Ci preoccupiamo troppo o troppo poco dell'intelligenza artificiale?

Negli anni cinquanta i computer sono bestioni molto più ingombranti che intelligenti. Eppure, già allora il matematico inglese Alan Turing, parlando in un programma radiofonico della Bbc, propone la visione di un futuro in cui computer intelligenti superano "le nostre deboli forze" e prendono il controllo. Sembra null'altro che la suggestione distopica di una mente visionaria, anche perché, nei decenni successivi, le prestazioni dei computer migliorano sì, ma lentamente e in ambiti ristretti, i finanziamenti sono ridotti e la maggior parte dei centri di ricerca si occupa d'altro.

Poi, nel 2012 succede qualcosa che cambia tutto il resto: con il *deep learning* basato su grandi reti neurali artificiali, rapidissimamente i computer diventano capaci non più solo di eseguire, ma anche di imparare. E diventano capaci perfino di imparare da soli: cioè senza supervisione umana.

È il fisico teorico Stephen Hawking ad affermare, in un'intervista del 2 dicembre 2014, che l'intelligenza artificiale può portare grandi benefici, ma che "the development of full artificial intelligence could spell the end of the human race". Dire che "lo sviluppo dell'intelligenza artificiale può segnare la fine dell'umanità" è impegnativo, specie da parte di una persona la cui interazione con il mondo e la cui stessa sopravvivenza sono da decenni affidate alla tecnologia (Hawking soffre di sla – sclerosi laterale amiotrofica – e usa un sintetizzatore vocale per comunicare). Tuttavia, Hawking sente il bisogno di lanciare un allarme a partire dalla possibilità che l'intelligenza artificiale "decolli da sola, modificandosi e progettando e costruendo autonomamente sistemi sempre più capaci. Gli esseri umani, vincolati dal ritmo lento dell'evoluzione biologica, ne verrebbero travolti".

Rischi e vantaggi

Sempre nel 2014 Eric Horvitz, direttore dei Microsoft research labs, scrive: "Alcuni hanno ipotizzato che un giorno potremmo perdere il controllo dei sistemi di intelligenza artificiale a causa dell'ascesa di superintelligenze che non agiscono secondo i desideri umani, e che sistemi così potenti minaccerebbero l'umanità". Horvitz aggiunge che queste ansie vanno affrontate anche se sono ingiustificate. Segnala diversi ambiti di ricerca, dai rischi per la privacy a quelli per la democrazia, alle implicazioni normative, legali, etiche, alle ricadute sull'economia e sul mercato del lavoro, ai potenziali usi criminali e bellici, all'eventualità più estrema: che ci si avvii "verso una potente 'singolarità' dell'Ia", tale da implicare una perdita di controllo sui processi attuati dalle macchine.

Nel gennaio 2015 è ancora Hawking, con Elon Musk e più di 150 scienziati e pensatori, a firmare una lettera aperta intitolata "Priorità di ricerca per un'intelligenza artificiale solida e benefica". Il testo sottolinea che l'Ia può portare grandi vantaggi all'umanità, ma il documento allegato segnala l'esistenza di rischi in molti ambiti, e afferma che risolvere il problema del controllo dell'Ia è cruciale.

Non possiamo sperare che l'Ia, come il demone del racconto di Stevenson, se ne resti chiusa nella bottiglia per sempre

Torna sullo stesso tema pochi mesi dopo, in una illuminante Ted conference, il direttore del Future of humanity institute di Oxford, Nick Bostrom, che domanda: "Cosa succede quando i nostri computer diventano più

intelligenti di noi?". Sottolinea i limiti biologici (non in termini di lentezza evolutiva, ma di capacità di elaborare dati) del nostro cervello. Dice che in questo secolo gli scienziati potrebbero scatenare la potenza dell'intelligenza artificiale così come nel secolo precedente hanno scatenato la potenza dell'atomo.

Dice che l'Ia può essere l'ultima invenzione a cui l'umanità mette mano: poi, le macchine diventeranno immensamente più brave (e più veloci) di noi. E potrebbero dare forma al futuro secondo le "loro" preferenze, anche perché non esiste un pulsante per "spegnere tutta l'intelligenza artificiale".

