Dario Amodei - CEO Anthropic - Ottobre 2024

Machines of Loving Grace -

(da https://www.darioamodei.com/essay/machines-of-loving-grace)

Machines of Loving Grace[1]
Come l'IA potrebbe trasformare il mondo in meglio
Ottobre 2024

Penso e parlo molto dei rischi di un'IA potente. L'azienda di cui sono CEO, Anthropic, svolge molte ricerche su come ridurre questi rischi. Per questo motivo, le persone a volte traggono la conclusione che io sia un pessimista o un "doomer" che pensa che l'IA sarà per lo più negativa o pericolosa. Non lo penso affatto. In realtà, uno dei miei motivi principali per concentrarmi sui rischi è che sono l'unica cosa che si frappone tra noi e quello che vedo come un futuro fondamentalmente positivo. Penso che la maggior parte delle persone stia sottovalutando quanto radicale possa essere il lato positivo dell'IA, proprio come penso che la maggior parte delle persone stia sottovalutando quanto gravi possano essere i rischi.
In questo saggio cerco di delineare come potrebbe essere quel lato positivo: come potrebbe apparire un mondo con un'IA potente se tutto andasse bene. Naturalmente nessuno può conoscere il futuro con certezza o precisione, e gli effetti di un'IA potente saranno probabilmente ancora più imprevedibili dei cambiamenti tecnologici passati, quindi tutto questo consisterà inevitabilmente in supposizioni. Ma miro ad almeno a fare supposizioni istruite e utili, che catturino il sapore di ciò che accadrà anche se la maggior parte dei dettagli finirà per essere errata. Includo molti dettagli principalmente perché penso che una visione concreta faccia di più per far progredire la discussione rispetto a una visione altamente cautelata e astratta.
In primo luogo, tuttavia, volevo spiegare brevemente perché io e Anthropic non abbiamo parlato molto dei vantaggi di un'IA potente e perché probabilmente continueremo, nel complesso, a parlare molto dei rischi. In particolare, ho fatto questa scelta per il desiderio di:
• Massimizzare l'influenza (leverage). Lo sviluppo di base della tecnologia IA e molti (non tutti) i suoi benefici sembrano inevitabili (a meno che i rischi non facciano deragliare tutto) e sono fondamentalmente guidati da potenti forze di mercato. D'altra parte, i rischi non sono predeterminati e le nostre azioni possono cambiare notevolmente la loro probabilità.
• Evitare la percezione di propaganda. Le aziende di IA che parlano di tutti i fantastici vantaggi dell'IA possono apparire come propagandisti, o come se stessero cercando di distrarre dagli svantaggi. Penso anche che, per principio, sia dannoso per l'anima passare troppo tempo a "promuovere il proprio interesse" (talking your book).
• Evitare la grandiosità. Sono spesso infastidito dal modo in cui molti personaggi pubblici esperti di rischi dell'IA (per non parlare dei leader delle aziende di IA) parlano del mondo post-AGI, come se fosse la loro missione realizzarlo da soli come

un profeta che conduce il suo popolo alla salvezza. Penso che sia pericoloso vedere le aziende come entità che plasmano unilateralmente il mondo, ed è pericoloso vedere gli obiettivi tecnologici pratici in termini essenzialmente religiosi.
• Evitare il bagaglio "sci-fi". Sebbene pensi che la maggior parte delle persone sottovaluti il potenziale positivo di un'IA potente, la piccola comunità di persone che discute di futuri radicali legati all'IA lo fa spesso con un tono eccessivamente "da fantascienza" (caratterizzato ad esempio da menti caricate digitalmente, esplorazione spaziale o vibrazioni cyberpunk generali). Penso che questo porti le persone a prendere meno sul serio le affermazioni e a intriderle di una sorta di irrealtà. Per essere chiari, il punto non è se le tecnologie descritte siano possibili o probabili (il saggio principale ne discute in dettaglio granulare) [illegible] è più che il "vibee" contrabbanda implicitamente una serie di bagagli culturali e presupposti non dichiarati su quale tipo di futuro sia desiderabile, su come si risolveranno varie questioni sociali, ecc. Il risultato finisce spesso per sembrare una fantasia per una ristretta sottocultura, risultando sgradevole per la maggior parte delle persone.
Eppure, nonostante tutte le preoccupazioni di cui sopra, penso davvero che sia importante discutere di come potrebbe essere un mondo positivo con un'IA potente, facendo del nostro meglio per evitare le trappole sopra menzionate. In effetti, penso che sia fondamentale avere una visione del futuro genuinamente ispiratrice, e non solo un piano per spegnere gli incendi. Molte delle implicazioni di un'IA potente sono conflittuali o pericolose, ma alla fine di tutto, ci deve essere qualcosa per cui stiamo combattendo, un risultato a somma positiva in cui tutti stanno meglio, qualcosa per spingere le persone a elevarsi al di sopra dei loro bisticci e affrontare le sfide future. La paura è un tipo di motivazione, ma non è sufficiente: abbiamo bisogno anche della speranza.
L'elenco delle applicazioni positive di un'IA potente è estremamente lungo (e comprende robotica, produzione, energia e molto altro), ma mi concentrerò su un piccolo numero di aree che mi sembrano avere il maggior potenziale per migliorare direttamente la qualità della vita umana. Le cinque categorie per cui sono più entusiasta sono:
1. Biologia e salute fisica
2. Neuroscienze e salute mentale
3. Sviluppo economico e povertà
4. Pace e governance
5. Lavoro e significato
Le mie previsioni saranno radicali se giudicate dalla maggior parte degli standard (tranne che dalle visioni di "singolarità" fantascientifiche[2]), ma le intendo con serietà e sincerità. Tutto ciò che dico potrebbe benissimo essere sbagliato (per ribadire il mio punto precedente), ma ho almeno tentato di basare le mie opinioni su una valutazione semi-analitica di quanto il progresso in vari campi potrebbe accelerare e cosa ciò potrebbe significare in pratica. Ho la fortuna di avere un'esperienza professionale sia in biologia che in neuroscienze, e sono un dilettante informato nel campo dello sviluppo economico, ma sono sicuro che sbaglierò molte cose. Una cosa che la scrittura di questo saggio mi ha fatto capire è che sarebbe prezioso riunire un gruppo di esperti di dominio (in biologia, economia, relazioni internazionali e altre aree) per

