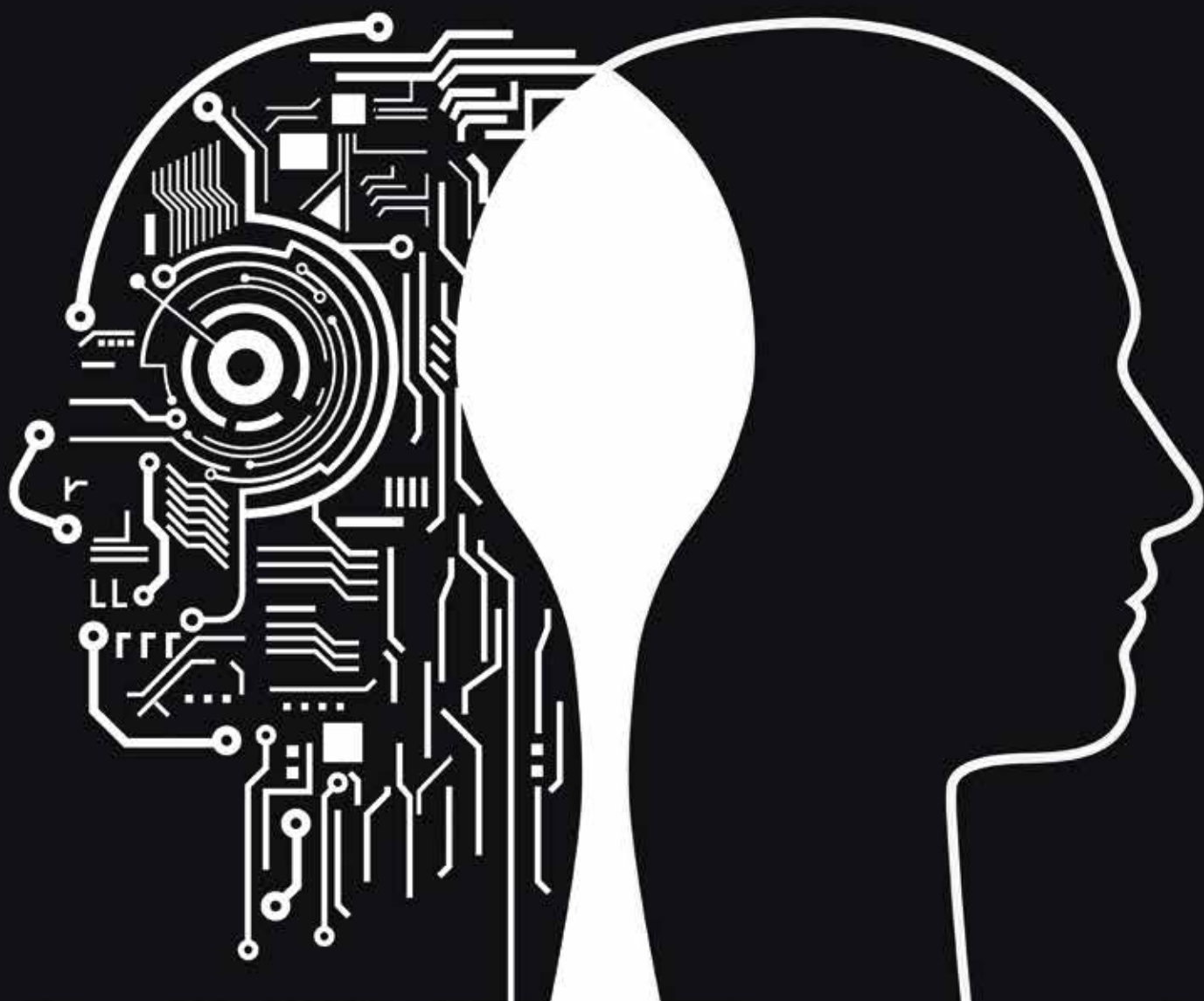


L'intelligenza ci seppellirà

È la fine dell'età dell'uomo. L'avvento dell'IA sembra dominare l'inizio del terzo millennio. Tutto questo sviluppo potrebbe però sfuggirci di mano. E allora ci troveremmo per la prima volta nella storia a competere con un nemico più astuto, più potente e più imprevedibile di quanto avessimo mai immaginato



