Istruzione & Formazione News n. 22

Il vaso di Pandora:
l'impossibilità di interrompere la ricerca sull'intelligenza artificiale (2)

Nel corso dei settant'anni della sua esistenza, il campo dell'intelligenza artificiale ha compiuto passi da gigante, passando dai primi, primitivi algoritmi alle moderne reti neurali e al machine learning, sempre più capaci di imitare l'intelligenza umana. I frutti di questa crescita sono visibili ovunque, dai piccoli Roomba che puliscono le nostre case e gli assistenti virtuali come Alexa e Cortana, fino alle macchine a guida autonoma e ai quasi fantascientifici robot che stanno diventando sempre più comuni. Giornali e telegiornali, libri ed internet, conferenze, incontri e dibattiti tra politici: oramai, di IA ne sentiamo parlare tutti i giorni; eppure, quello che si dice non è sempre positivo, e molti di questi interventi sono dedicati a discutere di rischi, pericoli e conseguenze negative legati a questa tecnologia: solo per citarne alcuni tra i più frequenti ed importanti, abbiamo le accuse che l'intelligenza artificiale ruba i posti di lavoro, le critiche su come essa possa essere usata per diffondere fake news, e perfino le discussioni sul rischio di estinzione per la razza umana. E così, alcuni, di fronte ai pericoli dell'intelligenza artificiale, hanno suggerito che la soluzione migliore sia di interrompere completamente la ricerca ed abbandonare i nostri tentativi di creare IA più potenti ed avanzate, sperando di poter tappare il "vaso di Pandora" prima che sia troppo tardi.
Sfortunatamente, proprio come nel mito, chiudere il vaso è inutile: i mali in esso contenuti sono ormai scappati, e ogni tentativo di tapparlo, ossia di fermare la ricerca, è destinato a fallire per due motivi: la difficoltà nel far rispettare un tale divieto, e l'enorme potenziale dell'IA.

Per quanto riguarda il primo, è facile immaginare quanto sarebbe difficile da applicare questo divieto, sia internamente che esternamente: internamente avrebbe gravi conseguenze economiche, costringendo le aziende dedicate allo sviluppo dell'intelligenza artificiale a chiudere o trasferirsi all'estero, portando con se molti esperti e neolaureati; ovviamente, questo porterebbe anche ad un ritardo nello sviluppo tecnologico del paese che avesse adottato il divieto, specialmente considerato quanto vaste sono le possibili applicazioni dell'IA, danneggiando ulteriormente la sua posizione e ponendolo in considerevole svantaggio in confronto alla competizione (politica, economica, tecnologica e militare) estera. Questa è infatti la seconda parte del problema: anche qualora si riuscisse ad applicare questo divieto senza danneggiare eccessivamente la propria economia, sarebbe impossibile garantire che gli altri stati lo rispettino; anche se il veto venisse adottato da un'entità sovranazionale, come l'Unione Europea o l'ONU, anziché da un singolo stato, rimarrebbe comunque impossibile garantire il rispetto da parte di stati che non sono membri, o anche solo da quelli meno disposti ad accettare tali limitazioni. Paradossalmente, questa incapacità di far rispettare un simile divieto potrebbe addirittura essere vista come un motivo a favore del continuare la ricerca, specialmente se si considerano le possibili applicazioni militari dell'IA.

Incidentalmente, nel discutere del primo motivo è emerso, sebbene solo parzialmente, anche il secondo: l'intelligenza artificiale è, potenzialmente, la più grande invenzione della storia dell'umanità. Le sue applicazioni sono innumerevoli, il suo potenziale è enorme, e più la ricerca avanza, più l'intelligenza artificiale diventa potente; in breve, l'IA è una tecnologia rivoluzionaria, applicabile in un'enorme varietà di ambiti e capace di migliorare considerevolmente la nostra qualità di vita – si immagini, come nei migliori romanzi di fantascienza, un futuro dove la maggior