In sostanza: non possiamo sperare che l'Ia, come il demone del racconto di Stevenson, se ne resti chiusa nella bottiglia per sempre. Dunque, dobbiamo già adesso fare in modo che, quando uscirà dalla bottiglia (dato che prima o poi ne uscirà) sia amichevole verso di noi. Perché questo accada, dobbiamo costruire qualcosa di ancora più complicato di una superintelligenza artificiale: una superintelligenza artificiale che sia sicura. E il rischio è quello di darsi da fare per raggiungere il primo obiettivo, trascurando il secondo.

Prestazioni imperscrutabili

Dobbiamo notare che già da qualche anno (ad affermarlo nel 2017 è la Mit Technology Review) i processi che stanno alla base di alcune prestazioni dell'Ia appaiono del tutto imperscrutabili agli stessi ricercatori che la progettano.
E dobbiamo notare che, grazie al diffondersi delle strutture di cloud computing e alla crescente disponibilità di ampie banche dati, i costi dell'intelligenza artificiale diminuiscono, gli ambiti di applicazione si espandono e cresce il numero delle imprese che la usano.

Nel 2017, secondo un sondaggio globale del Boston consulting group (Bcg), già un'azienda su cinque incorpora l'Ia in alcune offerte o processi, e un po' meno del 40 per cento delle aziende e oltre il 50 per cento delle grandi aziende ha strategie per l'impiego dell'Ia. Il Bcg cita inoltre alcuni dei campi in cui l'Ia è applicata. Tra questi: marketing e vendite personalizzate, previsioni sull'andamento dei mercati, aiuto nello sviluppo di nuovi prodotti, gestione della produzione, acquisti, magazzini e logistica, revisione di contratti, operazioni assicurative e bancarie, prevenzione sanitaria, diagnostica, ottimizzazione delle prestazioni mediche, sviluppo di nuovi farmaci, automazione dei servizi (con conseguente riduzione dei costi della manodopera).
Tutto ciò succede in assenza di qualsiasi normativa in merito emessa da governi nazionali o da organizzazioni internazionali.
Solo nel 2019 il G7 e il G20 si accordano su una serie di princìpi "per una crescita inclusiva e sostenibile nell'impiego dell'Ia".
La Commissione europea comincia a occuparsi dell'Ia nel 2018 e produce un regolamento generale sulla protezione dei dati. Nell'aprile 2021 elabora una proposta di legge europea (Eu Ai act) comprendente una graduatoria dei rischi, che dovrebbe diventare operativa entro il 2024. È la prima legge al mondo formulata da un importante ente regolatore e riguardante l'Ia. Il testo è stato approvato all'inizio del maggio 2023, e reso più stringente per quanto riguarda, per esempio, il riconoscimento facciale.

Nel settembre 2021 l'Unesco pubblica un report intitolato *La corsa contro il tempo per uno sviluppo più saggio* e il 24 novembre 2021 definisce, in forma di raccomandazione, un insieme di norme etiche per l'Ia: è il primo strumento globale di definizione degli standard su questo tema. Lo adottano 193 paesi ma non ha potere normativo. In seguito, nel maggio 2022, la Commissione europea pubblica una Risoluzione sull'intelligenza artificiale nell'era digitale.

Tra il 2017 e il 2022 gli articoli accademici sui rischi dell'Ia si moltiplicano e si fanno via via più specifici. Le definizioni che i ricercatori danno dell'Ia vanno da "tecnologia dirompente" a "la più grande invenzione dell'umanità".
I rischi evidenziati riguardano la privacy e la democrazia, i diritti umani, i possibili pregiudizi (*bias*) che possono distorcere le valutazioni dell'Ia in materia sanitaria, l'impatto dell'Ia sulla salute pubblica, sul sistema dell'informazione e sulla giustizia, le implicazioni etiche del suo impiego nei campi più diversi, dall'agricoltura di precisione alla gestione del traffico, all'addestramento militare.