La superintelligenza artificiale non esiste ancora né esiste l'intelligenza artificiale generale, quell'intelligenza in grado di imparare e per molti aspetti eguagliare e superare l'intelligenza umana. Tuttavia siamo circondati dall'intelligenza artificiale comune, che svolge centinaia di mansioni che all'uomo fanno comodo. Talvolta chiamata intelligenza artificiale debole, esegue utilissime ricerche online (Google), fornisce agli utenti consigli di lettura sulla base dei libri acquistati in precedenza (Amazon) e gestisce dal 50 al 70 per cento delle compravendite della Borsa di New York (Nyse) e del Nasdaq. Svolgendo un'unica missione, seppure con ottimi risultati, anche pezzi da novanta del calibro di Deep Blue, lo scacchista della Ibm, e Watson, il concorrente di *Jeopardy!*, rientrano nella categoria dell'IA debole. Finora l'IA ha mantenuto le promesse. Tra i numerosi chip che affollano la mia automobile, l'algoritmo che traduce la pressione del piede nella frenata più opportuna (sistema anti bloccaggio o Abs) riesce meglio di me a evitare gli slittamenti. Google Search è ormai il mio assistente virtuale, e probabilmente anche il vostro. Con l'IA, la vita ha un sapore migliore. E presto potrebbe migliorare ancora.

Immaginate squadre di un centinaio di computer, con la cultura di un ricercatore, impegnate ventiquattro ore su ventiquattro, sette giorni su sette, ad affrontare questioni come il cancro, la ricerca farmaceutica, l'estensione della vita, i carburanti sintetici e il cambiamento climatico. Pensate che rivoluzione sarebbe per la robotica se macchine intelligenti e adattive si sostituissero ai minatori, ai vigili del fuoco, ai soldati e agli astronauti al momento di svolgere lavori pericolosi. (...) Ma di cosa parliamo esattamente quando ci riferiamo alla straordinaria qualità di queste invenzioni, ossia un'intelligenza pari a quella dell'uomo? Che cosa, grazie all'intelligenza, siamo in grado di fare rispetto agli altri animali? Be', l'intelligenza ci permette di parlare al telefono. Di guidare. Di identificare migliaia di oggetti comuni, descriverli e utilizzarli. Sappiamo navigare in Internet. Qualcuno sa contare fino a dieci in più di una lingua, persino parlarne molte correntemente. (...) La stragrande maggioranza degli esseri viventi dispone fin dalla nascita delle doti di cui avrà bisogno in futuro. Noi no. Questa vasta gamma di capacità complesse è in sostanza l'intelligenza umana, l'intelligenza generale di cui gli sviluppatori dell'AGI (intelligenza artificiale generale) intendono dotare le macchine. (...)

Ora, quanto tempo occorrerà per sviluppare l'AGI? Alcuni esperti che ho consultato ritengono che dovremmo aspettarci un'intelligenza artificiale pari a quella dell'uomo entro e non oltre il 2020. Ma, nel complesso, stando a recenti sondaggi gli informatici e i professionisti dei settori in qualche modo connessi all'IA, come l'ingegneria, la robotica e le neuroscienze, sono più prudenti. Stimano infatti una probabilità di inventare l'AGI entro il 2028 superiore al 10 per cento, e superiore al 50 per cento di svilupparla entro il 2050. Vi sarebbe il 90 per cento delle probabilità che ciò avvenga prima della fine del secolo. Riguardo ai pro e ai contro, i risultati sono prevedibili: l'AGI ci ricompenserà con enormi benefici, ma al tempo stesso sarà causa di enormi disastri, e da uno di questi disastri il genere umano non uscirà vivo. La tragedia peggiore avverrà in seguito alla transizione dall'AGI - intelligenza pari a quella dell'uomo - all'ASI - superintelligenza. (...)