scrivere una versione molto migliore e più informata di quella che ho prodotto qui. Probabilmente è meglio vedere i miei sforzi qui come uno spunto iniziale per quel gruppo.

Presupposti di base e quadro di riferimento

Per rendere l'intero saggio più preciso e fondato, è utile specificare chiaramente cosa intendiamo per IA potente (ovvero la soglia alla quale inizia il conto alla rovescia di 5-10 anni), oltre a delineare un quadro di riferimento per pensare agli effetti di tale IA una volta presente.

Come sarà un'IA potente (non mi piace il termine AGI)[3] e quando (o se) arriverà, è un argomento enorme in sé. È un tema di cui ho discusso pubblicamente e su cui potrei scrivere un saggio completamente separato (probabilmente lo farò prima o poi). Ovviamente, molte persone sono scettiche sul fatto che un'IA potente sarà costruita presto e alcune sono scettiche sul fatto che sarà mai costruita. Penso che potrebbe arrivare già nel 2026, anche se ci sono modi in cui potrebbe volerci molto più tempo. Ma ai fini di questo saggio, vorrei mettere da parte queste questioni, assumere che arriverà ragionevolmente presto e concentrarmi su ciò che accadrà nei 5-10 anni successivi. Voglio anche assumere una definizione di come sarà un tale sistema, quali siano le sue capacità e come interagisca, anche se c'è spazio per il disaccordo su questo.

Per IA potente, ho in mente un modello di IA [?] probabilmente siimile agli attuali LLM nella forma, sebbene possa basarsi su un'architettura diversa, possa coinvolgere diversi modelli interagenti e possa essere addestrato diversamente [?] con le seguenti proprietà:

• In termini di pura intelligenza[4], è più intelligente di un vincitore di premio Nobel nella maggior parte dei campi rilevanti: biologia, programmazione, matematica, ingegneria, scrittura, ecc. Ciò significa che può dimostrare teoremi matematici irrisolti, scrivere romanzi estremamente validi, scrivere basi di codice difficili da zero, ecc.

• Oltre a essere solo una "cosa intelligente con cui parli", ha tutte le "interfacce" a disposizione di un essere umano che lavora virtualmente, inclusi testo, audio, video, controllo di mouse e tastiera e accesso a Internet. Può intraprendere qualsiasi azione, comunicazione o operazione remota abilitata da questa interfaccia, incluso compiere azioni su Internet, dare o ricevere istruzioni da esseri umani, ordinare materiali, dirigere esperimenti, guardare video, realizzare video e così via. Svolge tutti questi compiti con, ancora una volta, una competenza superiore a quella degli esseri umani più capaci al mondo.

• Non risponde solo passivamente alle domande; invece, gli possono essere assegnati compiti che richiedono ore, giorni o settimane per essere completati, e poi parte e svolge quei compiti in modo autonomo, come farebbe un dipendente intelligente, chiedendo chiarimenti se necessario.

• Non ha un corpo fisico (se non vivere su uno schermo di computer), ma può controllare strumenti fisici esistenti, robot o attrezzature di laboratorio tramite un computer; in teoria potrebbe persino progettare robot o attrezzature per il proprio uso.

• Le risorse utilizzate per addestrare il modello possono essere riutilizzate per

