

Informatica Umanistica e Cultura Digitale: il blog dell' AIUCD.

POSTS

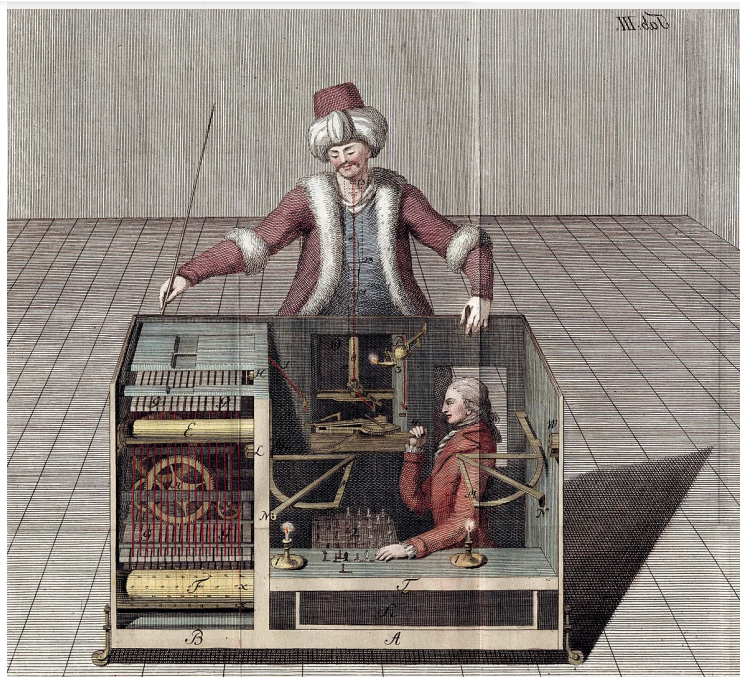
 4

Automi di ieri e sistemi di Intelligenza Artificiale di oggi: umani, troppo umani

DI MAURIZIOLANA · PUBBLICATO 5 LUGLIO 2023 · AGGIORNATO 5 LUGLIO 2023

Il tema di questo articolo di Maurizio Lana è il lavoro sottopagato e precario di annotazione manuale dei contenuti utilizzato da aziende produttrici di sistemi di intelligenza artificiale, direttamente o indirettamente, per la creazione di dataset per il training dei sistemi e per annotare i contenuti da filtrare.

Nel 1769 Wolfgang von Kempelen (1734-1804) inventò per il diletto di Maria Teresa d'Austria e della sua corte un «Turco meccanico»: un automa che giocava a scacchi, che in realtà era manovrato al suo interno da un giocatore umano (Wikipedia, «Il Turco», https://it.wikipedia.org/wiki/Il_Turco).



Disegno del "Turco meccanico", di Joseph Racknitz 1789 – Humboldt University

Library. Tratto da Wikipedia.

Amazon nel 2005 ha iniziato a vendere un servizio che aveva chiamato «mechanical turk» (MTurk), descritto in questi termini: «a crowdsourcing marketplace that makes it easier for individuals and businesses to outsource their processes and jobs to a distributed workforce who can perform these tasks virtually ... MTurk enables companies to harness the collective intelligence, skills, and insights from a global workforce to streamline business processes, augment data collection and analysis, and accelerate machine learning development» (<https://www.mturk.com/>). «Collective intelligence» richiama il titolo di un famoso saggio di Pierre Lévy (Lévy, Pierre. *Collective Intelligence: Mankind's Emerging World in Cyberspace*. Cambridge, Mass: Perseus Books, 1997 / Lévy, Pierre. *L'intelligenza collettiva: per un'antropologia del cyberspazio*. Milano: Feltrinelli, 2002), ma la «distributed workforce who can perform these tasks virtually» è una folla di lavoratori anonimi di cui chi si serve non sa quanto vengono pagati, se vengono rispettati i loro diritti, o se sono sfruttati, come quando comperiamo per pochi euro (uh, bella, costa davvero poco!) delle polo di cotone di cui non sappiamo in quale fabbrica bangla sia stata prodotta da donne e bambini che vi consumano la vita.

Oggi al tempo dei social la «distributed workforce» è formata da «giovani uomini e donne provenienti da paesi di tutta l'Africa [che] siedono a scrivanie incollate ai monitor dei computer, dove devono guardare video di omicidi, stupri, suicidi e abusi sessuali su minori» perché filtrano e taggano appropriatamente i contenuti che vedono. Lavorano per **Sama**, un'azienda di outsourcing di "AI etica", che vende i suoi servizi a Google,

Microsoft, Walmart, e Facebook ([Perrigo, Billy. «Inside Facebook's African Sweatshop». Time, 17 febbraio 2022](#)). Fanno questo lavoro che al pubblico viene descritto come filtri che individuano e filtrano i contenuti non appropriati. Il pubblico crede che ad operare sia un automa, in realtà chi opera sono esseri umani.