E il lasso di tempo tra l'AGI e l'ASI potrebbe essere breve. Ma, paradossalmente, se i rischi della condivisione del pianeta con un'IA superintelligente sono oggetto del dibattito internazionale tra gli esperti nel campo dell'IA, di tali rischi non si fa menzione nel dialogo con il pubblico. Come mai? Forse è proprio che chiamiamo "il bias della disponibilità" a impedirci di associare l'intelligenza artificiale allo sterminio dell'uomo. Non abbiamo ancora esperienza né siamo mai venuti a conoscenza di incidenti provocati dall'IA, mentre gli altri 'soliti sospetti' ci sono ben noti. Conosciamo i terribili virus dell'Hiv, della Sars e dell'influenza spagnola del 1918. Abbiamo assistito agli effetti delle armi nucleari sui centri abitati. Siamo rimasti atterriti dalle evidenze geologiche di antichi asteroidi grandi quanto il Texas. E le tragedie di Three Mile Island (1979), Chernobyl (1986) e Fukushima (2011) testimoniano che si può trarre insegnamento anche dalle lezioni più dolorose. Il nostro radar non ha ancora imparato a qualificare l'IA come minaccia esistenziale. Ripeto, un incidente cambierebbe tutto, così come l'11 settembre ha insegnato al mondo che gli aerei possono essere branditi come armi. (...)

Ma vi è un altro aspetto importante che differenzia un'IA fuori controllo dai comuni incidenti tecnologici. Quelli causati dagli aerei e dalle centrali nucleari sono eventi circoscritti: a disastro avvenuto, si provvede a rimediare. Una catastrofe dovuta all'IA interessa software intelligenti in grado migliorarsi e riprodursi a velocità elevate. In pratica, si propaga da

JAMES BARRAT



Tratto dal libro
La nostra invenzione finale
(Nutrimenti)

sola. Come arginare un disastro che va al di là dell'arma più potente di cui disponiamo: il cervello? Come rimediare a una tragedia che, una volta innescata, potrebbe non avere mai fine? Un'altra ragione per cui, quando si parla di rischi esistenziali, in modo alquanto sospetto non si fa mai menzione dell'IA è il fatto che a dominare i dibattiti su quest'ultima sia la Singolarità. "Singolarità" è ormai un termine così popolare da essere spesso usato a casaccio, benché abbia varie accezioni tra loro intercambiabili. Ray Kurzweil, noto inventore, scrittore e fautore della Singolarità, la definisce come un periodo di tempo "singolare" (che inizierà intorno all'anno 2045) dopo il quale il ritmo del cambiamento tecnologico modificherà irreversibilmente la vita dell'uomo. L'intelligenza sarà in gran parte informatizzata e miliardi di volte più potente rispetto a oggi. La Singolarità darà inizio a una nuova era durante la quale l'uomo porrà fine alla maggior parte dei problemi globali: la fame, la malattia, addirittura la morte. L'intelligenza artificiale è la star dello spettacolo mediatico della Singolarità, tra i cui protagonisti spicca la nanotecnologia. (...)

Uno dei rischi è che lo sviluppo di macchine più intelligenti dell'uomo dia il via all'estinzione della razza umana. È uno dei tanti inconvenienti dello spettacolo della Singolarità di cui non si parla abbastanza. Il silenzio potrebbe essere dovuto a quello che definisco "problema dei due minuti". Ho assistito a decine di conferenze sulla superintelligenza tenute da scienziati, inventori e filosofi. Molti, tra questi, la considerano inevitabile e ne celebrano le geniali promesse di abbondanza e prosperità. Dopodiché, di solito negli ultimi due minuti del discorso, aggiungono per inciso che un'inappropriata gestione dell'IA potrebbe portare all'estinzione del genere umano. Al che il pubblico si mette a ridacchiare nervosamente, impaziente di tornare alle buone notizie.

Gli scrittori che delineano la rivoluzione tecnologica si dividono in due tipologie. La prima è rappresentata da libri come quello di Kurzweil, *La singolarità è vicina*. Il loro scopo è di approntare un lavoro teorico preparatorio in vista di un futuro tutto sommato positivo. Se anche si accennasse a conseguenze disastrose, nessuno ci farebbe caso data la mole di ottimismo contenuta nel testo. *Wired for Thought* di Jeff Stibel esemplifica invece la seconda tipologia. Il libro guarda al futuro tecnologico attraverso la lente del business. In modo alquanto suggestivo, Stibel presenta Internet come un cervello sempre più connesso,

cosa di cui le start-up online dovrebbero tenere conto. Opere come quella di Stibel ambiscono a insegnare agli imprenditori a gettare una rete nel mare dei consumatori e delle tendenze online e pescare un mucchio di soldi. La maggior parte degli esperti e degli scrittori di tecnologia non tiene però conto di una terza prospettiva, meno rosea, che invece è proprio ciò che intendo fare io, con questo libro. L'ipotesi in analisi è che lo sviluppo di macchine intelligenti, prima, e di macchine più intelligenti dell'uomo, poi, non porterà all'integrazione di queste ultime nel nostro stile di vita, ma al loro predominio su di noi.

