


James Barrat

La nostra invenzione finale

L'intelligenza artificiale e la fine dell'età dell'uomo

Traduzione di Daniela Pezzella e Monica Pezzella

 Nutrimenti

*A mia moglie, Alison Barrat,
per l'amore e il sostegno che mi fanno andare avanti*

Titolo originale: *Our Final Invention: Artificial Intelligence and the End
of the Human Era*

Copyright © 2013 by James Barrat

Traduzione dall'inglese di Daniela Pezzella e Monica Pezzella

© 2019 Nutrimenti srl

Prima edizione gennaio 2019

www.nutrimenti.net

via Marco Aurelio, 44 – 00184 Roma

In copertina: illustrazione di Giovanna Mathis

ISBN 978-88-6594-642-8

ISBN 978-88-6594-664-0 (ePub)

ISBN 978-88-6594-665-7 (MobiPocket)

Indice

Prefazione all'edizione italiana	9
Introduzione	15
La creatura iperattiva	21
Il problema dei due minuti	37
Uno sguardo al futuro	51
La strada più difficile	65
Programmi che scrivono programmi	85
Quattro pulsioni primarie	93
L'esplosione di intelligenza	115
Il punto di non ritorno	135
La legge dei ritorni acceleranti	149
Il singolaritarista	167
L'impennata	181
L'ultima complicazione	209
Natura inconoscibile	237
La fine dell'età dell'uomo?	255
L'ecosistema cibernetico	269
AGI 2.0	293

Negli anni Duemila, quando ho cominciato a dedicarmi all'intelligenza artificiale, gli unici a sottolinearne puntualmente gli aspetti negativi erano gli intellettuali, per esempio Eliezer Yudkowsky e Nick Bostrom. Gli scritti di Yudkowsky mi hanno indirizzato verso il lavoro svolto dal Machine Intelligence Research Institute di San Francisco, a quei tempi Singularity Institute, che ho poi visitato. Tramite il Miri sono entrato in contatto con Michael Vassar, Michael Anissimov, Luke Muehlhauser, Bill Hibbard, Anna Salamon, Ben Goertzel e altri eclettici intellettuali. Per cui, all'uscita di *La nostra invenzione finale*, ero convinto che chiunque avrebbe potuto intrattenere una conversazione sul 'rischio dell'IA' con gli addetti ai lavori in una cabina telefonica. Peccato che tutti i miei studenti mi domandassero: "Cos'è una cabina telefonica?".

Ho deciso di scrivere *La nostra invenzione finale* perché nessuno aveva ancora scritto un testo divulgativo su quello che considero il 'rischio dell'IA': lo studio dell'intelligenza artificiale in quanto tecnologia 'a duplice utilizzo', capace di apportare grandi benefici e grandi danni. In tal senso l'IA somiglia alla fissione nucleare, ma è addirittura più delicata e pericolosa. L'intelligenza artificiale possiederà tutte le inclinazioni machiavelliche dell'intelligenza umana, amplificate un milione di volte. Ero convinto,

e lo sono tuttora, che l'IA rappresentasse la minaccia principale alla sopravvivenza dell'uomo, peggio del riscaldamento globale, dell'Armageddon nucleare e dell'impatto di un asteroide. Avvertivo la necessità di mettere a disposizione del mondo un libro che esponesse con chiarezza questi problemi.

Soprattutto, sapevo che il libro in questione avrebbe dovuto essere adatto a tutti, privo cioè del gergo informatico degli 'addetti ai lavori', del linguaggio accademico e di termini che non fossero universalmente comprensibili. Perché? Perché l'intelligenza artificiale ci riguarda tutti.

Il nostro destino è strettamente legato alla rivoluzione dell'IA che si sta verificando in questo esatto momento. Tutti noi abbiamo il diritto di esserne al corrente e, quando ciascuno di noi sarà consapevole del rischio, potremo forse unire le nostre forze per difenderci.

In qualità di documentarista particolarmente interessato alla storia e alla scienza, mi sono preparato a scrivere questo libro per trent'anni. Ho girato documentari, tra gli altri, per National Geographic, Discovery, Pbs, trattando gli argomenti più disparati, dal Vangelo di Giuda alla scoperta della tomba di Erode, dall'E-bola alla genetica. Scrivere e produrre un documentario vuol dire prendere una tematica complessa e renderla accessibile a un pubblico più ampio. Vuol dire anche intervistare gli esperti in materia e tradurre le loro ricerche in una lingua che sia comprensibile per un tredicenne mediamente sveglio. Avevo tutte le carte in regola per lanciarmi nell'impresa.