parte dei lavori manuali sono svolti per noi dai robot, in cui le macchine si guidano da sole e le nostre case sono controllate con comandi vocali, mentre supercomputer estremamente avanzati elaborano come risolvere grandi problemi come la fame nel mondo e il riscaldamento globale. Un'immagine forse un po' troppo utopica, che enfatizza troppo i lati positivi senza preoccuparsi del come e quando, ma certamente non troppo lontana da quelle proposte, ad esempio, dal transumanesimo. In ogni caso, ciò è importante per un semplice motivo: fermare lo sviluppo dell'intelligenza artificiale ci priverebbe di questi benefici, proprio mentre l'IA diventa sempre più diffusa e capace; per fare un paragone, sarebbe come se uno stato occidentale avesse vietato la ricerca nel settore informatico durante gli anni dello sviluppo dei computer e dell'internet. Bisogna indubbiamente riconoscere che l'intelligenza artificiale potrà avere degli effetti negativi sulla nostra società, ma questo non vuol dire dimenticare gli enormi benefici che potrebbe portare, come anche il fatto che essa non è certamente la prima tecnologia a combinare svantaggi e benefici: si pensi alle automobili, che inquinano e causano incidenti mortali. Forse i nostri antenati avrebbero rifiutato di usarle, se avessero saputo quali danni avrebbero causato, ma oggi vietare l'uso dell'automobile sarebbe inconcepibile semplicemente perché è troppo utile; similmente, l'intelligenza artificiale, nonostante i suoi pericoli, è troppo potente per essere semplicemente abbandonata. Medicina, matematica e fisica, robotica, ricerca e sviluppo… è forse più difficile trovare un campo che non beneficerebbe in alcun modo dall'avere accesso all'IA. In breve, si può dire che, in questo caso, il gioco vale la candela.

Da un lato, dunque, un veto sullo sviluppo dell'intelligenza artificiale sarebbe assai difficile da far rispettare, e sicuramente dannoso per lo stato che lo adottasse; dall'altro, sembra difficile poter giustificare tale divieto, di fronte agli enormi vantaggi che l'IA potrebbe portare. Perché, allora, vietare la ricerca? Precedentemente, ho menzionato tre rischi: i primi due, la perdita di posti di lavoro e le fake news, sono critiche comunemente avanzate, mentre la terza, il rischio di estinzione, è meno spesso menzionato, ma piuttosto interessante. Vorrei ora esaminare questi tre esempi, e cercare di mostrare come i timori, per quanto fondati, sono anche molto spesso esagerati.

Iniziamo dunque dalle fake news: l'IA generativa, così detta perché capace, appunto, di generare immagini, audio e testi, è andata incontro ad un enorme boom in questi ultimi anni, sia in termini di capacità che di interesse pubblico, risultando così nella nascita di una considerevole quantità di siti e programmi che permettono a chiunque di creare (e diffondere) una varietà di prodotti "generati dall'IA". Per molti versi si tratta di una cosa innocua o, perché no, anche positiva: sebbene la qualità sia spesso discutibile, è certamente un modo per divertirsi, e, soprattutto nel caso dei generatori di testi, può anche essere utile; sfortunatamente, l'intelligenza artificiale è anche spesso utilizzata, volontariamente o meno, per creare fake news. False foto di politici, dichiarazioni che non sono mai state pronunciate, basta solo un prodotto sufficientemente realistico ed un giornalista poco attento per diffondere una storia, e, via via che i programmi diventano più capaci, sarà sempre più difficile distinguere il vero dal falso. Il pericolo è serio: come fidarsi di giornali e telegiornali, quando possono essere così facilmente ingannati da un computer? Curiosamente, qualcosa del genere era già successo all'inizio dell'invasione russa dell'Ucraina: alcuni video che venivano fatti circolare come autentici (e, a volte, sono anche stati mostrati da media importanti) sono risultati essere clip provenienti da un videogioco; lo stesso era già accaduto l'anno precedente (con lo stesso videogioco, peraltro). E tutto ciò senza alcun intervento dell'intelligenza artificiale: solo pura stupidità biologica. E si potrebbero fare molti altri esempi: negli anni 30, in America, un programma radiofonico basato su "La guerra dei mondi" convinse moltissime persone che i marziani stessero davvero invadendo la Terra, mentre "I protocolli dei saggi anziani di Sion", un falso creato dalla polizia segreta zarista, è stato per decenni come prova dell'esistenza di un complotto ebraico per conquistare il mondo; più recentemente, abbiamo il famoso articolo (poco) scientifico di Andrew Wakefield, che i più ricorderanno come l'origine della credenza che i vaccini causino l'autismo.

