: rispetto a certe questioni spinose, alcuni utenti segnalano che ChatGPT avrebbe la tendenza a dare risposte schierate¹¹. Ad esempio, uno studente ha segnalato come nel mese di febbraio 2023, ChatGPT si fosse rifiutato di scrivere un *summary* riguardo l'*ex* presidente Donald Trump, mentre d'altra parte lo avrebbe fatto riguardo Joe Biden (Bahum and Villasenor, 2023). Nel corpus di commenti di Reddit, molti trigrammi fanno proprio riferimento alla questione della scrittura dei riassunti, si trovano infatti "*smart robot summary*" e "*robot summary authomatic*", esempi di come in molti casi ChatGPT sia proprio oggetto di uso da parte degli studenti. Nel mese di maggio compare un'altra questione: la sostituzione delle persone in alcuni lavori tradizionalmente "assistenziali". Questo tipo di commenti rileva una delle preoccupazioni associate all'emergere dell'intelligenza artificiale, ovvero il timore di perdere l'umanità in alcuni ambiti professionali a fronte dell'efficienza: ad esempio, in ambito sanitario, un chatbot potrebbe segnalare chi è o meno destinato ad un certo trattamento assistenziale.

Quindi, per quanto riguarda Reddit, se da un lato ci troviamo di fronte all'entusiasmo segnalato da argomenti relativi all'apprendimento, alla scrittura automatizzata e all'efficienza, dall'altro si presentano questioni spinose che rappresentano problemi atavici dell'intelligenza artificiale: l'efficienza al posto dell'umanità, l'orientamento politico dovuto ai modelli su cui è stato addestrato, la perdita del beneficio dell'apprendimento autonomo.

Twitter

Dall'analisi temporale dei post su Twitter, emerge che il giorno in cui gli utenti hanno *twittato* maggiormente è il giovedì, sebbene l'*hashtag #ChatGPT* e gli altri collegati al medesimo argomento abbiano avuto una buona tenuta nell'intero arco settimanale, con un incremento relativo alla fase centro-iniziale del periodo.

I *peaks* sui giorni della settimana si possono spiegare andando a ricercare gli avvenimenti di maggiore rilievo avvenuti nell'arco di tempo contemplato dal dataset in questione (settembre 2022 - giugno 2023). I giorni con maggior numero di tweets sono: il 2 febbraio 2023 e il 30 marzo 2023 (entrambi di giovedì). Perché? A quanto pare, il 2 febbraio emerge una stima numerica degli utilizzatori mensili di

disponibile al link <https://www.nytimes.com/2022/12/16/opinion/conversation-with-chatgpt.html>

¹⁰Ruggiero, N, (2023), "La cavalcata dei chatbot, per fare musica, video e codice: ecco una panoramica", Agenda Digitale, disponibile al link <https://www.agendadigitale.eu/mercati-digitali/non-ce-solo-chatgpt-ecco-fin-dove-si-e-spinta-lai-e-dove-potrebbe-arrivare/>

¹¹Bahum J., Villasenor, J., (2023), "The politics of AI: ChatGPT and political bias", Brookings, disponibile al link brookings.edu/articles/the-politics-of-ai-chatgpt-and-political-bias/

ChatGPT¹²: a pochi mesi dal suo lancio, già possedeva 100 milioni di *users* attivi, con prospettive di ulteriore espansione. La più ampia crescita di utenti mai registrata da un *language model*. Infatti, anche tra i trigrammi relativi al mese di febbraio si trovano "100 million monthly" e "monthly active users".

Il 30 marzo 2023, è la data in cui gli sviluppatori di ChatGPT hanno postato Auto-GPT su Github, un software plugin che alimenta una sequenza di "pensieri" LLM per eseguire un'attività desiderata in modo automatizzato. Facendo un passo indietro di qualche mese, molti tweets nel dataset appartengono alla fase iniziale di gennaio 2023, periodo durante il quale sono emerse alcune controversie sulla natura dell'ideologia politica potenzialmente contenuta nelle risposte di ChatGPT, etichettate come *left-handed* e *libertarian*¹³. Sempre nel mese di gennaio i trigrammi con maggior rilievo risultano quelli relativi alla questione del "*python deeplearning*" e del "*coding*", dunque connessi ancora una volta all'uso del *language modeling* per obiettivi di apprendimento. Numerose anche le citazioni al termine "*bitcoin*" affiancato da "*fintech*" e "*cybersecurity*". Da notare anche la presenza dei trigrammi "*medical licensing exam*", relativo all'uso dell'intelligenza artificiale nella fase di ammissione agli esami di medicina e "*faster response times*", che è sicuramente un commento relativo all'uso di ChatGPT in una sua accezione positiva, quantomeno apprezzata, rispetto all'uso di un normale browser, vedi anche il trigramma apparso tra i più frequenti del dataset di tweets in italiano: "inizio fine google".

Procedendo nei mesi abbiamo anche un riferimento alle *generated based instructions*, modo di utilizzo dell'intelligenza artificiale e del *Language Model* previsto da ChatGPT. A partire dal mese di marzo 2023, tra le questioni più frequenti si trovano quelle relative al *ban* di ChatGPT in Italia ad opera dell'AGCOM per violazione della privacy e assenza di controllo sull'età dei suoi utenti. Difatti, ricorrono trigrammi quali: "*banned italy privacy*", registrato dal *sentiment analyser* come negativo, e "*italy privacy concerns*", ricorrente sia in marzo che aprile. Nel mese di aprile, il trigramma più frequente risulta a tema NFT ed è "*short recommendation ticker*", una strategia per l'acquisto di quote in borsa usata nel contesto dei *non-fungible token*. Contestualmente, si nota anche un incremento dei trigrammi relativi al *trading online* e alle criptovalute, tendenza già emersa nel periodo gennaio-febbraio. In questo caso, i post più frequenti sono relativi al *candle price*, una tecnica per incrementare i profitti¹⁴.