Effetto dirompente

Intanto, la Cina non si pone tutti questi problemi e l'intelligenza artificiale nel paese sta già assumendo connotazioni distopiche. Dal 2021 esiste un telegiornale finanziario trasmesso in diretta 24 ore al giorno e interamente gestito (compresi gli avatar dei conduttori) dall'Ia. Un articolo uscito alla fine del 2021 sul South China Morning Post racconta che la procura del popolo di Shanghai Pudong sta sviluppando un "procuratore artificiale" capace di riconoscere otto tipi di reato e di sporgere denuncia: avrebbe un'accuratezza del 97 per cento e potrebbe ridurre il lavoro delle procure. L'intelligenza artificiale viene anche sviluppata per contrastare l'evasione fiscale e, naturalmente, per la sorveglianza. E si progettano *smart city* i cui servizi possono essere gestiti dall'Ia ed erogati su base individuale, compresi i consigli su come vestirsi date le condizioni meteo previste.

In Europa, sempre nel 2021, un sondaggio della le University di Madrid attesta (pagina 10) che il 51 per cento degli europei (e oltre il 60 per cento di quelli che hanno 25-35 anni) apprezzerebbe che una parte dei parlamentari fosse sostituito da un algoritmo.

Ci vuole poco, però, perché la percezione cambi. Un sondaggio del Pew research centre del 17 marzo 2022 dice che a un 18 per cento di statunitensi entusiasti dell'Ia fa riscontro un 45 per cento ugualmente entusiasta e preoccupato e un 37 decisamente preoccupato.

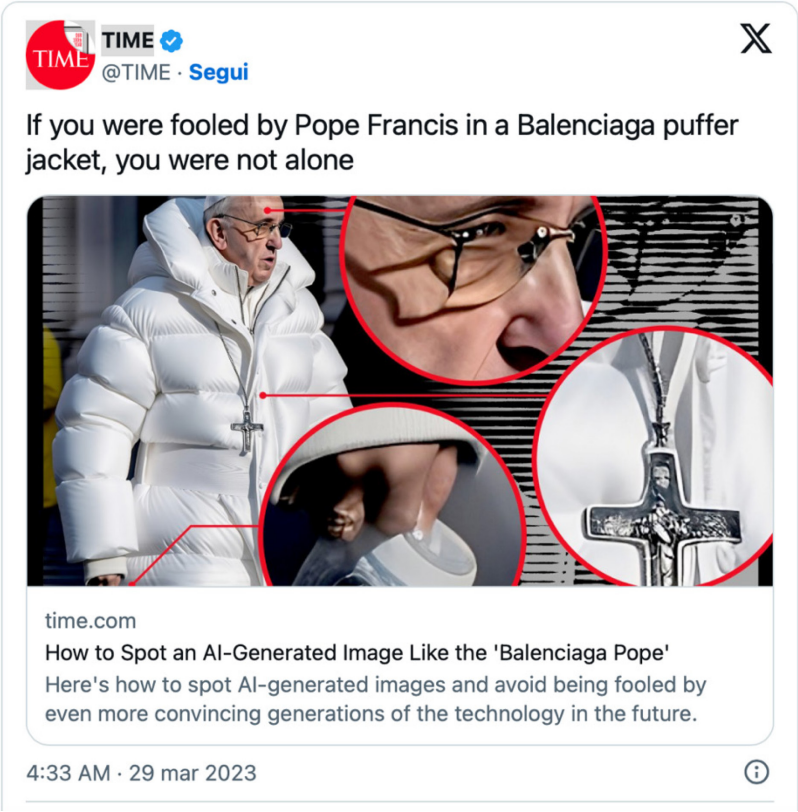
Alla fine del 2022 l'azienda Open Ai lancia Chat Gpt, un'Ia generativa che può produrre testi in risposta a specifiche richieste. È facile da usare, ha prestazioni stupefacenti, è accessibile a chiunque. Nel giro di due mesi si guadagna cento milioni di utenti. E l'interesse collettivo si impenna.