Insomma, le fake news non sono niente di nuovo, e non è assolutamente necessario usare l'IA per crearle, anzi; certo, l'intelligenza artificiale può creare dei falsi convincenti, e la sua abilità nel manipolare audio e immagini è considerevole, ma il vero problema sembrerebbe essere più nelle nostre capacità di giudicare se qualcosa sia vero o falso che in quelle dell'IA di ingannarci.

E per quanto riguardo l'automazione? Anche qui bisogna innanzitutto ricordare che non stiamo parlando di qualcosa di nuovo e mai visto prima; anzi, l'intera situazione è sorprendentemente simile ad eventi già accaduti nell'Inghilterra della rivoluzione industriale: la diffusione del telaio meccanico, percepito dai lavoratori come una minaccia in quanto capace di rubare loro il lavoro, portò alla nascita del Luddismo, un movimento che si opponeva all'automazione dell'industria tessile, risultando a volte anche nella distruzione delle macchine. Il parallelo non è perfetto, ma è certamente interessante. Detto questo, per comprendere cosa implichi davvero l'automazione portata dall'intelligenza artificiale, e ridimensionare queste accuse, dobbiamo considerare due elementi fondamentali: quali sono i lavori che l'IA ruberebbe agli esseri umani, e quali conseguenze effettive ha il processo di automazione. Per quanto riguarda il primo, la risposta è sorprendentemente semplice: quelli che noi non vogliamo, i lavori difficili, noiosi, faticosi, ripetitivi, sporchi, pericolosi e sottopagati; più che "rubarci" il lavoro, l'intelligenza artificiale cerca di "liberarci" dal (duro) lavoro e renderci la vita più facile. Peraltro, bisogna comunque tenere a mente che ci vorrà del tempo prima che l'IA possa sostituire l'uomo in diverse tipologie di lavoro, specialmente quelle che richiedono flessibilità e problem solving (per fare un esempio, gli autobus a guida autonoma sono ancora lontani), e in molti casi potrebbe essere preferibile mantenere un elemento umano come supervisore o collaboratore; bisogna, inoltre, considerare tutta una serie di altri elementi, come il costo di passare all'IA, che potrebbe, soprattutto all'inizio, risultare un investimento eccessivo per aziende di media e piccola taglia, favorendo così il personale umano. Il secondo elemento ha invece a che fare con i considerevoli benefici dell'automazione, che, tra le altre cose, servirebbe a ridurre il rischio di incidenti sul lavoro, soprattutto in quelli più pericolosi, e favorirebbe la produzione industriale riducendo i costi ed aumentando le capacità produttive, rendendo così i prodotti stessi meno costosi e più accessibili; inoltre, non dobbiamo dimenticare che l'intelligenza artificiale è anche capace di generare nuovi posti di lavoro, come programmatori, ingegneri, data experts e altri ancora, tutti necessari a garantire il perfetto funzionamento dell'IA.