Lo stesso fenomeno si ripete con i sistemi di Intelligenza Artificiale: GPT-3 aveva mostrato capacità di costruire frasi ma era incline a commenti violenti, sessisti e razzisti perché di questi erano pieni i dataset tirati su da Internet. Occorreva che questi contenuti venissero taggati come tali, perché GPT li potesse riconoscere e trattare in modo appropriato. Per far questo OpenAI ha adottato lo stesso metodo di Facebook: far valutare e taggare da soggetti umani un set sufficientemente ampio di contenuti deprecabili (abusi sessuali su minori, bestialità, omicidio, suicidio, tortura, autolesionismo, incesto), e usarlo per addestrare GPT a riconoscere tali contenuti. Sama, l'azienda già menzionata qui sopra, ha incaricato del lavoro persone assunte in Kenya che venivano pagate tra 1,32 and 2 dollari / ora. È quanto risulta da un'indagine svolta anche questa volta da Billy Perrigo e pubblicata da Time ([Perrigo, Billy. «OpenAI Used Kenyan Workers on Less Than \\$2 Per Hour to Make ChatGPT Less Toxic». Time, 18 gennaio 2023](#))

Ma il principio base, il nocciolo duro, in realtà era già noto: gli esiti scelti dagli utenti concorrono a raffinare i risultati che vengono mostrati da Google Search.

Globalmente sull'incidenza del nostro uso dei sistemi di Intelligenza Artificiale per raffinarne il funzionamento è uscito un articolo recente su MIT Technology Review ([Heikkilä, Melissa. «We are all AI's free data workers | MIT Technology Review». MIT Technology review, 13 giugno 2023](#)). L'autrice, Melissa Heikkilä, spiega che per far sì che le i sistemi conversazionali di Intelligenza Artificiale sembrino intelligenti e producano meno stupidaggini tossiche viene utilizzata una tecnica chiamata "apprendimento rinforzato dal feedback umano" (reinforcement learning from human feedback), che utilizza i suggerimenti delle persone per migliorare le risposte del modello.

Sul medesimo argomento, una prospettiva ed esposizione più tecnica si può dire che l'approccio «human in-the-loop» mira a formare un modello di previsione accurato con costi minimi, integrando la conoscenza e l'esperienza umana; gli esseri umani possono fornire dati di addestramento per le applicazioni di apprendimento automatico e svolgere direttamente compiti difficili per i computer nella pipeline con l'aiuto di approcci basati sulle macchine ([Wu, Xingjiao, Luwei Xiao, Yixuan Sun, Junhang Zhang, Tianlong Ma, e Liang He. «A Survey of Human-in-the-](#)

[Loop for Machine Learning». Future Generation Computer Systems 135 \(ottobre 2022\): 364–81](#)). Una situazione particolarmente difficile da gestire con i sistemi di Intelligenza Artificiale è quella degli argomenti su cui sono disponibili poche fonti valide, casi nei quali i sistemi di Intelligenza Artificiale danno luogo alla costruzione di «spurious patterns», schemi di ragionamento difettosi. Anche per queste specifiche situazioni si pensa che una soluzione possa essere l'intervento di annotatori umani ([Lu, Jinghui, Linyi Yang, Brian Namee, e Yue Zhang. «A Rationale-Centric Framework for Human-in-the-loop Machine Learning». In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics \(Volume 1: Long Papers\), 6986–96. Dublin, Ireland: Association for Computational Linguistics, 2022](#)).

Concludendo questo sintetico percorso, condivido quanto afferma, Adrienne Williams: «l'idea di macchine superintelligenti dotate di un proprio potere decisionale ci distrae dai rischi reali per le vite umane che lo sviluppo e la diffusione dei sistemi di IA comportano. Mentre l'opinione pubblica viene distratta dallo spettro di macchine senzienti inesistenti, un esercito di lavoratori precarizzati sta dietro ai presunti risultati dei sistemi di intelligenza artificiale di oggi» ([Williams, Adrienne. «The Exploited Labor Behind Artificial Intelligence». Noema, 13 ottobre 2022](#)).

