

ANTHROPIC

Claude di Anthropic, Mijansk76, shutterstock.com



DIVERSE TESTATE AFFRONTANO LA QUESTIONE SE ANTHROPIC SIA DAVVERO ETICA, DA QUANDO IL PAPA HA VOLUTO ACCANTO A SÉ LO SCIENZIATO CHRISTOPHER OLAH DI ANTHROPIC PER PRESENTARE L'ENCICLICA MAGNIFICA HUMANITAS. SI CHIEDONO SE DAL PROFILO ETICO CI

SIA DAVVERO UNA DIFFERENZA FRA ANTHROPIC, OPENAI E LE ALTRE BIG TECH, O SE IL PAPA SI SIA SBAGLIATO. PROVO A RISPONDERE.

L'IA in sintesi: una sfida per l'umanità e una svolta epocale

L'intelligenza artificiale è presente ormai in tutti i settori dell'attività umana anche se questo fatto non è evidente

per moltissime persone. Dalla medicina all'economia i vantaggi sono straordinari anche se in molti settori questa svolta epocale crea difficoltà relativamente a professioni che sono costrette a reinventarsi. Ma è accaduto con tutte le innovazioni tecnologiche, dalla scrittura con Gutenberg, alla prima industrializzazione, all'avvento di internet, alla digitalizzazione e ora all'IA. Questa volta la novità, il cambiamento profondo è legato alla gestione di questa tecnologia che è totalmente controllata da un numero esiguo di grandi aziende private. Quindi non ci sono più governi o organismi internazionali che supervisionano, ma la libertà totale ed assoluta.

La questione etica

In tempi recenti sono emersi a livello generale interrogativi su quanto l'IA

potesse rispondere a criteri etici o solo a criteri economici. Alcuni fatti come il suicidio di un quattordicenne della Florida accompagnato e sostenuto dall'IA (vedi l'articolo *Etica e intelligenza artificiale*, pg.34) hanno posto la questione etica a livello del grande pubblico. Ma la questione è molto complicata perché, come affrontato nell'articolo, c'è un problema di addestramento dell'IA che dovrebbe essere rivisto completamente affinché i riferimenti etici diventino irrinunciabili. Se si esclude l'IA cinese controllata da un governo dittatoriale e l'IA con Grok di Elon Musk, personaggio pericoloso privo di qualsiasi preoccupazione etica, per il resto ci sono Meta di FB e Whatsapp, Google con Gemini, OpenAI che gestisce ChatGpt e Anthropic che ha creato l'AI Claude. Di queste quattro l'unica che abbia manifesta-

Leone XIV scommette sull'unica possibilità attuale, andando incontro a questi miliardari tecnologici perché potrebbero diventare gli interlocutori del futuro. Non ha santificato Anthropic portandola al suo tavolo, ma ha affermato la necessità di "disarmare l'AI", cosa fattibile solo collaborando con chi la controlla davvero

to, non solo a parole ma anche con qualche fatto, un interesse alla questione etica è Anthropic come risulta da alcuni testi di Dario Amodè (uno dei fondatori) e da alcuni documenti interni. Poi i fatti. Anthropic ha rinunciato a un contratto da centinaia di milioni col Pentagono perché non voleva che con la sua AI si facessero due cose: controllo della popolazione e costruzione di armi autonome senza controllo umano. E così Anthropic è finita in una lista nera dove il governo americano normalmente mette ditte cinesi e russe ritenute pericolose. OpenAI ha ripreso al suo posto le trattative col governo americano affermando che anche loro vogliono le stesse due condizioni chieste da Anthropic, ma non hanno bisogno che siano scritte in un contratto perché si fidano dell'amministrazione Trump.

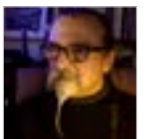
L'intuizione geniale del Papa

Credo che l'intuizione del Papa stia nell'aver capito che la sfida sul piano etico è gigantesca e il rischio di non poter incidere sul controllo dell'AI è altissimo. E deve essersi detto che forse con quelli di Anthropic, che hanno manifestato più degli altri un interesse per la questione etica, si poteva tentare di stabilire un ponte, nel tentativo, di collaborare. Sono certo che il Papa sappia benissimo che Anthropic, ha una idea dell'AI ben diversa dalla sua espressa nell'enciclica, come del resto è emerso dall'intervento di Christopher Olah, alla presentazione di *Magnifica Humanitas*; Leone XIV scommette sull'unica possibilità attuale, andando incontro a questi mi-

liardari tecnologici perché potrebbero diventare gli interlocutori del futuro. Non ha santificato Anthropic portandola al suo tavolo, ma ha affermato la necessità di "disarmare l'AI", cosa fattibile solo collaborando con chi la controlla davvero. Ma la lista delle malefatte di Anthropic è lunga: non è contro la guerra e le armi, continua a far contratti milionari con Palantir, guadagna con Mithos usato ancora dal Pentagono, e consuma quantità smisurate di energia per i suoi server.

Ma allora Anthropic è più etica degli altri?

Lo è solo un po' ma forse abbastanza perché ci sia uno spiraglio di cambiamento. Quelli di Anthropic sono gli unici con cui il Papa può tentare di stabilire una relazione di collaborazione partendo da un interesse comune per alcune questioni etiche, anche se le posizioni ideologiche sono molto distanti. E se poi questo avvenisse solo perché questa immagine etica porta dei vantaggi economici ad Anthropic, come sostengono alcuni, pragmaticamente non mi preoccuperei: nella storia infatti si sono ottenuti dei cambiamenti di natura etica, perché c'erano in gioco valori e profonde convinzioni ideali, talvolta mischiate ad interessi ben poco nobili. Non è una visione machiavellica ma sano realismo pragmatico. ■



testo di
Roby Noris

ETICA, ANTHROPIC E PAPA LEONE

ANTHROPIC È PIÙ ETICA DELLE ALTRE AZIENDE LEADER NEL SETTORE DELL'INTELLIGENZA ARTIFICIALE?