L'intelligenza artificiale, come tutte le tecnologie informatiche, si evolve a velocità esponenziale. Dalla pubblicazione di *La nostra invenzione finale*, nel campo dell'IA sono cambiate molte cose, ma i principi fondamentali del rischio che essa comporta sono rimasti invariati. Le nuove tecniche di programmazione e l'hardware d'avanguardia renderanno le macchine più abili dell'uomo in tutti i settori, compreso quello di giochi come gli scacchi e il Go, il riconoscimento di oggetti, la navigazione, la ricerca e la

raccolta di informazioni, il ragionamento logico, i processi decisionali e via dicendo. L'IA ci rimpiazzerà in migliaia di mansioni. Prima o poi, in un futuro non troppo lontano, gli scienziati costruiranno macchine in grado di ricercare e sviluppare l'intelligenza artificiale meglio di noi. Dopodiché saranno le macchine e non più l'uomo a dettare il ritmo del progresso dell'intelligenza. Le macchine diverranno milioni di volte più intelligenti di noi, soprattutto in fatto di logica e matematica. Come noi, sfrutteranno l'intelligenza in svariati settori, non soltanto in un paio. Questo concetto è definito 'esplosione di intelligenza', e mi ci soffermerò con particolare attenzione in quest'opera. Se non conoscete ancora l'esplosione di intelligenza, sarò lieto di esporvi quello che è un concetto chiave per la sopravvivenza dell'uomo.

In seguito all'esplosione di intelligenza l'uomo sarà declassato ad abitante di second'ordine del pianeta; saremo trattati con la stessa attenzione e la stessa dignità che attualmente riserviamo ai nostri immediati avversari nella classifica dell'intelligenza: le scimmie. Vale a dire *nessuna*. Chiedete a un qualsiasi esperto di primati; vi dirà che le scimmie sono ormai agli sgoccioli. Se non riusciamo a creare macchine superintelligenti compatibili con l'uomo, l'*homo sapiens* è condannato all'estinzione.

Ma non è necessario aspettare così tanto prima che l'IA metta a repentaglio la nostra sopravvivenza. In una lettera scritta a più mani con altri scienziati, il fisico Stephen Hawking, oggi deceduto, ha dichiarato: "...l'impatto a breve termine dell'IA dipende da chi la gestisce, l'impatto a lungo termine dipende invece dalla possibilità o dall'impossibilità di gestirla".

Chi gestisce l'IA è fondamentale. Le corporazioni e i paesi che la sviluppano devono crearla e utilizzarla rispettando principi etici, salvaguardando il benessere dell'uomo. Tra non molto, una o più di queste entità controlleranno il software che innescherà l'esplosione d'intelligenza. Assicurarsi che quest'evento cruciale avvenga in totale sicurezza o non avvenga affatto è la principale responsabilità da cui il futuro dipende.

Al momento dell'esplosione d'intelligenza, potremo fidarci dei colossi della tecnologia che sono al timone della più potente delle IA?

Che dire di Google? Già prima che il suo portavoce mi mentisse, nel periodo in cui scrivevo questo libro, dubitavo di Google, oggi Alphabet. Negli ultimi anni i miei sospetti sono stati confermati. Google ha assunto quattrocento avvocati dopo essere stata citata in giudizio in più di duecento paesi per svariate questioni che vanno dalla violazione della privacy, alla violazione del copyright, alle pratiche imprenditoriali predatorie. Per la Cina, Google sta creando un motore di ricerca che censura i siti considerati offensivi da un regime nazionale sempre più accentrato. Di recente, la politica adottata da Google riguardo le molestie sessuali ha alimentato scioperi e proteste negli uffici della società da un capo all'altro del pianeta. Ed è confortante sapere che, non appena Google ha cominciato a interessarsi ai progetti del Dipartimento della Difesa per la costruzione di armi dotate di IA, molti dipendenti si sono rifiutati di partecipare.

Dobbiamo quindi fidarci della posizione di Google Inc. rispetto all'esplosione di intelligenza? Non credo proprio.