Riferimenti

Heikkilä, Melissa. «We are all AI's free data workers | MIT Technology Review». MIT Technology review, 13 giugno 2023. <https://www.technologyreview.com/2023/06/13/1074560/we-are-all-ais-free-data-workers/>

Lévy, Pierre. *Collective Intelligence: Mankind's Emerging World in Cyberspace*. Cambridge, Mass: Perseus Books, 1997 / Lévy, Pierre. *L'intelligenza collettiva: per un'antropologia del cyberspazio*. Milano: Feltrinelli, 2002

Lu, Jinghui, Linyi Yang, Brian Namee, e Yue Zhang. «A Rationale-Centric Framework for Human-in-the-loop Machine Learning». In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 6986–96. Dublin, Ireland: Association for Computational Linguistics, 2022. <https://doi.org/10.18653/v1/2022.acl-long.481>

Perrigo, Billy. «Inside Facebook's African Sweatshop». Time, 17 febbraio 2022. <https://time.com/6147458/facebook-africa-content-moderation-employee-treatment/>

Perrigo, Billy. «OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic». Time, 18 gennaio 2023.

<https://time.com/6247678/openai-chatgpt-kenya-workers/>

Wikipedia, «Il Turco», https://it.wikipedia.org/wiki/Il_Turco

Williams, Adrienne. «The Exploited Labor Behind Artificial Intelligence». Noema, 13 ottobre 2022.

<https://www.noemamag.com/the-exploited-labor-behind-artificial-intelligence>

Wu, Xingjiao, Luwei Xiao, Yixuan Sun, Junhang Zhang, Tianlong Ma, e Liang He. «A Survey of Human-in-the-Loop for Machine Learning». Future Generation Computer Systems 135 (ottobre 2022): 364–81. <https://doi.org/10.1016/j.future.2022.05.014>



Tag: [annotazione](#) [artificial intelligence](#) [intelligenza artificiale](#)

4 RISPOSTE

[Commenti](#) 4 [Pingback](#) 0

Maurizio Lana 31 Luglio 2023 alle 21:21

Billy Perrigo continua a seguire il tema dei lavoratori che mettono a punto i sistemi di IA.

Il 27 luglio 2023 ha pubblicato su Time un articolo su un'azienda indiana che fornisce servizi per lo sviluppo di sistemi IA capaci di lavorare con le lingue di quel paese e che pare operare in modo corretto, a differenza di quanto avvenuto nel recente passato in Africa:

«Al momento, le IA all'avanguardia – ad esempio i modelli linguistici di grandi dimensioni come ChatGPT – funzionano meglio in lingue come l'inglese, dove i dati testuali e audio sono abbondanti online. Funzionano molto meno bene in lingue come il Kannada che, pur essendo parlato da milioni di persone, è poco diffuso su Internet. (Wikipedia ha 6 milioni di articoli in inglese, per esempio, ma solo 30.000 in Kannada). Quando funzionano, le IA in queste lingue "meno dotate" possono essere distorte, ad esempio assumendo regolarmente che i medici siano uomini e le infermiere donne, e possono faticare a comprendere i dialetti locali. Per creare un'IA efficace in lingua inglese, è sufficiente raccogliere i dati da dove si sono già accumulati. Ma per lingue come il Kannada, è necessario andare a cercarne altri. Ciò ha creato un'enorme domanda di set di dati – raccolte di dati testuali o vocali nelle lingue parlate da alcune delle popolazioni più povere del mondo. Una parte della domanda

proviene dalle aziende tecnologiche che cercano di sviluppare i loro strumenti di intelligenza artificiale. Un'altra parte consistente proviene dal mondo accademico e dai governi, soprattutto in India, dove l'inglese e l'hindi hanno avuto a lungo una posizione di preminenza in una nazione di circa 1,4 miliardi di persone con 22 lingue ufficiali e almeno altre 780 lingue indigene. Questa crescente domanda significa che centinaia di milioni di indiani hanno improvvisamente il controllo di un bene scarso e di recente valore: la loro lingua madre.

...

Nei villaggi di Alahalli e Chilukavadi, una startup indiana sta testando un nuovo modello. Chandrika lavora per Karya, un'organizzazione non profit lanciata nel 2021 a Bengaluru (ex Bangalore) che si autodefinisce "la prima società di dati etici al mondo". Come i suoi concorrenti, vende dati alle grandi aziende tecnologiche e ad altri clienti al prezzo di mercato. Ma invece di trattenere gran parte del denaro come profitto, copre i suoi costi e convoglia il resto verso i poveri delle campagne indiane.»