eseguire milioni di istanze di esso (questo corrisponde alle dimensioni dei cluster previste per il ~2027), e il modello può assorbire informazioni e generare azioni a una velocità circa 10x-100x superiore a quella umana[5]. Potrebbe tuttavia essere limitato dal tempo di risposta del mondo fisico o del software con cui interagisce.
• Ognuna di queste milioni di copie può agire indipendentemente su compiti non correlati, o se necessario possono tutte lavorare insieme nello stesso modo in cui collaborerebbero gli umani, magari con diverse sottopopolazioni ottimizzate per essere particolarmente brave in compiti specifici.
Potremmo riassumere questo come un "paese di geni in un datacenter".
Chiaramente un'entità del genere sarebbe in grado di risolvere problemi molto difficili, molto velocemente, ma non è banale capire quanto velocemente. Due posizioni "estreme" mi sembrano entrambe false. In primo luogo, si potrebbe pensare che il mondo verrebbe trasformato istantaneamente su una scala di secondi o giorni ("la Singolarità"), poiché un'intelligenza superiore si auto-alimenta e risolve quasi immediatamente ogni possibile compito scientifico, ingegneristico e operativo. Il problema è che esistono limiti fisici e pratici reali, ad esempio per quanto riguarda la costruzione di hardware o la conduzione di esperimenti biologici. Anche un nuovo paese di geni si scontrerebbe con questi limiti. L'intelligenza può essere molto potente, ma non è polvere magica.
In secondo luogo, e per contro, si potrebbe credere che il progresso tecnologico sia saturo o limitato dalla velocità dei dati del mondo reale o da fattori sociali, e che un'intelligenza superiore a quella umana aggiungerà molto poco[6]. Questo mi sembra altrettanto implacabile posso pensare a centinaia di problemii scientifici o persino sociali in cui un folto gruppo di persone davvero intelligenti accelererebbe drasticamente il progresso, specialmente se non sono limitate all'analisi e possono far accadere cose nel mondo reale (cosa che il nostro postulato paese di geni può fare, anche dirigendo o assistendo team di umani).
Penso che la verità sia probabilmente una mescolanza disordinata di questi due quadri estremi, qualcosa che varia a seconda del compito e del campo ed è molto sottile nei suoi dettagli. Credo che servano nuovi quadri di riferimento per pensare a questi dettagli in modo produttivo.
Gli economisti parlano spesso di "fattori di produzione": cose come lavoro, terra e capitale. La frase "rendimenti marginali del lavoro/terra/capitale" cattura l'idea che in una data situazione, un dato fattore può o meno essere quello limitante per esempio, un"aeronautica ha bisogno sia di aerei che di piloti, e assumere più piloti non aiuta molto se hai finito gli aerei. Credo che nell'era dell'IA dovremmo parlare dei rendimenti marginali dell'intelligenza[7] e cercare di capire quali siano gli altri fattori complementari all'intelligenza che diventano fattori limitanti quando l'intelligenza è molto alta. Non siamo abituati a pensare in questo modo a chiederci "quanto aiuta essere più intelligenti in questo compito e in quale lasso di tempo?" ma sembra il modo giusto di concettualizzare uun mondo con un'IA molto potente.
La mia ipotesi su un elenco di fattori che limitano o sono complementari all'intelligenza include:
1. Velocità del mondo esterno. Gli agenti intelligenti devono operare in modo

interattivo nel mondo per realizzare cose e anche per apprendere[8]. Ma il mondo si muove solo a una certa velocità. Le cellule e gli animali funzionano a una velocità fissa, quindi gli esperimenti su di essi richiedono un certo tempo che può essere irriducibile. Lo stesso vale per l'hardware, la scienza dei materiali, tutto ciò che comporta la comunicazione con le persone e persino la nostra infrastruttura software esistente. Inoltre, nella scienza molti esperimenti sono spesso necessari in sequenza, ognuno dei quali impara o si basa sull'ultimo. Tutto ciò significa che la velocità con cui un progetto importante ad esempio lo svilupppo di una cura per il cancro può essere completato può avere un minimo irriducibile che non può essere ulteriormente ridotto anche se l'intelligenza continua ad aumentare.
2. Necessità di dati. A volte mancano i dati grezzi e in loro assenza una maggiore intelligenza non aiuta. I fisici delle particelle di oggi sono molto ingegnosi e hanno sviluppato una vasta gamma di teorie, ma mancano dei dati per scegliere tra esse perché i dati degli acceleratori di particelle sono molto limitati. Non è chiaro se farebbero drasticamente meglio se fossero superintelligenti se non forse accelerando laa costruzione di un acceleratore più grande.
3. Complessità intrinseca. Alcune cose sono intrinsecamente imprevedibili o caotiche e anche l'IA più potente non può prevederle o districarle sostanzialmente meglio di quanto possa fare un umano o un computer oggi. Ad esempio, anche un'IA incredibilmente potente potrebbe prevedere solo marginalmente più avanti in un sistema caotico (come il problema dei tre corpi) nel caso generale[9], rispetto agli umani e ai computer di oggi.
4. Vincoli umani. Molte cose non possono essere fatte senza infrangere leggi, danneggiare gli esseri umani o scombussolare la società. Un'IA allineata non vorrebbe fare queste cose (e se abbiamo un'IA non allineata, torniamo a parlare di rischi). Molte strutture sociali umane sono inefficienti o addirittura attivamente dannose, ma sono difficili da cambiare rispettando vincoli come i requisiti legali sulle sperimentazioni cliniche, la volontà delle persone di cambiare le proprie abitudini o il comportamento dei governi. Esempi di progressi che funzionano bene in senso tecnico, ma il cui impatto è stato sostanzialmente ridotto da regolamenti o paure fuori luogo, includono l'energia nucleare, il volo supersonico e persino gli ascensori.
5. Leggi fisiche. Questa è una versione più netta del primo punto. Ci sono alcune leggi fisiche che sembrano infrangibili. Non è possibile viaggiare più veloci della luce. Il budino non si "smescola". I chip possono avere solo un certo numero di transistor per centimetro quadrato prima di diventare inaffidabili. Il calcolo richiede una certa energia minima per bit cancellato, limitando la densità del calcolo nel mondo.
C'è un'ulteriore distinzione basata sulle scale temporali. Cose che sono vincoli rigidi a breve termine possono diventare più malleabili all'intelligenza a lungo termine. Ad esempio, l'intelligenza potrebbe essere utilizzata per sviluppare un nuovo paradigma sperimentale che ci consenta di imparare in vitro ciò che un tempo richiedeva esperimenti su animali vivi, o per costruire gli strumenti necessari per raccogliere nuovi dati (ad esempio, l'acceleratore di particelle più grande), o per (entro i limiti etici) trovare modi per aggirare i vincoli basati sull'uomo (ad esempio, aiutando a migliorare il sistema delle sperimentazioni cliniche, aiutando a creare nuove

giurisdizioni dove le sperimentazioni cliniche abbiano meno burocrazia, o migliorando la scienza stessa per rendere le sperimentazioni cliniche umane meno necessarie o più economiche).
Pertanto, dovremmo immaginare un quadro in cui l'intelligenza è inizialmente fortemente limitata dagli altri fattori di produzione, ma nel tempo l'intelligenza stessa aggira sempre più gli altri fattori, anche se non si dissolvono mai completamente (e alcune cose come le leggi fisiche sono assolute)[10]. La domanda chiave è quanto velocemente accade tutto questo e in quale ordine.
Con il quadro sopra esposto in mente, cercherò di rispondere a questa domanda per le cinque aree menzionate nell'introduzione.