E che dire di Facebook, altro concorrente nella corsa all'IA? Facebook ha venduto i dati personali di più di 87 milioni di utenti degli Stati Uniti alla Cambridge Analytica, una società del Regno Unito. Nel 2016 la CA ha utilizzato questi dati a sostegno delle campagne di estrema destra condotte, tra gli altri, da Donald Trump e Ted Cruze. Contemporaneamente, Facebook ha facilitato la divulgazione di informazioni menzognere tra gli utenti statunitensi al fine di influenzare le elezioni presidenziali. E, con il consenso di Facebook, la Cambridge Analytica ha venduto i dati privati degli utenti alle imprese statali della Cina. Questo complesso di macchinazioni hanno richiesto l'uso di programmi di *data mining* di IA avanzata. Facebook ha contribuito a influenzare le elezioni negli Usa? È probabile, ma è impossibile saperlo se non indagando la mente di milioni di elettori. Ci si potrà fidare di Facebook al momento dell'esplosione d'intelligenza?

Assolutamente no.

A dispetto del titolo di questo libro, sono un gran sostenitore dell'intelligenza artificiale. *La nostra invenzione finale* analizza e celebra le potenti tecnologie dell'IA tra cui figurano, per esempio, gli algoritmi genetici, che sfruttano il potere dell'evoluzione e rappresentano un appassionante oggetto di riflessione. Questo libro esplora le reti neurali artificiali, ma non voglio anticipare in che modo queste ultime, unite ai grandi dati e ai nuovi processori, daranno vita ai sistemi di apprendimento approfondito. Sono molte le aziende, comprese quelle sanitarie, bancarie e dei trasporti, che hanno adottato con straordinario successo l'apprendimento approfondito. Particolarmente interessanti sono i progressi fatti dalla Deep Mind, oggi società semi-indipendente sussidiaria di Google. Gli esperti prevedevano che le macchine non avrebbero battuto i campioni di Go, il gioco di strategia orientale, per almeno un decennio. Eppure AlphaGo della Deep Mind ci è riuscita nel 2016 battendo il coreano Lee Sedol, diciotto volte campione del mondo. AlphaGo, come le sorelle algoritmiche AlphaGo Zero e AlphaZero, impara da sola. AlphaGo Zero si è allenata con gli *zero games* di Go. Ha giocato contro sé stessa in un processo chiamato *deep adversarial learning*. Nel giro di qualche giorno ha imparato a giocare a Go meglio di qualsiasi uomo o macchina. A mio parere si tratta di un esempio di esplosione d'intelligenza circoscritto a un unico settore. Una sorta di prototipo. E proprio per questo è terrificante.

Sono lieto di affermare che la pubblicazione di *La nostra invenzione finale* ha coinciso con un'ondata di consapevolezza in merito ai rischi dell'IA, e il libro è diventato un bestseller. La cabina telefonica sopra menzionata si è riempita davvero! *La nostra invenzione finale* è stato tradotto in otto lingue, compreso l'italiano. Oggi le organizzazioni che promuovono uno sviluppo e un'attuazione etica dell'IA sbocciano a ritmo sorprendente. Mentre scrivevo *La nostra invenzione finale* gli unici gruppi a fare questo tipo di informazione erano il Machine Intelligence

Research Institute e il Future of Humanity Institute. Oggi, a questi, se ne sono aggiunti almeno altri sei. Particolarmente interessante è l'AI Now Institute, che nella sua breve vita ha già dimostrato di non temere il confronto con potenti corporazioni quali Google, Facebook e Amazon, e non ha esitato a sottolineare le evidenti pecche nella loro gestione dell'IA. Ancora meglio fanno sperare i dipendenti dei colossi della tecnologia che hanno espresso la propria visione etica con scioperi e proteste. Sempre più spesso, quando qualcuno mi chiede come potremmo aggirare il problema di una IA che minaccia di sfruttarci o distruggerci, mi appello a questi atti di disobbedienza civile. Sono convinto della necessità della supervisione e del monitoraggio del governo sulle architetture cognitive avanzate. Ma ho il sospetto che i singoli individui potrebbero evitare che l'innovazione dell'IA ci sfugga di mano ancor meglio delle società di vigilanza e delle regolamentazioni governative. Come ho scritto nel libro che avete tra le mani, abbiamo un'unica possibilità di riuscita.

