

Dove va l'AI?

Il dibattito sulla Intelligenza Artificiale sta uscendo fuori dai suoi tradizionali confini scientifici e tecnologici per investire il campo delle compatibilità culturali e, in particolare, etiche.

Le perplessità degli studiosi in merito si scontrano con gli interessi delle maggiori aziende informatiche.

Alessandro Ovi

Dal computer HAL (ricordate: una lettera indietro rispetto a IBM) di *2001 Odissea dello spazio* a *Io Robot* (con le 3 leggi di Asimov), l'Intelligenza Artificiale, (AI) era stata sempre vista come una conquista molto lontana, ma comunque straordinaria dell'uomo. Il fatto che a un certo punto gli si rivoltasse contro, era vissuto più come la naturale trama di un film, che non come una situazione da temere realmente.

Non c'era preoccupazione, non c'era paura. Come non c'erano preoccupazione e paura quando Apple ha lanciato Siri, (la segretaria virtuale di iPhone, della quale si poteva al massimo rischiare di innamorarsi, come in *Lei*, anche questo un film), o quando, su scala molto più grande, IBM ha presentato Watson (un sistema di grandi elaboratori e software in grado di risolvere molteplici problemi reali della vita).

E neppure quando Google si è lanciato in una rincorsa quasi "affannosa" ad acquistare tutte le società disponibili sul mercato, impegnate a sviluppare forme, ancora primordiali, ma comunque sempre più avanzate, di AI. L'aggregazione a Google di Jetpac, Vision Factory, Dark Blue Labs, Deep Mind, Boston Robotics non ha impressionato nessuno.

Negli anni Ottanta e Novanta i libri di esperti e di scrittori di fantascienza erano stati positivi nei confronti della AI, anche se con visioni diverse. Alan Turing, il matematico e crittografo inglese che aveva progettato la macchina che violò il codice di guerra nazista Enigma, aveva postulato una imitazione della intelligenza umana.

Ray Kurzweil, invece, ingegnere, futurologo, imprenditore, oggi a Google, pensava più in grande: Autore nel 1999 del libro *The Age of Spiritual Machines*, sostiene che quando un computer diventa capace di trovare, indipendentemente

da precedenti istruzioni, modi propri per raggiungere i suoi obiettivi, è molto probabilmente capace di introspezione e quindi di modificare il suo software e migliorare la sua intelligenza. Comunque la visione era quella positiva di una nuova era di meraviglie in arrivo.

La scossa è venuta quando un visionario come Elon Musk, il 27 ottobre dello scorso anno, mentre era al MIT, ha dichiarato che «L'AI è il più grande rischio per la sopravvivenza dell'umanità». Musk è il fondatore di due società "impossibili", poi divenute di successo: Tesla, per automobili elettriche di lusso, e Space X, per trasporti spaziali a basso costo. Non è uno scienziato della AI, ma si sa che vede lontano e che ha da poco investito assieme a Zuckerberg (Facebook) nella nascita di Deep Mind, fondata da quel genio straordinario che è Hassabis (un suo profilo si legge a p. 46), anch'essa poi acquistata da Google. Si dice che Larry Page (fondatore e CEO di Google) la abbia voluta per lasciarla a lavorare a Londra, lontano dalla Silicon Valley, per fare da guida intellettuale al gregge delle altre acquisizioni, restandone esterna.

Hassabis è un ricercatore della AI, ma anche neuro-scienziato, progettista di *computer games* e maestro di scacchi all'età di otto anni. I suoi mentori sono lontani nel tempo. Turing, che nel 1950 aveva ipotizzato un computer a cui si potesse insegnare come a un bambino, e, ancora più lontano, Charles Babbage (1791-1871), il genio inglese che per primo aveva espresso il concetto del computer programmabile. Mentori che danno profondità al suo pensiero e credibilità alle sue proposte.

Hassabis ha convinto prima Musk e poi Page della sua idea rivoluzionaria, e cioè che nell'uomo è possibile ci siano elementi unici e singolari, ma non nella intelligenza, perché il cervello è, a tutti gli effetti, assimilabile a un computer,

anche nella creatività, nella immaginazione e perfino nella coscienza, tutte riconducibili a programmi software.

Dice Hassabis: «L'AI che vediamo applicata oggi, alla diagnosi, alla terapia medica, alle previsioni del tempo, alle automobili senza pilota, consiste sostanzialmente in istruzioni date al computer che le applica alle diverse situazioni per il cui riconoscimento è stato programmato, e le esegue. La "vera AI" è quella in cui il computer apprende dall'esperienza le relazioni ottimali tra problemi e soluzioni ed è in grado di cambiare autonomamente software e anche hardware per migliorarsi».