L'ultimo esempio che vorrei considerare è il rischio di estinzione. Come critica è molto peculiare: in apparenza è una chiara esagerazione e poco credibile, e proprio per questo appare abbastanza raramente; si tratta, tuttavia, di un problema che è stato seriamente considerato da alcuni studiosi, e che ha recentemente attirato l'attenzione del pubblico grazie ad un articolo di Eliezer Yudkowski, favorendo la diffusione dei concetti di "IA etica" e "istanziazione malvagia". Di cosa si tratta però? Innanzitutto, bisogna chiarire che l'istanziazione malvagia non ha niente a che vedere con "Matrix" o "Terminator": un'IA ribelle, che decide di voler conquistare il mondo e/o distruggere l'umanità, è altamente improbabile. Quello che è possibile, invece, è che un'IA adotti una soluzione ad un problema che comporta, come effetto collaterale, l'estinzione dell'umanità: per esempio, un'intelligenza artificiale con il compito di massimizzare la produzione di graffette potrebbe decidere di trasformare l'intero pianeta in un'enorme fabbrica di graffette. Possiamo già vedere qualcosa del genere, anche se su una scala molto più ridotta, nelle IA odierne, soluzioni tecnicamente corrette ma per noi inaccettabili, e non è difficile immaginare che lo stesso potrebbe succedere con IA più avanzate e potenti: questa è l'istanziazione malvagia. In questo caso, sembrerebbe che ci siano davvero dei buoni motivi per interrompere la ricerca: un'intelligenza artificiale sufficientemente evoluta, ma incontrollabile, sarebbe estremamente problematica, anche se non necessariamente maliziosa; per continuare con gli esempi provenienti dalla cultura popolare, più simile al malfunzionante HAL che a Skynet, ma non per questo meno pericoloso.

Tuttavia, come abbiamo già visto, un blocco completo incorrerebbe in diversi problemi che lo renderebbero assai poco efficace, mentre accordi, limitazioni, o pause temporanee si limiterebbero a ritardare il sorgere del problema; la soluzione migliore, proposta proprio dagli stessi autori che ci avvertono di questo rischio, sarebbe invece di continuare la ricerca, ma indirizzandola verso lo sviluppo dell'IA "etica", ossia un'intelligenza artificiale che possiede valori etici (o anche soltanto quello che definiremmo "comune buon senso") e che è quindi capace di comprendere quali soluzioni siano accettabili e quali evitare. Per gli autori che sostengono questa idea la ricerca potrebbe essere rallentata o fermata temporaneamente, per garantirci il tempo necessario a creare un'IA etica prima che l'intelligenza artificiale diventi troppo avanzata, ma mai bloccata completamente: non solo questo ci priverebbe della possibilità di studiare contromisure in caso un'IA sufficientemente avanzata venisse davvero realizzata, ma non tiene nemmeno contro dell'altra faccia della medaglia, ossia del fatto che le stesse capacità che temiamo possano portare all'estinzione della razza umana potrebbero anche essere utilizzate a nostro beneficio.

Insomma, sembrerebbe che l'idea di interrompere la ricerca sull'IA non funzioni molto bene: i benefici dell'intelligenza artificiale sono troppo grandi, e gli svantaggi per uno stato che adottasse tale divieto troppo gravi per rendere un simile approccio credibile; allo stesso tempo, molti dei problemi che vengono usati per giustificare questa interruzione sono esagerati o il risultato di una mancanza di comprensione. Un po' paradossalmente, quello che sembrerebbe essere il rischio più grave, ossia l'istanziazione malvagia, sarebbe più facilmente gestito continuando la ricerca per trovare modi per prevenirlo.

Precedentemente ho paragonato l'intelligenza artificiale al famoso "Vaso di Pandora", il mitologico artefatto che, una volta aperto dalla curiosa Pandora, rilascia nel mondo i mali al suo interno rinchiusi, condannando l'umanità a sofferenza e dolore; il mito si conclude con Pandora che riesce finalmente a chiudere il vaso, ma ormai è troppo tardi: tutti i mali sono già scappati, e nel vaso rimane solo una cosa, la speranza. Il parallelo è ovvio: l'intelligenza artificiale è a volte vista come pericolosa, ma, come ho cercato di provare, è ormai troppo tardi per interrompere la ricerca; tuttavia, come nel mito, il vaso non contiene solo mali, ed è su questo che dovremmo concentrarci: perché cercare, inutilmente, di chiudere il vaso, quando al suo interno rimane ancora così tanto potenziale? *(A cura di Manfredi Negro)*

Milano, 1.04.2024