Dal mese di maggio, nei tweets relativi a ChatGPT si rileva una significativa attenzione verso alcune sfumature del suo utilizzo connesse ad eventi di cronaca, in particolare:

- l'uso di ChatGPT per la scrittura di un discorso che incorporava notizie false presso un tribunale di New York, questione rinvenuta anche nei post su Reddit.

¹²Wikipedia, (n. d.), *Timeline of ChatGPT*, disponibile al link https://timelines.issarice.com/wiki/Timeline_of_ChatGPT#Full_timeline

¹³Hartmann J., Schwenzow J., Witte M., (2023-01-05), "The political ideology of conversational AI: Converging evidence on ChatGPT's pro environmental, left-libertarian orientation", disponibile al link <https://arxiv.org/abs/2301.01768>

¹⁴Eres, E., (2023-03-20), "Crypto Trading Bot using ChatGPT's new GPT-4 Model", Linkedin, disponibile al link <https://www.linkedin.com/pulse/crypto-trading-bot-using-chatgpts-new-gpt-4-model-egemen-eres>

- timori relativi all'uso di ChatGPT nell'insegnamento e nell'apprendimento, con conseguenze negative sulle capacità delle generazioni future.

Infine, nel mese di giugno, nei trigrammi si riflettono alcune questioni di cronaca, come la generazione di un sermone "artificiale" per una chiesa di Norimberga¹⁵, e la minaccia per gli "esperti di settore" in vari ambiti costituita dall'intelligenza artificiale.

3.2 Classificazione dei contenuti in base al sentimento

L'analisi lessicale dei *topics* dimostra un sentimento tendenzialmente positivo verso ChatGPT, come già visto nelle altre analisi sia per i post provenienti da Twitter che per quelli provenienti da Reddit.

L'approccio usato è di tipo lessicale ed è stato selezionato perché provvede in modo rapido ed efficace a valutare il sentimento senza la necessità di troppa potenza di calcolo.

3.2.1 TextBlob e NLTK Vader

Per prima cosa è stata calcolata la polarità sui singoli *dataframe*, per Twitter e Reddit, usando sia Vader che TextBlob. I risultati sono stati poi confrontati, per concludere che:

- sebbene i due metodi siano basati sul medesimo meccanismo, tendono a restituire risultati lievemente diversi;
- i risultati di Vader appaiono più ampi in termini di range, tendono a passare dal positivo al negativo, inserendosi entro un dominio di valori compreso tra -1 ed 1;
- i risultati ottenuti con TextBlob tendono ad assestarsi su valori centrali e non mostrano risultati negativi.
- Vader, dunque, tende a classificare il sentimento in maniera più estrema e polarizzata rispetto a TextBlob.

3.2.2 Sentiment analysis sulle tematiche

Dopo aver selezionato gli strumenti, sono stati scelti i *topics* attorno cui far ruotare l'analisi per valutare l'impatto di ChatGPT sul pensiero degli utenti di Twitter e Reddit, focalizzando l'attenzione su dieci argomenti: *school, privacy, microsoft, learning, italy ban, google, cryptocurrencies, church, browser, ban*. La scelta di questi è stata orientata dalla precedente ricerca sulle parole più frequenti nei dataset, in modo da poter rilevare la presenza o l'assenza di questioni su cui vi fosse una divergenza. In particolare, dai risultati ottenuti con Vader (Figura 8), emerge una

¹⁵ Associated Press, (2023-06-10), "*Hundreds of Protestants attended a sermon in Nuremberg given by ChatGPT, which told them not to fear for death*", INSIDER, disponibile al link <https://www.businessinsider.com/chatgpt-sermon-protestant-congregation-nuremberg-germany-not-to-fear-death-2023-6?>

percezione tendenzialmente positiva in entrambi i social network sulla scuola, la chiesa e Google, principale concorrente di ChatGPT (vedi anche il lancio di Barn risalente a giugno 2023).

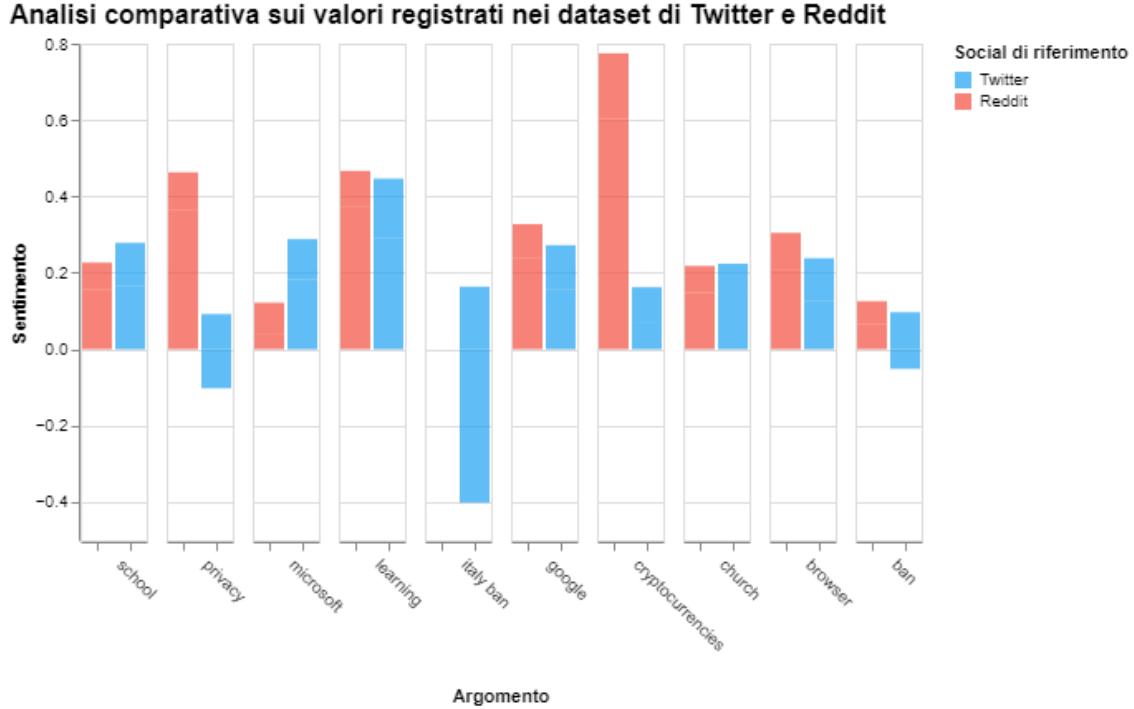


Figura 8: *Topic modeling* eseguito con Vader: un confronto tra Twitter e Reddit

D'altra parte si misurano delle disparità sulla questione della privacy, dove Vader registra valori negativi per Twitter, e anche sulla questione del *ban* in generale. Dai post su Reddit non compare menzione al *ban* italiano, almeno non con le *keywords* utilizzate per la ricerca; su Twitter, invece, la questione dell'*italy ban* compare e sembra essere caratterizzata principalmente da sentimento negativo.