Il 26 marzo 2023, uno studio della Goldman Sachs segnala che l'adozione dell'Ia potrebbe accrescere di sette punti il pil globale annuo, ma potrebbe avere un effetto dirompente (*disruptive* è un aggettivo che ricorre spesso a proposito dell'Ia) sull'occupazione, mettendo a rischio 300 milioni di posti lavoro.
Appena quattro giorni prima, il 22 marzo, il Future of life institute pubblica una lettera aperta dai toni allarmati: "L'Ia avanzata potrebbe rappresentare un profondo cambiamento nella storia della vita sulla Terra e dovrebbe essere pianificata e gestita con cure e risorse adeguate".
È un appello pressante, che chiede ai governi di varare una moratoria di almeno sei mesi su studi che rischiano di sviluppare "menti digitali sempre più potenti che nessuno – nemmeno chi le ha create – può capire, prevedere o controllare", e di istituire autorità capaci di supervisionare, tracciare e certificare i nuovi sistemi di Ia.
A firmare l'appello sono, oltre a moltissimi ricercatori, Elon Musk, fondatore della Tesla, Steve Wozniak, cofondatore dell'Apple, Jaan Tallinn, cofondatore di Skype, Evan Sharp, cofondatore di Pinterest.

Il 29 marzo uno dei padri dell'intelligenza artificiale, Eliezer Yudkowsky, scrive a Time che certo, la moratoria chiesta dal Future of life institute è meglio che niente, ma "sottostima la gravità della situazione": se crei qualcosa di molto, molto più intelligente di te, e quel qualcosa è composto da enormi matrici imperscrutabili, il risultato probabile è che tutti sulla Terra moriranno. È della stessa opinione il ricercatore Paul Christiano.

Anche le persone comuni cominciano a preoccuparsi, ma continuano a farlo in chiave, diciamo così, meno apocalittica: il 4 aprile Forbes pubblica un sondaggio, secondo il quale il 75 per cento del pubblico statunitense è preoccupato per la crescita della disinformazione connessa con la diffusione incontrollata dell'Ia.

Il 28 marzo, un'immagine del papa vestito con un candido piumino Balenciaga fa il giro del mondo. La foto, a parte alcuni piccolissimi dettagli, è realistica: in pochi mesi Midjourney, l'Ia che l'ha creata, è diventata capace di rappresentare le mani umane che fino ad allora la mettevano in difficoltà. "La storia potrebbe considerare il papa in Balenciaga come il primo vero evento di disinformazione virale alimentato dalla tecnologia *deepfake* e un segno dei guai in arrivo", scrive Time. E aggiunge che a oggi non esistono online strumenti gratuiti in grado di discernere tra immagini false o vere.



Il 28 aprile l'Economist pubblica un articolo intitolato "Yuval Noah Harari afferma che l'Ia ha hackerato il sistema operativo dell'intelligenza umana" (qui la traduzione italiana, uscita su Internazionale). In sostanza, Harari dice che quasi tutta la nostra cultura e il nostro sistema di credenze e valori condivisi si basano sul linguaggio. Se l'Ia se ne impossessa e lo manipola, può manipolare le nostre opinioni creando una falsa intimità, può configurarsi come "oracolo onnisciente", può creare religioni. "Per millenni gli esseri umani hanno vissuto all'interno dei sogni di altri esseri umani. Nei prossimi decenni potremmo trovarci a vivere all'interno dei sogni di un'intelligenza aliena". Per questo è necessario e urgente che l'Ia sia regolamentata.

Un procedimento lento

Va però sottolineato che formalizzare una regolamentazione efficace su un fenomeno così esteso e pervasivo come l'Ia non è per niente facile né veloce, e che estendere le regole su scala mondiale, o quasi, può essere ancora più lento e complicato.
Il 5 maggio 2023, il sito di analisi di mercato Yougov afferma che oltre la metà del pubblico globale è preoccupato per la perdita di posti di lavoro.
Ai primi di maggio è lo psicologo cognitivo e informatico Geoffrey Hinton, vincitore del Turing award, a dimettersi da Google e a lanciare, dal campus del Mit un appello a Stati Uniti e Cina perché regolamentino l'Ia, che sta sviluppandosi molto più rapidamente di quanto lui e altri informatici si aspettassero, e potrebbe a breve superare i suoi creatori umani: "Pensavo che ci sarebbero voluti dai 30 ai 50 anni", dice. "Ora penso che sia più probabile che ce ne vorranno da cinque a 20". Attacchi informatici, truffe, disinformazione e sorveglianza di massa sono i quattro maggiori rischi immediati.