È vero che nel 2013 James Barrat aveva scritto *Our Final Invention: Artificial Intelligence and the End of the Human Era* e Stuart Armstrong, *Smarter Than Us: The Rise of Machine Intelligence*. Ma non si era scatenato un grande dibattito.

Il pensiero più preoccupato era quello di Nick Bostrom, un filosofo che dirige l'Istituto per il futuro dell'umanità, che a settembre 2014, nel suo libro *Superintelligence. Paths, Dangers, Strategies*, si è domandato cosa potrà succedere quando le macchine supereranno l'intelligenza dell'uomo.

Gli "agenti artificiali" ci salveranno o ci distruggeranno? Forse è proprio la lettura di Bostrom, calata sulle visioni inquietanti di Hassabis, a fare scattare in Musk la visione della catastrofe in arrivo. Musk aveva poi fatto seguire alle sue dichiarazioni una donazione di 10 milioni di dollari all'Istituto Future of Life, da non confondere con il centro di Bostrom, che dice di lavorare in generale alla mitigazione dei rischi che pendono sulla umanità.

Il mondo degli esperti si è stupito della sua dichiarazione, ma, subito dopo, non a stupirsi, ma a spaventarsi, è stato tutto il resto del mondo. Perché? Perché



a dire esattamente la stessa cosa è stato Steve Hawking che, con la sua voce sintetica e dalla sua sedia a rotelle, ha detto: «L'intelligenza artificiale può essere la più grande conquista dell'uomo. Ma anche l'ultima».

E quando Hawking parla, il mondo ascolta. Può essere d'accordo o no, ma ascolta. Così in pochi giorni l'AI da un sogno è diventato un incubo. I due sono stati subito seguiti da 300 scienziati da tutto il mondo, che hanno sottoscritto un manifesto di allarme contro i rischi per l'umanità impliciti nel diffondersi di forme di AI sempre più avanzate, raccomandando di prendere tutte le cautele necessarie a evitarli.

Oggi si accavallano le riflessioni attorno all'interrogativo se il computer possa pensare.

Nessuno però ritiene che qualcosa di simile a una intelligenza artificiale "generale" già esista.

In realtà non esiste neppure il disegno completo di come arrivarci in tempi "medi" (5 -10 anni).

Il cervello di un bambino può facilmente riconoscere un gatto in una fotografia. Ma le reti neurali artificiali debbono venire istruite facendo loro vedere migliaia di foto di gatti. C'è ancora tanta strada da fare.

Qui entrano in gioco gli scettici, come Brooks, uno dei fondatori di iRobot e Rethink Robotica.

Anche se per un computer il progresso nel riconoscere l'immagine di un gatto è stato enorme, la macchina non ha il senso di cosa sia la "gattità" o cos'altro stia accadendo nella fotografia. Brooks ha scritto di recente su Edge.org. che «la *malevolent AI* non costituisce nulla di cui preoccuparsi, per centinaia di anni almeno».

Una persona che, invece, condivide le preoccupazioni di Bostrom è Stuart J. Russell, professore di informatica presso la University of California, Berkeley. Russell è l'autore, di *Intelligenza Artificiale: un approccio moderno*, che è stato il libro di testo standard di AI per due decenni.

Scrivono Russell: «Poiché Google, Facebook e altre aziende stanno attivamente cercando di creare una macchina intelligente, una delle cose che non dobbiamo fare è andare avanti a tutto vapore senza pensare ai rischi potenziali. Se si vuole una intelligenza illimitata, è meglio capire come allineare i computer con i valori e i bisogni umani».

Bostrom richiama l'idea di un pensatore di nome Eliezer Yudkowsky, che parla di una "volontà estrapolata coerente", cioè il consenso del "meglio di sé" di tutte le persone. Come sarebbe possibile programmare questi valori nelle nostre (potenziali) superintelligenze? Questo è il problema che, secondo Bostrom, i ricercatori dovrebbero affrontare.

Dopo tutto, se svilupperemo una intelligenza artificiale che non condivide i migliori valori umani, vorrà dire che non saremo stati abbastanza intelligenti nel controllare le nostre creazioni. ■

Alessandro Ovi è editore e direttore di MIT Technology Review Italia.