Nel creare l'AGI, i ricercatori inventeranno un tipo di intelligenza più avanzato della propria e, pertanto, non saranno in grado di gestirlo né comprenderlo adeguatamente. Sappiamo cosa accade quando esseri tecnologicamente avanzati si imbattono in altri più arretrati: Cristoforo Colombo contro i Taino, Pizarro contro gli Inca, gli europei contro i nativi americani. Preparatevi al prossimo scontro. L'intelligenza artificiale contro l'uomo. (...) Ma per ora preferisco soffermarmi sulle allarmanti questioni relative allo scenario della "creatura iperattiva". Ecco cos'è. Gli informatici, specialmente quelli al servizio della Difesa e dei servizi segreti, si sentiranno in dovere di velocizzare lo sviluppo dell'AGI perché le alternative (per esempio, che il governo cinese la metta a punto per primo) sono ancor più inquietanti dello sviluppo accelerato e avventato della propria tecnologia. Gli stessi informatici, inoltre, potrebbero avere l'esigenza di velocizzare lo sviluppo dell'AGI al fine di gestire al meglio altre tecnologie più pericolose, come la nanotecnologia, che probabilmente si affermeranno entro questo secolo. In questa prospettiva, la sospensione del lavoro per eseguire verifiche sull'andamento dell'evoluzione autonoma delle macchine non sarebbe contemplata. Un'intelligenza artificiale in grado di migliorarsi potrebbe saltare dall'AGI all'ASI con un'impennata molto simile "all'esplosione di intelligenza". Nell'impossibilità di prevedere il comportamento di un'intelligenza più evoluta di noi, concepiamo solo in minima parte quali mezzi questa userà contro di noi, per esempio l'auto-replicazione per disporre di più intelligenze che si dedichino simultaneamente al problem

Se eserciti di computer e robot sostituissero l'uomo nelle attività più pericolose, pensate che svolta sarebbe per la razza umana

solving e alla ricerca di strategie di fuga e sopravvivenza senza tener conto dei principi dell'onestà e della correttezza. Per concludere, non dovremmo limitarci a presumere che la prima ASI sarà o amichevole o ostile nei nostri confronti, ma valutare che potrebbe anche assumere una posizione ambivalente circa la felicità, il benessere e la sopravvivenza dell'uomo.

È possibile calcolare il rischio potenziale dell'ASI? Nel libro *Il rischio tecnologico*, H.W. Lewis individua alcune categorie di rischio e le classifica in base al loro indice di prevedibilità.

Gli eventi più prevedibili sono quelli altamente probabili e ad alto rischio, come spostarsi in auto da una città all'altra. I dati di cui tenere conto sono innumerevoli. Gli eventi con bassa probabilità e alto rischio, come i terremoti, sono più rari, e di conseguenza prevederli è più difficile. Ma hanno conseguenze tali che anticiparli è fondamentale. Vi sono poi rischi la cui probabilità è bassa perché legati a eventi che non si sono mai verificati finora e le cui conseguenze sono, anche in questo caso, gravi. I grossi cambiamenti climatici dovuti all'inquinamento da attività umana sono un ottimo esempio. Prima dell'esperimento del 16 luglio 1945 a White Sands, nel Nuovo Messico, l'esplosione di una bomba atomica era anch'essa un esempio di un evento mai sperimentato.

In teoria, la super intelligenza andrebbe collocata in questa categoria. L'esperienza non ci aiuta. Non possiamo calcolarne la probabilità con i tradizionali metodi statistici. Tuttavia credo che, stando all'attuale ritmo di sviluppo dell'IA, l'invenzione della superintelligenza rientri nella prima categoria: elevata probabilità e alto rischio.