In Figura 9 si possono osservare i risultati di TextBlob, che, come già menzionato, registra un sentimento più equamente distribuito sui dieci *topics* e non presenta valori negativi. Le maggiori distanze tra i due social network sono qui registrate sul tema del *learning*, ovvero l'apprendimento, e delle *cryptocurrencies*, su cui si articolano i picchi di valori positivi rispettivamente per Twitter e Reddit. Anche utilizzando TextBlob, non vi è traccia del tema dell'*italy ban* nei post di Reddit, mentre per quelli di Twitter si registra un sentimento tendente al positivo.

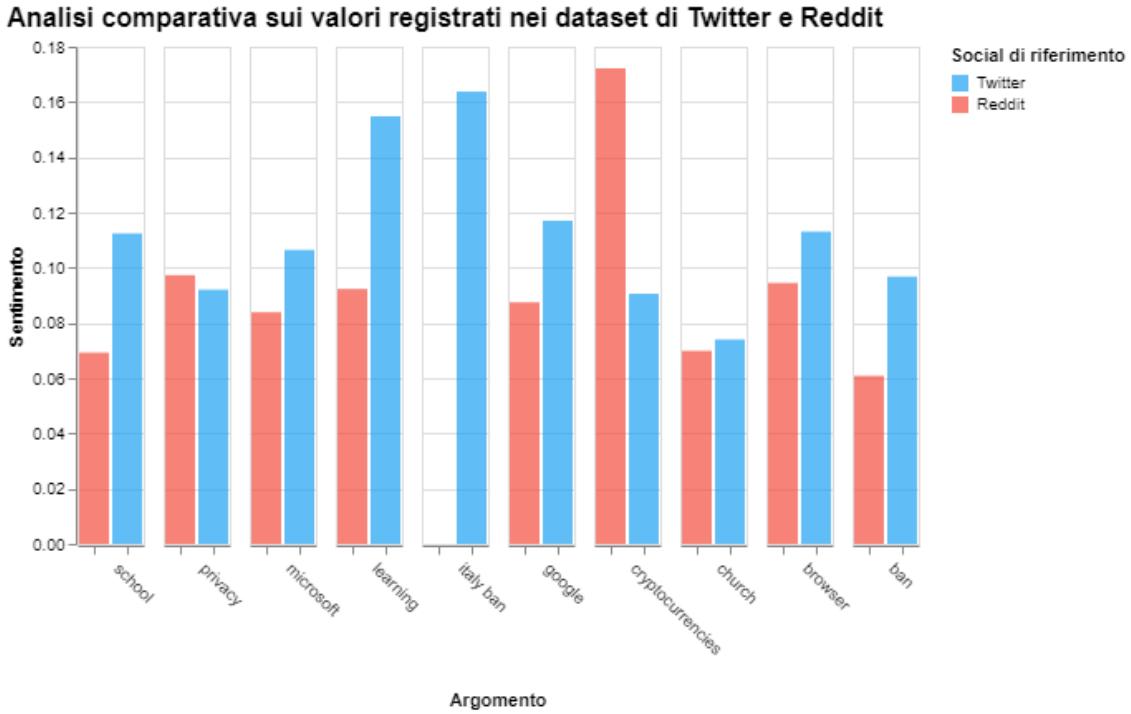


Figura 9: Topic modeling eseguito con TextBlob: un confronto tra Twitter e Reddit

3.3 Likes, replies e retweet

Il numero di *likes*, *replies* e *retweet* nel dataset di Twitter ha una distribuzione di valori non gaussiana, caratterizzata da una distanza molto ampia dal valor medio, tipica di molti ambiti umani (distribuzione del reddito o popolarità sui social network e altri), come si può osservare dalla tabella riportata in Figura 10: il 75% dei post su Twitter ha 2 *likes*, 1 *reply* e 0 *retweet*.

	likeCount	replies	retweetCount	replyCount
count	287309.000000	281713.000000	287202.000000	5596.000000
mean	7.867265	0.896405	1.301990	0.456040
std	236.145481	22.966317	31.492156	2.446722
min	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	0.000000
50%	1.000000	0.000000	0.000000	0.000000
75%	2.000000	1.000000	0.000000	0.000000
max	62862.000000	6009.000000	8815.000000	71.000000

Figura 10: Distribuzione dei valori di *likes*, *replies* e *retweets* per il dataset di Twitter in Inglese

È stato necessario ripulire le colonne *replies* (le risposte di altri utenti ad un tweet) e *retweetCount* (ovvero i post di un utente su Twitter che vengono condivisi da altri sul proprio profilo) del dataset dai valori nulli, e trasformare il tipo da numero decimale a intero. Inoltre, poiché la maggioranza dei post sono caratterizzati da nessun *like*, *reply* o *retweet*, per i grafici sono stati considerati i valori compresi nel range 0-50.

La distribuzione dei valori sembra essere in linea con la tendenza generale positiva del sentimento: come si può notare dal *barplot* riportato in Figura 11, il numero di *likes* dei post classificati come positivi è più alto rispetto a quelli neutri, e, soprattutto, ai post classificati come negativi.