1. Biologia e salute

La biologia è probabilmente l'area in cui il progresso scientifico ha il maggior potenziale per migliorare direttamente e inequivocabilmente la qualità della vita umana. Nell'ultimo secolo alcune delle più antiche afflizioni umane (come il vaiolo) sono state finalmente sconfitte, ma molte altre rimangono, e sconfiggerle sarebbe un enorme traguardo umanitario. Oltre a curare le malattie, la scienza biologica può in linea di principio migliorare la qualità di base della salute umana, estendendo la durata della vita sana, aumentando il controllo e la libertà sui nostri processi biologici e affrontando problemi quotidiani che attualmente consideriamo parti immutabili della condizione umana.
Nel linguaggio dei "fattori limitanti" della sezione precedente, le principali sfide nell'applicazione diretta dell'intelligenza alla biologia sono i dati, la velocità del mondo fisico e la complessità intrinseca (infatti, tutte e tre sono correlate tra loro). Anche i vincoli umani giocano un ruolo in una fase successiva, quando sono coinvolte le sperimentazioni cliniche. Esaminiamoli uno per uno.
Gli esperimenti su cellule, animali e persino processi chimici sono limitati dalla velocità del mondo fisico: molti protocolli biologici comportano la coltura di batteri o altre cellule, o semplicemente l'attesa che si verifichino reazioni chimiche, e questo a volte può richiedere giorni o addirittura settimane, senza alcun modo ovvio per accelerarlo. Gli esperimenti sugli animali possono richiedere mesi (o più) e gli esperimenti sull'uomo spesso richiedono anni (o addirittura decenni per studi sugli esiti a lungo termine). In qualche modo correlata a ciò, spesso mancano i dati non tantoo in quantità, quanto in qualità: c'è sempre carenza di dati chiari e univoci che isolino un effetto biologico di interesse dalle altre 10.000 cose confondenti che stanno accadendo, o che intervengano causalmente in un dato processo, o che misurino direttamente un effetto (invece di dedurre le sue conseguenze in qualche modo indiretto o rumoroso). Persino dati molecolari massicci e quantitativi, come i dati di proteomica che ho raccolto lavorando su tecniche di spettrometria di massa, sono rumorosi e mancano di molto (in quali tipi di cellule si trovavano queste proteine? In quale parte della cellula? In quale fase del ciclo cellulare?).
In parte responsabile di questi problemi con i dati è la complessità intrinseca: se avete mai visto un diagramma che mostra la biochimica del metabolismo umano, saprete che è molto difficile isolare l'effetto di una qualsiasi parte di questo sistema complesso, e ancora più difficile intervenire sul sistema in modo preciso o

prevedibile. E infine, oltre al tempo intrinseco necessario per eseguire un esperimento sugli esseri umani, le sperimentazioni cliniche effettive comportano molta burocrazia e requisiti normativi che (secondo l'opinione di molte persone, me compreso) aggiungono ulteriore tempo non necessario e ritardano il progresso.

Dato tutto ciò, molti biologi sono stati a lungo scettici sul valore dell'IA e dei "big data" più in generale in biologia. Storicamente, matematici, informatici e fisici che hanno applicato le loro competenze alla biologia negli ultimi 30 anni hanno avuto un discreto successo, ma non hanno avuto l'impatto veramente trasformativo inizialmente sperato. Parte dello scetticismo è stato ridotto da scoperte importanti e rivoluzionarie come AlphaFold (che ha appena meritatamente fruttato ai suoi creatori il premio Nobel per la Chimica) e AlphaProteo[11], ma c'è ancora la percezione che l'IA sia (e continuerà a essere) utile solo in un insieme limitato di circostanze. Una formulazione comune è: "L'IA può fare un lavoro migliore nell'analizzare i tuoi dati, ma non può produrre più dati o migliorare la qualità dei dati. Spazzatura dentro, spazzatura fuori (Garbage in, garbage out)".

Ma penso che questa prospettiva pessimistica stia considerando l'IA nel modo sbagliato. Se la nostra ipotesi principale sul progresso dell'IA è corretta, allora il modo giusto di pensare all'IA non è come un metodo di analisi dei dati, ma come un biologo virtuale che svolge tutti i compiti che fanno i biologi, incluso progettare ed eseguire esperimenti nel mondo reale (controllando i robot di laboratorio o semplicemente dicendo agli umani quali esperimenti eseguire come farebbe un ricercatore principale con i suuoi studenti laureati), inventando nuovi metodi biologici o tecniche di misurazione e così via. È accelerando l'intero processo di ricerca che l'IA può veramente accelerare la biologia. Voglio ripeterlo perché è il malinteso più comune che sorge quando parlo della capacità dell'IA di trasformare la biologia: non sto parlando dell'IA come di un mero strumento per analizzare i dati. In linea con la definizione di IA potente all'inizio di questo saggio, sto parlando di usare l'IA per eseguire, dirigere e migliorare quasi tutto ciò che fanno i biologi.