Spero che *La nostra invenzione finale. L'intelligenza artificiale e la fine dell'età dell'uomo* vi piaccia.

James Barrat
4 dicembre 2018

Introduzione

Qualche anno fa ho scoperto, con mia grande sorpresa, di avere qualcosa in comune con un gran numero di estranei. Uomini e donne che non avevo mai incontrato: scienziati e professori universitari, imprenditori della Silicon Valley, ingegneri, programmatori, blogger e via dicendo. Provenivano dai luoghi più disparati del Nord America, dell'Europa e dell'India; senza Internet, di loro non avrei mai saputo niente. Ad accomunarci era il fatto che, dopo aver a lungo riflettuto, dubitavamo fortemente che fosse possibile sviluppare un'intelligenza artificiale avanzata senza incorrere in gravi conseguenze. Individualmente o in piccoli gruppi avevamo studiato la letteratura in materia e ciascuno si era fatto la propria idea al riguardo. Mi sono così ritrovato a far parte di una rete di intellettuali e piccole associazioni, queste ultime più all'avanguardia e sofisticate di quanto immaginassi. Non condividevamo però soltanto un vago scetticismo circa l'IA; eravamo convinti che il tempo necessario per intervenire ed evitare il disastro stesse per scadere.

Ho girato documentari per più di vent'anni. Nel 2000 ho intervistato il maestro della fantascienza Arthur C. Clarke, l'inventore Ray Kurzweil e il pioniere della robotica Rodney Brooks. Kurzweil e Brooks dipingevano un quadro roseo, addirittura

entusiasta, della futura convivenza tra esseri umani e macchine intelligenti. Ma Clarke accennava alla possibilità che queste ultime potessero coglierci impreparati. Prima di allora ero inebriato dal potenziale dell'IA. Oggi, la sfiducia in un futuro roseo mi si è insinuata nella mente, contagiandola.

Nella mia professione, a fare la differenza è il pensiero critico: un documentarista deve stare in guardia dalle storie troppo allettanti per essere vere. Il rischio è sprecare mesi o anni a realizzare un documentario su una bufala, se non addirittura contribuire a diffonderne una. Tra le altre cose, ho saggiato l'attendibilità di un vangelo secondo Giuda Iscariota (autentico), di una tomba attribuita a Gesù di Nazareth (bufala), della tomba di Erode il Grande nei pressi di Gerusalemme (incontestabile) e della tomba di Cleopatra in un tempio di Osiride in Egitto (molto dubbia). In un'occasione un'emittente mi chiese di presentare con un'aura di credibilità un servizio sugli Ufo. Mi accorsi che il servizio in questione era una sequenza di bufale già smascherate: piatti volanti, sovrimpressioni, effetti e illusioni ottiche. Proposi di girare un documentario sui creatori di bufale anziché sugli Ufo. Mi licenziarono.

Diffidare dell'IA si è rivelato un'ardua impresa per due ragioni. Quello che prometteva aveva piantato nella mia mente un seme che ero intenzionato a coltivare, non confutare. In secondo luogo, non mettevo in dubbio né l'esistenza né il potere dell'IA. A rendermi scettico erano l'affidabilità dell'IA avanzata e l'imprudenza dell'attuale società nel mettere a punto tecnologie pericolose. Temevo che gli esperti fermamente convinti dell'affidabilità dell'IA si stessero illudendo. Non facevo che parlare con persone che l'intelligenza artificiale la conoscevano bene, e la loro esperienza prospettava scenari più allarmanti del previsto. Ho deciso di scrivere un libro in cui riportare le loro impressioni e preoccupazioni per condividerle con quante più persone possibile.

*

Durante la stesura di questo testo ho interpellato scienziati che producono intelligenza artificiale per la robotica, la ricerca in Internet, il *data mining*, il riconoscimento vocale e facciale e altre applicazioni di questo tipo. Ho parlato con scienziati determinati a creare un'intelligenza artificiale in grado di competere con quella umana, che avrà un infinito numero di applicazioni e ci cambierà la vita (se non la distruggerà prima). Mi sono confrontato con i direttori tecnici delle società di IA e con i consulenti tecnici per le operazioni segrete del Dipartimento della Difesa. Tutti prevedevano che tra non molto saranno le macchine, o esseri umani la cui intelligenza è da queste ultime aumentata, a prendere le decisioni fondamentali che regolano la vita dell'uomo. Quando? Molti sono convinti che vivranno abbastanza per esserne testimoni.