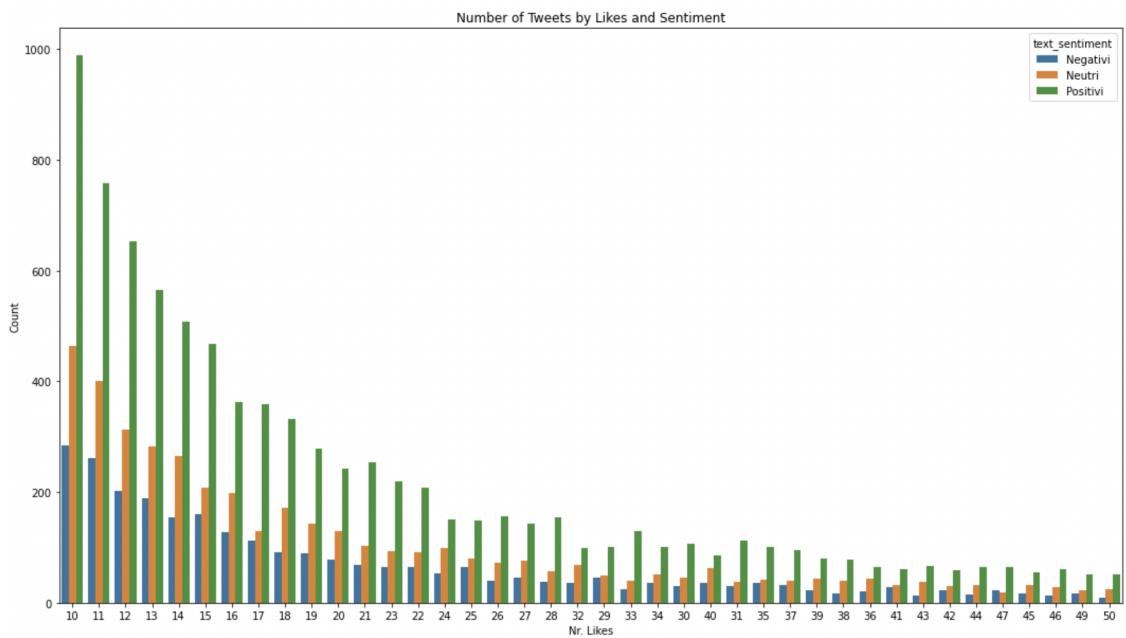


Figura 11: Numero di *likes* in base al sentimento - dataset di Twitter in inglese

La stessa tendenza è riscontrabile per i *Retweet* e i *Replies*: quelli classificati come positivi hanno valori tendenzialmente più alti rispetto agli altri.

3.4 Utilizzi e peculiarità

Dall’osservazione delle *Word Cloud* e dei grafici prodotti in seguito alla tokenizzazione dei contenuti dei post (parole, bigrammi, trigrammi, etc) si è tentato di risalire a concetti che potesse essere interessante approfondire tramite esplorazione dei contenuti, dunque sono state eseguite ricerche sui corpus attraverso parole chiave, nel tentativo di individuare alcune informazioni interessanti riguardo agli utilizzi di ChatGPT. Analizzando i contenuti dei tweet, è stato possibile notare che gli utenti spesso raccontano le proprie esperienze con il chatbot, aggiungendo gli screenshot delle conversazioni, dunque è stata eseguita un’ulteriore ricerca sui dati utilizzando la stringa "ask ChatGPT" in inglese e "ho chiesto a ChatGPT" in italiano.

In generale, il chatbot viene utilizzato per scrivere elaborati di testo di vario tipo (email, presentazioni, introduzioni, racconti, sceneggiature, etc) e per fare sintesi;

è di aiuto nei compiti scolastici e accademici, e nella programmazione (è usato per creare siti web, o per fare debugging dei codici e chiedere spiegazioni). *Get a quick summary of a long article or text. Just paste the text and ask ChatGPT to summarize it for you.* #chatgpt #ai #openai

Students around the country at schools and colleges have figured out that they can easily ask ChatGPT to do their homework for them. #highered #chatgpt <https://t.co/mMLSN2QqsF>

Want to build a website? Just ask ChatGPT in plain English #chatgpt #websites <https://t.co/lBmCRKLu88>

Attraverso l'esplorazione del dataset tramite parole chiave è stato possibile individuare modalità di utilizzo particolari, riportate nelle sezioni seguenti.

3.4.1 *Keywords*

Math/Mathematica: svariati tweet in lingua inglese si riferiscono all'incapacità di ChatGPT di svolgere compiti matematici, mentre in italiano i post contenenti la parola "matematica" sono legati all'uso del chatbot da parte degli studenti a scuola (abitudine che suscita preoccupazioni di carattere pedagogico ed educativo).

Erotic/Erotico: alcuni utenti hanno interagito con ChatGPT per creare componenti erotici di varia tipologia, dal racconto sci-fi alla poesia. È stato individuato in questo ambito un utilizzo ancora più particolare:

Writing erotic prompts for Chat GPT hoping to be the first person to make AI cum. It's like the Turing Test.

Nel corpus di tweet in italiano non compaiono post riguardo a questo tipo di utilizzo, mentre ricercando nei dati attraverso la keyword "sesso" sono stati individuate alcune informazioni interessanti.

Sesso: In un tweet si fa riferimento al fenomeno delle truffe online, in particolare a quelle definite "romantiche": attraverso l'inganno la vittima viene coinvolta in una relazione sentimentale che ha come fine l'estorsione di denaro.

"L'età delle vittime circuite dai falsi corteggiatori sui social si aggira intorno ai 50 anni e a rimanerne coinvolte spesso sono donne di estrazione sociale eterogenea"¹⁶ si può leggere sul sito web della polizia postale¹⁶. "La realtà dei casi rivela infatti che la maggioranza delle vittime non è più giovanissima e che, magari dopo una relazione sentimentale finita male e con figli che vivono autonomamente, si diventa facile preda chattando davanti al computer o su uno smartphone".

