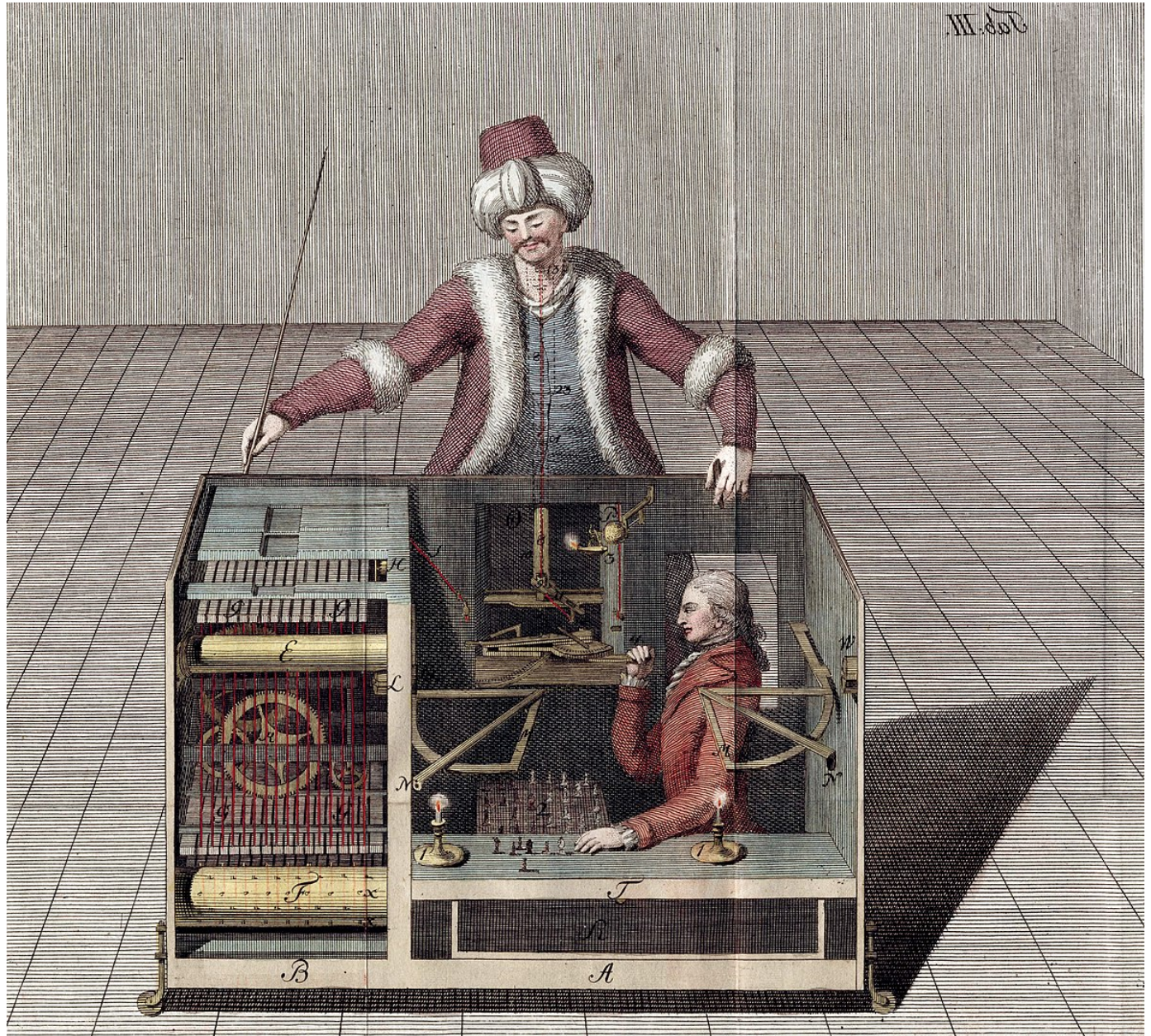


Automi di ieri e sistemi di Intelligenza Artificiale di oggi: umani, troppo umani

mauriziolana

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>

OpenEdition vi suggerisce di citare questo post nel modo seguente:

mauriziolana (5 Luglio 2023). Automati di ieri e sistemi di Intelligenza Artificiale di oggi: umani, troppo umani. *Informatica Umanistica e Cultura Digitale: il blog dell'AIUCD*. Recuperato il 12 Dicembre 2024 da <https://doi.org/10.58079/qcdg>

-
-
-
-