È un'affermazione sconvolgente ma non del tutto opinabile. I computer sono già indispensabili al sistema finanziario, alle infrastrutture idriche ed energetiche e ai trasporti. Sono onnipresenti negli ospedali, nelle automobili e negli elettrodomestici. Molti di essi, per esempio quelli che generano gli algoritmi di compravendita di Wall Street, funzionano autonomamente senza l'assistenza dell'uomo. Il prezzo da pagare in cambio dello svago e delle comodità offerti dai computer è la dipendenza. Di giorno in giorno, siamo sempre più dipendenti. Fin qui tutto bene.

Ma l'intelligenza artificiale dà vita ai computer e li trasforma in qualcosa di diverso. Ammesso che in futuro le macchine decideranno al posto nostro, mi sono chiesto: *quando* otterranno questo potere? Lo otterranno con il nostro consenso? *In che modo* assumeranno il controllo, e quanto in fretta? In questo libro, affronterò ciascuno di questi problemi.

Qualche scienziato ritiene che il nostro spodestamento sarà pacifico e collaborativo, più simile a un trasferimento di potere. Accadrà per gradi, solo i provocatori si impunteranno, ma gli altri dovranno riconoscere che disporre di un'intelligenza incomensurabilmente superiore in grado di decidere cosa è meglio per noi non fa che migliorarci la vita. Per di più, non è detto che la o

le IA che deterranno il controllo saranno robot freddi e disumani; potrebbero essere uomini aumentati; potrebbe essere un unico cervello umano potenziato e caricato su un computer. In tal modo sarà più facile digerirne l'autorità. Il trasferimento di consegne alle macchine, come descritto da alcuni scienziati, è praticamente identico a quello che stiamo sperimentando al momento: graduale, indolore, divertente.

La progressiva transizione all'egemonia dei computer procederebbe senza troppo clamore e, forse, senza pericoli non fosse che per un piccolo particolare: l'intelligenza. L'intelligenza non è imprevedibile solo *qualche* volta o in determinati casi. Per ragioni che esamineremo, è ipotizzabile che sistemi informatici avanzati al punto da sfoggiare un'intelligenza pari a quella umana saranno *sempre* imprevedibili e imperscrutabili. Al momento non possiamo sapere cosa faranno i sistemi consapevoli né come lo faranno. L'imperscrutabilità si sommerà ai casi fortuiti generati dalla complessità e agli eventi imprevedibili determinati dall'intelligenza, come quello che approfondiremo e che è noto come 'esplosione di intelligenza'.

Ma *in che modo* le macchine subentreranno all'uomo? Persino l'ipotesi migliore, tra quelle più realistiche, rappresenta una minaccia per l'umanità?

Alcuni tra i più illustri scienziati che ho intervistato hanno risposto alla domanda citando le tre leggi della robotica dello scrittore di fantascienza Isaac Asimov. Queste regole, hanno risposto senza pensarci troppo, saranno 'incorporate' nelle IA, quindi non avremo nulla da temere. Neanche stessero parlando di una scienza esatta. Vedremo le tre leggi nel primo capitolo, per ora basti dire che se qualcuno propone le leggi di Asimov come soluzione al dilemma delle macchine superintelligenti, vuol dire che non ci ha riflettuto poi tanto né si è confrontato abbastanza con altri. Come fare per creare macchine intelligenti *amichevoli* e cosa temere da quelle superintelligenti sono problemi non contemplati

dal tropo di Asimov. Esperienza e capacità nel settore dell'IA non immunizzano dal sottovalutarne i rischi.

Non sono certo il primo a ritenere che siamo in rotta di collisione. La razza umana sta per schiantarsi contro un'incognita mortale. In quest'opera valuto la possibilità che l'uomo perda il controllo del proprio futuro a beneficio di macchine che non necessariamente lo odieranno, ma che agiranno in maniera inaspettata allorché acquisiranno la forza più imprevedibile e potente che esista sviluppata ad altissimi livelli, livelli cui neanche l'uomo può aspirare, ed è probabile che tali facoltà si riveleranno incompatibili con la sopravvivenza del genere umano. Una forza talmente instabile e misteriosa che la natura ha sviluppato appieno una sola volta: l'intelligenza.