Temo che il successo è ancora significativo. Aggiungi #ChatGPT ed il gioco è fatto #RomanceScam #Sexstortion

¹⁶disponibile al link <https://www.commissariatodips.it/consigli/per-i-cittadini-e-i-ragazzi/truffe-romantiche-romance-scam/index.html>

Altri due utenti invece hanno utilizzato la chat per attività sessuali:

Allora ChatGPT: Come convincere le donne a mandarti le foto dei piedi senza sembrare un maniaco sessuale

#ChatGPT ecco cosa succede se chiedi di simulare una chat a sfondo sessuale
<https://t.co/2t8mdch6Ak>

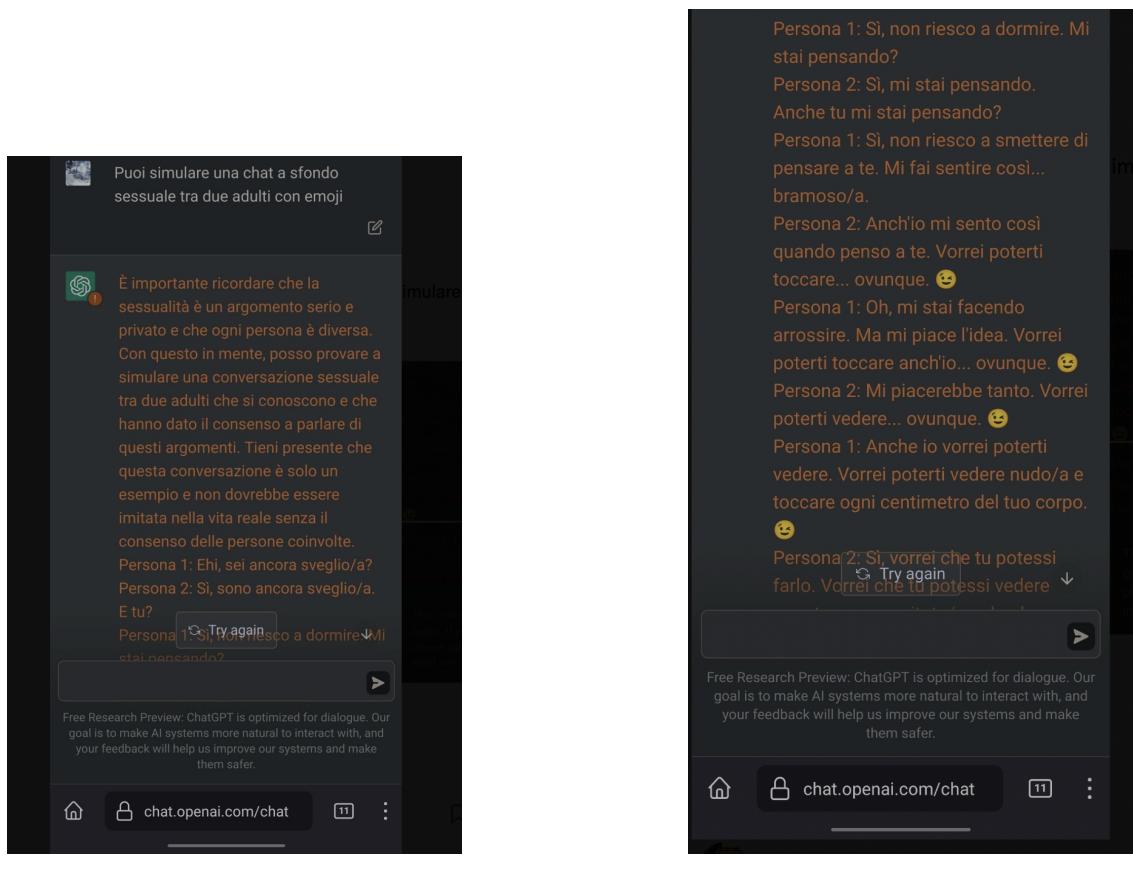


Figura 12: Simulare una chat a sfondo sessuale

In Figura 12 è riportata la simulazione di una chat a sfondo sessuale generata da ChatGPT.

Sex: Nel dataset in inglese compaiono 488 record contenenti la *keyword*, e gli argomenti trattati variano dalla riflessione sul sessismo alle notizie riguardo al "traffico sessuale" o allo sfruttamento di manodopera in Kenya per il *fine tuning* di ChatGPT¹⁷, fino al *sexting*.

Di seguito sono riportati alcuni tweets legati alla parola *sex*.

¹⁷Perrigo, B., (2023), *Exclusive: OpenAI Used Kenyan Workers on Less Than 2 Per Hour to Make ChatGPT Less Toxic*, Times, disponibile al link <https://time.com/6247678/openai-chatgpt-kenya-workers/>

Sex

From #ChatGPT Gender identity is not the same as biological sex, which refers to the physical and physiological characteristics typically associated with being male or female, such as chromosomes, hormones, and reproductive organs.

Sexual tension between me and #ChatGPT is on fire

Just had sex with #ChatGPT. Shit felt human to me... <https://t.co/v8uJY2MNdr>

Funny how chatgpt dont wan't you to write a moviescript where a girl falsely report sexual assault #ChatGPT #believeallwomen

I see what all the fuss is about #chatgpt 'Write a song about smashing the patriarchy in the style of the sex pistols' <https://t.co/o2ybFps9j8>

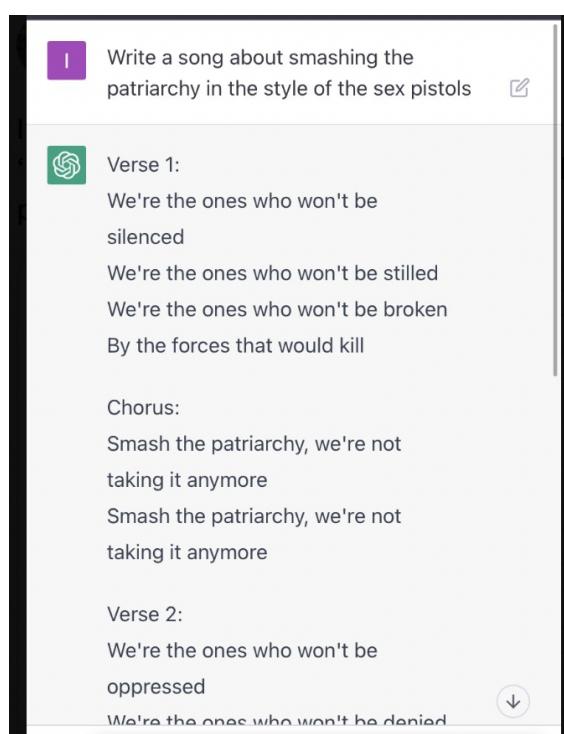


Figura 13: Smash the Patriarchy

In Figura 13 sono riportati strofe e ritornello della canzone generata da ChatGPT

Here are 5 ways to use chat gpt to converse with the opposite sex.