[Rispondi](#)

maurizio lana  17 Luglio 2023 alle 22:29

sul tema della forza-lavoro umana che concorre a mettere a punto i sistemi di IA e a filtrarne i contenuti sono usciti di recente: Alba, Davey. «Google's AI Chatbot Is Trained by Humans Who Say They're Overworked, Underpaid and Frustrated». Bloomberg.Com, 12 luglio 2023. <https://www.bloomberg.com/news/articles/2023-07-12/google-s-ai-chatbot-is-trained-by-humans-who-say-they-re-overworked-underpaid-and-frustrated>. Deck, Andrew. «The Workers at the Frontlines of the AI Revolution». Rest of World, 11 luglio 2023. <https://restofworld.org/2023/ai-revolution-outsourced-workers/>. Linebaugh, Kate. «The Hidden Workforce That Helped Filter Violence and Abuse out of ChatGPT». WSJ, 11 luglio 2023. <https://www.wsj.com/podcasts/the-journal/the-hidden-workforce-that-helped-filter-violence-and-abuse-out-of-chatgpt/ffc2427f-bdd8-47b7-9a4b-27e7267cf413>.

[Rispondi](#)

Maurizio Lana  11 Luglio 2023 alle 21:29

la situazione descritta per il Kenia è in evoluzione: "Huge ruling in Kenyan court threatens global model of outsourced content moderation – and says that Facebook is the "true employer" of its key safety workers" (<https://www.foxglove.org.uk/2023/06/06/kenyan-court-ruling-outsourced-content-moderation-facebook/>) "Il giudice Byram Ongaya ha stabilito che Facebook è il "vero datore di lavoro" dei suoi moderatori di contenuti in Kenya, e non Sama, la sua società di outsourcing. Ciò significa che Facebook è legalmente responsabile di questi lavoratori. La sentenza è – crediamo – la prima al mondo di questo tipo. Ed è una sentenza con potenziali conseguenze esplosive. Se Facebook è legalmente responsabile per i suoi moderatori, tutti i problemi a cui sono stati sottoposti: la repressione e lo sfruttamento sindacale, l'angoscia mentale causata dall'esposizione prolungata a contenuti tossici, il licenziamento fittizio illegale che li ha impoveriti e ha minacciato il diritto di molti di

vivere nel Paese, diventano tutti responsabilità di Facebook, agli occhi della legge.”

[Rispondi](#)



Andrea Bolioli 5 Luglio 2023 alle 14:41

Segnalo una reference molto citata del 2010 sul Mechanical Turk di Amazon, tra le tante esistenti:

Paolacci, G., Chandler, J., & Ipeirotis, P. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5(5), 411-419. doi:10.1017/S1930297500002205

<https://www.cambridge.org/core/journals/judgment-and-decision-making/article/running-experiments-on-amazon-mechanical-turk/BBD787F3B4DDB61119CBB215927CA39E>

Il Mechanical Turk di Amazon è stato usato spesso e viene usato da ricercatori NLP per l'annotazione di corpora e da ricercatori di altre discipline per vari esperimenti. Il prezzo per il lavoro veniva definito da chi offriva il lavoro.

Il Reinforcement Learning from Human Feedback (RLHF) parte nel 2019 . Segnalo un articolo divulgativo di blog, che a sua volta cita parecchi articoli:

Lambert, et al., “Illustrating Reinforcement Learning from Human Feedback (RLHF)”, Hugging Face Blog, 2022.

<https://huggingface.co/blog/rlhf>

L'annotazione manuale dei contenuti (testi, immagini, video) è sempre stata indispensabile per il machine learning supervisionato e più in generale per la preparazione di dataset, per il controllo dei risultati, per la correzione di errori. In alcuni casi che conosco personalmente il lavoro non era sotto-pagato. Purtroppo si sono diffuse pratiche diverse, di cui si parla anche in questo articolo di Josh Dzieza su The Verge (20 giugno 2023):

“AI Is a Lot of Work. As the technology becomes ubiquitous, a vast tasker underclass is emerging — and not going anywhere.”

<https://www.theverge.com/features/23764584/ai-artificial-intelligence-data-notation-labor-scale-surge-remotasks-openai-chatbots>

[Rispondi](#)



CC-BY-SA 2011-2023 AIUCD - Associazione per l'Informatica Umanistica e la Cultura Digitale

Powered by WordPress. Theme by Press Customizr.



Un blog proposto da Hypotheses - Vedi questo blog nel catalogo di

OpenEdition - Norme sulla privacy - Segnala un problema

Feed di diffusione - Credits - ISSN 2550-9594