Per essere più specifici su dove penso che arriverà l'accelerazione, una frazione sorprendentemente ampia del progresso in biologia è derivata da un numero veramente esiguo di scoperte, spesso legate a strumenti o tecniche di misurazione ad ampio spettro[12] che consentono interventi precisi ma generalizzati o programmabili nei sistemi biologici. C'è forse ~1 di queste grandi scoperte all'anno e collettivamente esse guidano verosimilmente oltre il 50% del progresso in biologia. Queste scoperte sono così potenti proprio perché tagliano la complessità intrinseca e i limiti dei dati, aumentando direttamente la nostra comprensione e il nostro controllo sui processi biologici. Alcune scoperte per decennio hanno permesso sia la maggior parte della nostra comprensione scientifica di base della biologia, sia guidato molti dei trattamenti medici più potenti.

Alcuni esempi includono:

• CRISPR: una tecnica che consente l'editing dal vivo di qualsiasi gene in organismi viventi (sostituzione di qualsiasi sequenza genica arbitraria con qualsiasi altra sequenza arbitraria). Da quando la tecnica originale è stata sviluppata, ci sono stati costanti miglioramenti per colpire tipi di cellule specifici, aumentando la precisione e riducendo le modifiche del gene sbagliato tutte cose necessarie per un uso sicuro

negli essseri umani.
• Vari tipi di microscopia per osservare cosa sta succedendo a un livello preciso: microscopi ottici avanzati (con vari tipi di tecniche fluorescenti, ottiche speciali, ecc.), microscopi elettronici, microscopi a forza atomica, ecc.
• Sequenziamento e sintesi del genoma, il cui costo è crollato di diversi ordini di grandezza negli ultimi due decenni.
• Tecniche optogenetiche che consentono di far attivare un neurone puntandogli contro una luce.
• Vaccini a mRNA che, in linea di principio, ci consentono di progettare un vaccino contro qualsiasi cosa e poi adattarlo rapidamente (i vaccini a mRNA sono ovviamente diventati famosi durante il COVID).
• Terapie cellulari come CAR-T che consentono di prelevare cellule immunitarie dal corpo e "riprogrammarle" per attaccare, in linea di principio, qualsiasi cosa.
• Intuizioni concettuali come la teoria dei germi o la realizzazione di un legame tra il sistema immunitario e il cancro[13].
Mi sto prendendo la briga di elencare tutte queste tecnologie perché voglio fare un'affermazione cruciale su di esse: penso che il loro tasso di scoperta potrebbe essere aumentato di 10 volte o più se ci fossero molti più ricercatori creativi e di talento. O, per dirla in un altro modo, penso che i rendimenti dell'intelligenza siano elevati per queste scoperte e che tutto il resto in biologia e medicina derivi principalmente da esse.
Perché lo penso? Per le risposte ad alcune domande che dovremmo abituarci a porre quando cerchiamo di determinare i "rendimenti dell'intelligenza". Primo, queste scoperte sono generalmente fatte da un numero esiguo di ricercatori, spesso le stesse persone ripetutamente, il che suggerisce competenza e non una ricerca casuale (quest'ultima potrebbe suggerire che gli esperimenti lunghi siano il fattore limitante). Secondo, spesso "avrebbero potuto essere fatte" anni prima di quando lo sono state: ad esempio, CRISPR era una componente naturale del sistema immunitario nei batteri nota dagli anni '80, ma ci sono voluti altri 25 anni perché le persone capissero che poteva essere riutilizzata per l'editing genico generale. Spesso vengono anche ritardate di molti anni dalla mancanza di supporto della comunità scientifica per direzioni promettenti (si veda il profilo dell'inventore dei vaccini a mRNA; storie simili abbondano). Terzo, i progetti di successo sono spesso frammentari o erano idee secondarie che inizialmente la gente non riteneva promettenti, piuttosto che sforzi massicciamente finanziati. Ciò suggerisce che non è solo la massiccia concentrazione di risorse a guidare le scoperte, ma l'ingegno.
Infine, sebbene alcune di queste scoperte abbiano una "dipendenza seriale" (devi fare la scoperta A per avere gli strumenti o le conoscenze per fare la scoperta B)  il che potrebbe ancora creare ritardi sperimentali molte, forse la maaggior parte, sono indipendenti, il che significa che molte possono essere portate avanti in parallelo. Sia questi fatti, sia la mia esperienza generale di biologo, mi suggeriscono fortemente che ci sono centinaia di queste scoperte in attesa di essere fatte se gli scienziati fossero più intelligenti e più bravi a fare collegamenti tra la vasta quantità di conoscenze biologiche possedute dall'umanità (si consideri di nuovo l'esempio CRISPR). Il successo di AlphaFold/AlphaProteo nel risolvere problemi importanti in modo molto