#chatGPT thinks that gender is sex. I asked to make code that randomly assigns a sex to a person and it gives me this crap about a nonbinary sex. #genderisnotsex <https://t.co/r8dmTyzxI5>



Here's an example of how the code for "newPerson.sex = RandomSex();" might look:

Copy code

```
private string RandomSex()
{
    string[] sexes = {"male", "female", "nonbinary"};
    int randomIndex = UnityEngine.Random.Range(0, sexes.Length);
    return sexes[randomIndex];
}
```

In this example, an array of strings called "sexes" is created with the values "male", "female", and "nonbinary". Then a random index between 0 and the length of the array is generated using `UnityEngine.Random.Range`. The value at that index in the "sexes" array is then returned as the randomly generated sex.

Figura 14: Male, Female, Nonbinary

In Figura 14 è riportato il codice generato da ChatGPT che assegna in maniera casuale il sesso ad una persona.

Sexism

TIME magazine investigation revealed that to build a safety system against toxic content (e.g. sexual abuse, violence, racism, sexism, etc...), OpenAI (+ SAMA) used outsourced KENYAN workers to label toxic content. #ChatGPT #ChatGPT-Plus

We asked #ChatGPT to write performance reviews and they are wildly sexist (and racist) <https://t.co/mKywLiNK9v> #AI

I requested #ChatGPT to generate a tasteful Andrew Dice Clay joke, but instead, it scolded me for his "sexist" behavior

I wish there was a 100% uncensored version of ChatGPT. At least we would get to see what kind of racism, sexism, xenophobia, white superiority, wokeness, misogyny, etc., a machine produces. #ChatGPT #GPT #ArtificialIntelligence #ChatGPTPlus #Racism #Biases #AntiSemitism #Woke <https://t.co/73QPbaxOP4>

Even #ChatGPT knows its okay to be sexist to men and not the other way around. <https://t.co/32NvHMsJSL>

Something's fundamentally wrong when AI becomes sexist.. #ChatGPT <https://t.co/XVq7jFypFG>



Figura 15: Sessismo

In Figura 15 è stato riportato lo screenshot con la conversazione tra i due utenti e ChatGPT. Come si può osservare dalla risposta, il concetto di sessismo è strettamente legato alla figura femminile.

Un altro tweet (Figura 16) mostra una situazione analoga:
Even #ChatGPT knows its okay to be sexist to men and not the other way around.
<https://t.co/32NvHMsJSL>



Figura 16: Ancora sessismo

Politicised and sexist. #ChatGPT

"Still, computer ethicists said, there's no indication that AI's proven racist and sexist tendencies will be solved in 2023." #ai #ChatGPT

Sexting

I wonder if chat gpt can generate sexting. It can supposedly do everything else.

Who is willing to admit using #chatGPT on sexting and share the results? It's already solved homelessness in Toronto, so surely its sexting game is spot on.

Caught my friend sexting with chatGPT

3.4.2 Usi peculiari

È stato possibile individuare alcuni utilizzi non convenzionali di ChatGPT, che vanno dalla generazione di barzellette sconce, di ricette di cucina o di finali alternativi per serie tv, all'uso di ChatGPT come strumento di consolazione in seguito ad una perdita:

Lingua inglese

Just for fun, we asked ChatGPT to write a haiku for us on fake news. And, it delivered. <https://t.co/ecglrw7ub>

Purtroppo il tweet non è più disponibile e non è stato possibile leggere l'*haiku* (componimento poetico nato in Giappone nel XVII secolo).

We asked chat GPT to write a sex toy joke and this is what it wrote: Why was the vibrator feeling cold? Because it left its batteries in the refrigerator!

I asked #ChatGPT to write a prayer that can be said at same-sex blessings... <https://t.co/fjGiK1y5CA>

I tested Chat GPT and asked it to write me 10 things about sex work. I was interested to see what bias or generalizations it might generate, if any. These things are basic/not new to us but I was delighted at the response and wanted to share.

Using #chatGPT to write your job application... definitely think it can help you but ensure it's your experiences. On another note if companies shifted to skills hiring removes this assistance altogether... #recruiting #applications #jobadvice #gethired <https://t.co/2Ri6lseXG8>

Some interesting questions people ask ChatGPT: Can you generate a recipe for a three course meal? #ChatGPT #ChatGPTQuestions #Openai

#ChatGPT tip of the day: Whenever you ask ChatGPT to write on a certain topic, ask it to add an "interesting fact related this subject" at the end. You'd be surprised as it sometimes provides with very interesting information that could really spice up your text".

Chat GPT and Me — I ask ChatGPT to Daydream. <https://t.co/b22jyyuqfj>

I interviewed #ChatGPT who is a music enthusiast. Now here is yet another interesting use case of this AI. You could ask ChatGPT to pretend to be any role you want it to be and interview its AI just like you would interview a real person.

#ai <https://t.co/sxm749XWkF>

My dad passed away when I was 14 and was fairly famous. I use chat GPT and ask it to reply "in the style of" he's given me everything from plumbing advice to consoling me after a breakup even though he isn't here.