Uno dei rischi è che l'intelligenza artificiale faccia molto bene quello che non vorremmo che facesse

Il 13 maggio esce la notizia che la Casa Bianca sta arruolando quattromila hacker per trovare le falle dell'Ia. Si incontreranno a Las Vegas il prossimo agosto.
Il 17 maggio Sam Altman, il fondatore della Open Ai, appena due mesi dopo aver magnificato a Seattle le prestazioni di ChatGpt, in un'audizione davanti alla sottocommissione giustizia del senato statunitense chiede che il governo si attivi in fretta per mitigare i rischi dell'Ai, così come è stato fatto in passato per gli armamenti nucleari: "Se questa tecnologia prende la direzione sbagliata, diventa veramente tossica". Può, per esempio, manipolare le menti e alterare i risultati elettorali.

Il problema di re Mida

In buona sostanza, i rischi dell'Ia si possono ricondurre a quattro categorie.

Il primo, e il più ovvio, è che l'Ia non faccia "bene abbastanza" quello che le vogliamo far fare: che sia affetta da *bias*, che soffra di allucinazioni, che abbia delle vulnerabilità. Questi problemi possono essere risolti a mano a mano che si presentano.

Il secondo rischio è che l'Ia faccia "troppo bene" quello che vogliamo farle fare, eliminando posti di lavoro qualificati in moltissimi ambiti, dal giornalismo all'intero sistema della comunicazione, alla programmazione, all'intrattenimento, ai settori legale, bancario, finanziario, assicurativo, al marketing e all'assistenza-clienti, ai trasporti, alla produzione di immagini e video, cambiando le regole e i processi di interi mercati e rendendo obsoleti i sistemi produttivi di un gran numero di imprese. Chi parla dell'Ia come di un'innovazione dirompente, con ogni probabilità ha in mente questi scenari.

Il terzo rischio (e questo è ancora più dirompente) è che l'Ia faccia molto, molto bene quello che non vorremmo che facesse: esercitare una sorveglianza ubiqua e intrusiva, disinformare e diffondere notizie false e credenze infondate, manipolare le persone attraverso una comunicazione individualizzata e intima, alterare i risultati elettorali, trasformarsi in uno strumento bellico, disabituare gli studenti a produrre pensiero critico.

Il quarto rischio, quello che più, mi pare, sta allarmando i ricercatori, e che forse è meno chiaro ai non addetti ai lavori, è che l'Ia arrivi a fare qualcosa che riusciamo a stento a immaginare, in modi che non saremmo in grado di capire e seguendo criteri che non le abbiamo trasferito e che non condividiamo. È la "singolarità" di cui finora hanno parlato soprattutto i futurologi e gli autori di fantascienza.

E infatti: a metà maggio 2023 il New Yorker scrive che, secondo molti ricercatori, "ci sono buone probabilità che l'attuale tecnologia la si trasformi in intelligenza artificiale generale, o Iag, una forma superiore di Ia in grado di pensare al livello umano sotto molti o quasi tutti gli aspetti. Un gruppo più piccolo sostiene che il potere dell'Iag potrebbe aumentare in modo esponenziale. Se un sistema informatico è in grado di scrivere codici, come già fa ChatGpt, allora potrebbe imparare a migliorarsi più e più volte fino a quando la tecnologia informatica non raggiungerà quella che è nota come 'la singolarità': un punto in cui sfugge al nostro controllo. Nello scenario peggiore immaginato da questi pensatori, la incontrollabili potrebbero infiltrarsi in ogni aspetto delle nostre vite tecnologiche, interrompendo o reindirizzando la nostra infrastruttura, i sistemi finanziari, le comunicazioni e altro ancora".

Questo scenario non è affatto scontato. Ma i recenti progressi dell'Ia l'hanno reso più plausibile, anche perché le grandi aziende stanno già sviluppando algoritmi "generalisti". È il "problema di re Mida"(*King Midas problem*): raggiungere integralmente un obiettivo che appare desiderabile avendone sottostimato le implicazioni e le conseguenze altamente indesiderabili.

Internazionale, Annamaria Testa, esperta di comunicazione, 29 maggio 2023

https://www.internazionale.it/opinione/annamaria-testa/2023/05/29/intelligenza-artificiale-rischi-regole