più efficace degli umani, nonostante decenni di modelli fisici accuratamente progettati, fornisce una prova di principio (sebbene con uno strumento ristretto in un dominio ristretto) che dovrebbe indicare la via da seguire.
Pertanto, la mia ipotesi è che un'IA potente potrebbe aumentare di almeno 10 volte il tasso di queste scoperte, regalandoci i prossimi 50-100 anni di progresso biologico in 5-10 anni[14]. Perché non 100 volte? Forse è possibile, ma qui sia la dipendenza seriale che i tempi degli esperimenti diventano importanti: ottenere 100 anni di progresso in 1 anno richiede che molte cose vadano bene al primo colpo, inclusi gli esperimenti sugli animali e cose come la progettazione di microscopi o costosi laboratori. In realtà sono aperto all'idea (che forse suona assurda) di poter ottenere 1000 anni di progresso in 5-10 anni, ma molto scettico sul fatto che si possano ottenere 100 anni in 1 anno. Un altro modo di dirlo è che penso ci sia un ritardo costante inevitabile: gli esperimenti e la progettazione dell'hardware hanno una certa "latenza" e devono essere iterati un numero "irriducibile" di volte per imparare cose che non possono essere dedotte logicamente. Ma oltre a ciò potrebbe essere possibile un parallelismo massiccio[15].
Che dire delle sperimentazioni cliniche? Sebbene vi sia molta burocrazia e rallentamenti associati ad esse, la verità è che gran parte (anche se non tutta!) della loro lentezza deriva in definitiva dalla necessità di valutare rigorosamente farmaci che funzionano a malapena o funzionano in modo ambiguo. Purtroppo questo è vero per la maggior parte delle terapie odierne: il farmaco medio contro il cancro aumenta la sopravvivenza di pochi mesi pur avendo effetti collaterali significativi che devono essere misurati attentamente (una storia simile vale per i farmaci contro l'Alzheimer). Ciò porta a studi enormi (per ottenere potenza statistica) e a difficili compromessi che le agenzie di regolamentazione generalmente non sono brave a gestire, sempre a causa della burocrazia e della complessità di interessi in conflitto.
Quando qualcosa funziona davvero bene, va molto più veloce: esiste un percorso di approvazione accelerato e la facilità di approvazione è molto maggiore quando le dimensioni dell'effetto sono più ampie. I vaccini a mRNA per il COVID sono stati approvati in 9 mesi – molto più velocemente del ritmo abituale. Detto questo, anche in queste condizioni le sperimentazioni cliniche sono ancora troppo lente – i vaccini a mRNA avrebbero probabilmente dovuto essere approvati in circa 2 mesi. Ma questi tipi di ritardi (~1 anno dall'inizio alla fine per un farmaco) combinati con una massiccia parallelizzazione e la necessità di alcune, ma non troppe, iterazioni ("pochi tentativi") sono molto compatibili con una trasformazione radicale in 5-10 anni. In modo ancora più ottimistico, è possibile che la scienza biologica abilitata dall'IA riduca la necessità di iterazioni nelle sperimentazioni cliniche sviluppando migliori modelli sperimentali animali e cellulari (o persino simulazioni) più accurati nel prevedere cosa accadrà negli esseri umani. Ciò sarà particolarmente importante nello sviluppo di farmaci contro il processo di invecchiamento, che si svolge nell'arco di decenni e dove abbiamo bisogno di un ciclo di iterazione più veloce.
Infine, sul tema delle sperimentazioni cliniche e delle barriere sociali, vale la pena sottolineare esplicitamente che per certi versi le innovazioni biomediche hanno un record insolitamente positivo di successo nell'implementazione, contrariamente ad altre tecnologie[16]. Come accennato nell'introduzione, molte tecnologie sono

ostacolate da fattori sociali nonostante funzionino bene tecnicamente. Ciò potrebbe suggerire una prospettiva pessimistica su ciò che l'IA può compiere. Ma la biomedicina è unica in quanto, sebbene il processo di sviluppo dei farmaci sia eccessivamente macchinoso, una volta sviluppati essi vengono generalmente implementati e utilizzati con successo.

Per riassumere quanto sopra, la mia previsione di base è che la biologia e la medicina abilitate dall'IA ci permetteranno di comprimere il progresso che i biologi umani avrebbero ottenuto nei prossimi 50-100 anni in 5-10 anni. Mi riferirò a questo come al "21° secolo compresso": l'idea che dopo lo sviluppo di un'IA potente, in pochi anni faremo tutti i progressi in biologia e medicina che avremmo fatto nell'intero 21° secolo.

Sebbene prevedere cosa possa fare un'IA potente in pochi anni rimanga intrinsecamente difficile e speculativo, c'è una certa concretezza nel chiedersi "cosa potrebbero fare gli umani senza aiuto nei prossimi 100 anni?". Semplicemente guardando a ciò che abbiamo compiuto nel 20° secolo, o estrapolando dai primi 2 decenni del 21°, o chiedendosi cosa ci darebbero "10 CRISPR e 50 CAR-T", tutto ciò offre modi pratici e fondati per stimare il livello generale di progresso che potremmo aspettarci da un'IA potente.

Di seguito provo a stilare un elenco di ciò che potremmo aspettarci. Questo non si basa su alcuna metodologia rigorosa e si rivelerà quasi certamente errato nei dettagli, ma cerca di trasmettere il livello generale di radicalismo che dovremmo aspettarci:

1. Prevenzione e trattamento affidabili di quasi tutte[17] le malattie infettive naturali. Dati gli enormi progressi contro le malattie infettive nel 20° secolo, non è radicale immaginare di poter più o meno "finire il lavoro" in un 21° secolo compresso. I vaccini a mRNA e tecnologie simili indicano già la strada verso "vaccini per qualsiasi cosa". Se le malattie infettive saranno completamente eradicate dal mondo (anziché solo in alcuni luoghi) dipende da questioni riguardanti la povertà e la disuguaglianza, discusse nella Sezione 3.
2. Eliminazione della maggior parte dei tumori. I tassi di mortalità per cancro sono diminuiti di circa il 2% all'anno negli ultimi decenni; quindi siamo sulla buona strada per eliminare la maggior parte dei tumori nel 21° secolo al ritmo attuale della scienza umana. Alcuni sottotipi sono già stati ampiamente curati (ad esempio alcuni tipi di leucemia con la terapia CAR-T), e sono forse ancora più entusiasta per farmaci molto selettivi che colpiscono il cancro nella sua infanzia e gli impediscono di crescere. L'IA renderà possibili regimi di trattamento finemente adattati al genoma individualizzato del cancro – questi sono possibili oggi, ma enormemeente costosi in termini di tempo e competenza umana, che l'IA dovrebbe permetterci di scalare. Riduzioni del 95% o più sia della mortalità che dell'incidenza sembrano possibili. Detto questo, il cancro è estremamente vario e adattivo, ed è probabilmente la più difficile di queste malattie da distruggere completamente. Non sorprenderebbe se persistesse un assortimento di tumori rari e difficili.
3. Prevenzione molto efficace e cure efficaci per le malattie genetiche. Uno screening embrionale notevolmente migliorato renderà probabilmente possibile prevenire la maggior parte delle malattie genetiche, e qualche discendente più sicuro e affidabile di CRISPR potrebbe curare la maggior parte delle malattie genetiche nelle persone

esistenti. Le afflizioni che colpiscono l'intero corpo e interessano una grande frazione di cellule potrebbero tuttavia essere le ultime a resistere.
4. Prevenzione dell'Alzheimer. Abbiamo avuto molta difficoltà a capire cosa causi l'Alzheimer (è in qualche modo correlato alla proteina beta-amiloide, ma i dettagli effettivi sembrano essere molto complessi). Sembra esattamente il tipo di problema che può essere risolto con strumenti di misurazione migliori che isolino gli effetti biologici; quindi sono ottimista sulla capacità dell'IA di risolverlo. C'è una buona probabilità che alla fine possa essere prevenuto con interventi relativamente semplici, una volta capito cosa sta succedendo. Detto questo, i danni derivanti dall'Alzheimer già esistente potrebbero essere molto difficili da invertire.
5. Miglioramento del trattamento della maggior parte degli altri disturbi. Questa è una categoria onnicomprensiva per altri disturbi tra cui diabete, obesità, malattie cardiache, malattie autoimmuni e altro ancora. La maggior parte di questi sembra "più facile" da risolvere rispetto al cancro e all'Alzheimer e in molti casi è già in forte calo. Ad esempio, i decessi per malattie cardiache sono già diminuiti di oltre il 50% e interventi semplici come gli agonisti GLP-1 hanno già fatto enormi progressi contro l'obesità e il diabete.
6. Libertà biologica. Gli ultimi 70 anni sono stati caratterizzati da progressi nel controllo delle nascite, nella fertilità, nella gestione del peso e molto altro. Ma sospetto che la biologia accelerata dall'IA amplierà notevolmente ciò che è possibile: il peso, l'aspetto fisico, la riproduzione e altri processi biologici saranno completamente sotto il controllo delle persone. Ci riferiremo a questi sotto il titolo di libertà biologica: l'idea che ognuno dovrebbe essere messo in grado di scegliere ciò che vuole diventare e vivere la propria vita nel modo che più lo attrae. Ci saranno ovviamente questioni importanti sull'uguaglianza globale di accesso; si veda la Sezione 3 per queste.
7. Raddoppio della durata della vita umana[18]. Questo potrebbe sembrare radicale, ma l'aspettativa di vita è aumentata quasi di 2 volte nel 20° secolo (da ~40 anni a ~75), quindi è "in linea con la tendenza" che il "21° secolo compresso" la raddoppi ancora a 150. Ovviamente gli interventi coinvolti nel rallentare il processo di invecchiamento effettivo saranno diversi da quelli necessari nel secolo scorso per prevenire le morti premature (soprattutto infantili) per malattia, ma l'entità del cambiamento non è senza precedenti[19]. Concretamente, esistono già farmaci che aumentano la durata massima della vita nei ratti del 25-50% con effetti negativi limitati. E alcuni animali (ad esempio alcuni tipi di tartarughe) vivono già 200 anni, quindi gli esseri umani non sono manifestamente a un limite superiore teorico. A occhio e croce, la cosa più importante di cui c'è bisogno potrebbero essere biomarcatori dell'invecchiamento umano affidabili e non manipolabili (non-Goodhart-able), poiché ciò consentirà un'iterazione rapida su esperimenti e sperimentazioni cliniche. Una volta che la durata della vita umana sarà di 150 anni, potremmo essere in grado di raggiungere la "velocità di fuga", guadagnando abbastanza tempo affinché la maggior parte di coloro che sono attualmente vivi oggi possa vivere quanto vuole, sebbene non vi sia certamente alcuna garanzia che ciò sia biologicamente possibile.
Vale la pena guardare questo elenco e riflettere su quanto sarà diverso il mondo se tutto ciò venisse raggiunto tra 7-12 anni (il che sarebbe in linea con una timeline