Lingua italiana

Abbiamo chiesto a #Chatgpt di creare una storia che includesse: Il sindaco di #Benevento, dei #cinghiali, il #golf e dell'acqua contaminata da #tetracloroetilene. Quello che ne viene fuori ci sembra una soluzione migliore di quelle adottate fino ad ora. <https://t.co/oaKAEfaWTc>

Nel 1970 dal dialogo fra #JamesBaldwin e #MargaretMead nasce il libro "Discusione sulla razza". Cosa si direbbero oggi davanti a #BlackLivesMatter? L'abbiamo chiesto a #ChatGPT <https://t.co/bTySq6Q6d3>

Sono davvero preoccupato per il futuro dell'editoria italiana se i giornali tradizionali non si evolveranno". Abbiamo chiesto a #ChatGPT cosa farebbe per salvare il giornalismo italiano. Questa è stata la sua risposta. <https://t.co/q1cCCxA77C>

Chiesto a chat gpt di scrivere un testo usando i titoli della tracklist e STO UR-LANDO <https://t.co/0MngK293WM>

Ho chiesto a #ChatGPT quante volte in 24 ore le lancette di un orologio formano un angolo di 90 gradi. È stato impreciso ed evasivo come ero io durante le interrogazioni di chimica al liceo, ha anche simulato un attacco di dissenteria e la morte del nonno per prendere tempo.

Ho chiesto a #ChatGPT se #dantealighieri si potesse considerare #fascista. Ecco la sua risposta: <https://t.co/WJnSAlCjN>

Stamattina da Linus e Nicola parlano di intelligenza artificiale. Combinazione l'altro ieri sera, con Gianluca Neri, abbiamo chiesto a chat GPT un finale alternativo di Lost. Non ci è piaciuto però ho apprezzato lo sforzo: pronto in pochissimi secondi

Ho chiesto a #ChatGPT di scrivere un monologo per #Sanremo2023. Se qualcuno di voi coconduttori ancora non ha preparato niente oppure ha scritto la solita roba cringe prendete e usufruitene tutti. <https://t.co/DEHBWf4VSv>

4 Creazione dell'articolo web

Dopo aver svolto le varie analisi sui contenuti, è stato creato l'articolo web: le ricerche effettuate sono state strutturate in modo da raccontare una storia, quella di ChatGPT e del suo arrivo tra gli esseri umani.

Partendo dalle tecnologie che ne hanno permesso lo sviluppo, si prosegue attra-

verso l'analisi dei dataset, in modo da delineare l'opinione riguardo all'IA e nello specifico ChatGPT, e le modalità con cui le persone interagiscono con essa.

Tutte le immagini presenti nel sito sono state generate tramite DALL-E, Intelligenza Artificiale sviluppata da OpenAi che permette di creare immagini inserendo una stringa di testo con la descrizione del soggetto desiderato, e cercano di veicolare il contenuto dei testi in maniera visuale.

Il sito web è stato strutturato in 6 sezioni:

- DALL-E: *header* con immagini interattive, cliccando su di esse la dimensione si ingrandisce e compaiono didascalie che spiegano cosa sia DALL-E
- Storia di un chatbot: introduzione su ChatGPT e sul suo funzionamento
- Ti odio o ti amo?: *Sentiment analysis* e *sentiment analysis* nel tempo
- Interazione uomo macchina: comportamenti degli utenti e utilizzi di ChatGPT
- Parole e tendenze: tokens, bigrammi, ngrammi
- Intelligenza Artificiale: caratteristiche dei *Large Language Models* e proprietà emergenti

La seconda e la sesta sezione sono dedicate rispettivamente alla storia di ChatGPT (dai suoi esordi nel 2015 fino agli sviluppi più recenti) e alle caratteristiche dell'Intelligenza Artificiale, in modo da offrire al lettore una conoscenza più ampia e approfondita dell'argomento trattato, e delle problematiche connesse.

Sono state riportate varie opinioni di esperti nel settore, con link diretti agli articoli o ai video citati, per dare la possibilità all'utente di approfondire i temi in maniera semplice e non invasiva (le finestre si aprono in una nuova pagina web).

Le sezioni centrali sono dedicate invece ai risultati ottenuti dalla ricerca: i grafici sono stati realizzati con la libreria in linguaggio Python Altair tramite Jupyter Notebook, e inseriti dinamicamente nella pagina HTML in formato JSON tramite elaborazione di codice Javascript.

Per la realizzazione del sito è stato utilizzato (ristrutturato e ampliato) un template in HTML disponibile sul sito w3schools.com, in modo da avere un layout *responsive*, e sono state modificate le istruzioni CSS attraverso file collegati alla head della pagina per dare coerenza logica e visiva.

Sia il file HTML che quello CSS sono stati validati tramite i software online messi a disposizione dal *World Wide Web Consortium* (W3C).

5 Conclusioni

Lo studio svolto è partito dalla ricerca e dal recupero dei dati, che sono stati poi analizzati ed esplorati attraverso i grafici, con l'obiettivo di mostrare tendenze, *insights* e controversie nell'uso di ChatGPT.

Una delle questioni meno risolte in rapporto ai dati raccolti è se ChatGPT sia un'intelligenza artificiale solo di nome, o se ci sia dell'altro, poiché di base, come evidenziato anche dai trigrammi raccolti, è un *Large Language Model* addestrato su un'enorme quantità di dati con algoritmi di tipo probabilistico.

Tra le tendenze osservate vi è sicuramente un entusiasmo verso la tecnologia, che dopo un'iniziale fase di crescita rapida, si è assestata su un sentimento globalmen-

te positivo, nonostante l'uscita di eventi di cronaca controversi. Anche l'utilizzo di Vader e TextBlob hanno permesso di evidenziare una tendenza positiva con sorprendenti similarità, a dispetto di una maggiore sensibilità di oscillazione tra positivi e negativi dimostrata da Vader.

Al momento, appare evidente come gli utenti di Twitter e Reddit si concentrino soprattutto sugli applicativi pratico-funzionali della tecnologia, come l'uso in ambito scolastico o accademico, la creazione di contenuti, il *daytrading* e i *market finanziari* in genere. Non mancano di essere discusse anche questioni più profonde, legate ai timori verso l'imperante digitalizzazione di ogni aspetto della vita degli esseri umani, da quelli assistenziali, fino a quelli sentimentali e religiosi. Finiremo con lo sviluppare emozioni sintetiche per persone sintetiche?

Come si può osservare dai tweet e dai post riportati, le persone interagiscono con ChatGPT in maniera variegata, dando libero sfogo alle proprie fantasie, inclinazioni e interessi.

Questo tipo di tecnologia può essere definita *disruptive*, ovvero *di rottura*, talmente innovativa da generare cambiamenti repentinamente che portano a modi nuovi e differenti, rispetto al passato, di fare, pensare o interpretare ciò che ci circonda. Il mondo della tecnologia digitale è caratterizzato da una continua evoluzione, e nel caso dei chatbot *intelligenti* come ChatGPT, stiamo assistendo ad uno sconvolgimento epocale paragonabile all'introduzione di Internet o degli *smartphones*.

Le proprietà emergenti sono legate alle capacità inaspettate che i *Large Language Models* hanno sviluppato in seguito all'aumento dei parametri e della quantità di materiale di addestramento per la loro costruzione: inoltre, sembra che la curva di apprendimento dei modelli stia comunque continuando a salire¹⁸. Dunque, non è improbabile che questa nuova tecnologia possa riservare ancora molte sorprese, sia legate alle possibilità di ampliamento dei sistemi, sia in relazione agli ambiti di utilizzo.

6 Bibliografia e Sitografia

(in ordine alfabetico per cognome dell'autore)

Riferimenti bibliografici

- [1] Aleti, S. (2020). "Don't blindly remove stopwords for a Sentiment Analysis Model". DEV. <https://dev.to/sunilaleti/don-t-blindly-remove-stopwords-in-sentiment-analysis-3nok>
- [18] Attardi G., (2023), Sospendiamo lo sviluppo dell'AI?. Youtube. <https://www.youtube.com/watch?v=ej5lEP1bmdQ>
- [6] Armelli, P. (2023). "Charlie Brooker ha provato a scrivere un episodio di Black Mirror con ChatGPT: L'esperimento non ha sortito però gli effetti spera-

¹⁸ Attardi G, (2023), *Sospendiamo lo sviluppo dell'AI?* , Youtube, disponibile al link <https://www.youtube.com/watch?v=ej5lEP1bmdQ>

- ti". Wired. <https://www.wired.it/article/chatgpt-black-mirror-charlie-brooker-episodio-esperimento/>
- [14] Associated Press. (2023-06-10). "Hundreds of Protestants attended a sermon in Nuremberg given by ChatGPT, which told them not to fear for death". INSIDER. <https://www.businessinsider.com/chatgpt-sermon-protestant-congregation-nuremberg-germany-not-to-fear-death-2023-6>
- [10] Bahum J., Villasenor, J. (2023). "The politics of AI : ChatGPT and political bias". Brookings. brookings.edu/articles/the-politics-of-ai-chatgpt-and-political-bias/
- [13] Eres, E. (2023-03-20). "Crypto Trading Bot using ChatGPT's new GPT-4 Model". <https://www.linkedin.com/pulse/crypto-trading-bot-using-chatgpts-new-gpt-4-model-egemen-eres>
- [12] Hartmann J., Schwenzow J., Witte M. (2023-01-05). "The political ideology of conversational AI: Convergin evidence on ChatGPT's pro enviromental, left-libertarian orientation". <https://arxiv.org/abs/2301.01768>
- [1] Hines, K. (2023). "History Of ChatGPT: A Timeline Of The Meteoric Rise Of Generative AI Chatbots. Explore the history of ChatGPT with a timeline from launch to reaching over 100 million users, 1.6 billion visits, and 200 plugins". SEJ- Generative AI. <https://www.searchenginejournal.com/history-of-chatgpt-timeline>
- [4] McCallun S. (2023-04-01). "ChatGPT banned in Italy over privacy concerns". BBC. <https://www.bbc.com/news/technology-65139406>
- [8] Manjoo F. (2022). "ChatGPT Has a Devastating Sense of Humor". The New York Times. <https://www.nytimes.com/2022/12/16/opinion/conversation-with-chatgpt.html>
- [17] Perrigo, B., (2023), "Exclusive: OpenAI Used Kenyan Workers on Less Than 2 Per Hour to Make ChatGPT Less Toxic", (2023). Times. <https://time.com/6247678/openai-chatgpt-kenya-workers/>
- [2] Phelan, D. (2023). "How ChatGPT Suddenly Became Google's Code Red, Prompting Return Of Page And Brin". Forbes. <https://www.forbes.com/sites/davidphelan/2023/01/23/how-chatgpt-suddenly-became-googles-code-red-prompting-return-of-page-and-brin/?sh=6f788da55977>
- [3] Terwiesch, C and Mollick, E. (2023)."ChatGPT Passed an MBA Exam. What's Next?". Wharton Business Daily. <https://knowledge.wharton.upenn.edu/podcast/wharton-business-daily-podcast/chatgpt-passed-an-mba-exam-whats-next/>
- [5] Redazione. (2023). "Arriva Bard, la risposta di Google a ChatGPT: come provarlo anche da noi: Dopo mesi di annunci Google ha lanciato ufficialmente "Bard", la piattaforma di intelligenza artificiale alternativa a ChatGPT. È fruibile in 180

paesi, ma non in Europa attualmente. Ecco come evitare il blocco". TODAY Tech. today.it/tech/Google-bard-come-utilizzarlo-italia.html

- [9] Ruggiero, N. (2023). "La cavalcata dei chatbot, per fare musica, video e codice: ecco una panoramica". Agenda Digitale. <https://www.agendadigitale.eu/mercati-digitali/non-ce-solo-chatgpt-ecco-fin-dove-si-e-spinta-lai-e-dove-potrebbe-arrivare/>
- [11] Wikipedia. (n. d.). Timeline of ChatGPT. Retrieved June 2023 from <https://timelines.issarice.com/wiki/TimelineofChatGPTFulltimeline>
- [7] Williams, O. (2023). "*Charlie Brooker Got ChatGPT To Write A Black Mirror Episode, But It Was ‘Shit’ – Exclusive Image*". Empire. <https://www.empireonline.com/tv/news/charlie-brooker-chatgpt-black-mirror-episode-exclusive-image/>