aggressiva per l'IA). Inutile dire che sarebbe un trionfo umanitario inimmaginabile, l'eliminazione in un colpo solo della maggior parte dei flagelli che hanno perseguitato l'umanità per millenni. Molti dei miei amici e colleghi stanno crescendo figli e, quando quei figli saranno grandi, spero che ogni menzione di malattia suoni per loro come lo scorbuto, il vaiolo o la peste bubbonica suonano per noi. Quella generazione beneficerà anche di una maggiore libertà biologica ed espressione di sé e, con fortuna, potrebbe anche essere in grado di vivere quanto desidera.
È difficile sopravvalutare quanto questi cambiamenti saranno sorprendenti per tutti, tranne che per la piccola comunità di persone che si aspettavano un'IA potente. Ad esempio, migliaia di economisti ed esperti di politiche negli Stati Uniti discutono attualmente di come mantenere solvibili la Previdenza Sociale e Medicare e, più in generale, di come contenere i costi dell'assistenza sanitaria (che viene consumata per lo più da chi ha più di 70 anni e specialmente da chi ha malattie terminali come il cancro). La situazione per questi programmi probabilmente migliorerà radicalmente se tutto questo si avverasse[20], poiché il rapporto tra popolazione in età lavorativa e popolazione in pensione cambierà drasticamente. Senza dubbio queste sfide saranno sostituite da altre, come quella di garantire un accesso diffuso alle nuove tecnologie, ma vale la pena riflettere su quanto cambierà il mondo anche se la biologia fosse l'unica area ad essere accelerata con successo dall'IA.

2. Neuroscienze e mente

Nella sezione precedente mi sono concentrato sulle malattie fisiche e sulla biologia in generale, e non ho trattato le neuroscienze o la salute mentale. Ma le neuroscienze sono una sottodisciplina della biologia e la salute mentale è altrettanto importante della salute fisica. In realtà, se non altro, la salute mentale influisce sul benessere umano in modo ancora più diretto della salute fisica. Centinaia di milioni di persone hanno una qualità di vita molto bassa a causa di problemi come dipendenza, depressione, schizofrenia, autismo a basso funzionamento, disturbo da stress post-traumatico (PTSD), psicopatia[21], o disabilità intellettive. Altri miliardi lottano con problemi quotidiani che possono spesso essere interpretati come versioni molto più lievi di uno di questi gravi disturbi clinici. E come per la biologia generale, potrebbe essere possibile andare oltre la risoluzione dei problemi per migliorare la qualità di base dell'esperienza umana.
Il quadro di base che ho esposto per la biologia si applica ugualmente alle neuroscienze. Il campo è spinto in avanti da un piccolo numero di scoperte spesso legate a strumenti per la misurazione o l'intervento preciso nell'elenco di quelli sopra, l'optogenetiica è stata una scoperta neuroscientifica, e più recentemente CLARITY e la microscopia a espansione sono progressi sulla stessa scia, oltre a molti dei metodi generali di biologia cellulare che si trasferiscono direttamente alle neuroscienze. Penso che il tasso di questi progressi sarà similmente accelerato dall'IA e quindi che il quadro di "100 anni di progresso in 5-10 anni" si applichi alle neuroscienze nello stesso modo in cui si applica alla biologia e per le stesse ragioni. Come in biologia, il progresso delle neuroscienze del 20° secolo è stato enorme ad esempio non capivamo nemmeno come o pperché i neuroni si attivassero fino agli anni '50. Pertanto, sembra ragionevole aspettarsi che le neuroscienze accelerate

dall'IA producano rapidi progressi nel giro di pochi anni.
C'è una cosa che dovremmo aggiungere a questo quadro di base: alcune delle cose che abbiamo imparato (o stiamo imparando) sull'IA stessa negli ultimi anni aiuteranno probabilmente a far progredire le neuroscienze, anche se continuassero a essere portate avanti solo da esseri umani. L'interpretabilità è un esempio ovvio: sebbene i neuroni biologici operino superficialmente in modo completamente diverso dai neuroni artificiali (comunicano tramite impulsi e spesso frequenze di impulsi, quindi c'è un elemento temporale non presente nei neuroni artificiali, e un sacco di dettagli relativi alla fisiologia cellulare e ai neurotrasmettitori ne modificano sostanzialmente il funzionamento), la domanda fondamentale di "come fanno reti distribuite e addestrate di unità semplici che eseguono operazioni lineari/non lineari combinate a lavorare insieme per eseguire calcoli importanti" è la stessa, e sospetto fortemente che i dettagli della comunicazione dei singoli neuroni saranno astratti nella maggior parte delle domande interessanti sul calcolo e sui circuiti[22]. Come un solo esempio di ciò, un meccanismo computazionale scoperto dai ricercatori di interpretabilità nei sistemi di IA è stato recentemente riscoperto nel cervello dei topi. È molto più facile fare esperimenti sulle reti neurali artificiali che su quelle reali (queste ultime richiedono spesso di incidere nel cervello degli animali), quindi l'interpretabilità potrebbe benissimo diventare uno strumento per migliorare la nostra comprensione delle neuroscienze. Inoltre, le IA potenti saranno probabilmente esse stesse in grado di sviluppare e applicare questo strumento meglio di quanto possano fare gli umani.
Oltre alla semplice interpretabilità, però, ciò che abbiamo imparato dall'IA su come vengono addestrati i sistemi intelligenti dovrebbe (anche se non sono sicuro che l'abbia già fatto) causare una rivoluzione nelle neuroscienze. Quando lavoravo nelle neuroscienze, molte persone si concentravano su quelle che ora considererei le domande sbagliate sull'apprendimento, perché il concetto di ipotesi di scalabilità (scaling hypothesis) / lezione amara (bitter lesson) non esisteva ancora. L'idea che una semplice funzione obiettivo unita a molti dati possa guidare comportamenti incredibilmente complessi rende più interessante capire le funzioni obiettivo e i bias architetturali e meno interessante capire i dettagli dei calcoli emergenti. Non ho seguito il campo da vicino negli ultimi anni, ma ho la vaga sensazione che i neuroscienziati computazionali non abbiano ancora assorbito appieno la lezione. Il mio atteggiamento verso l'ipotesi di scalabilità è sempre stato: "Aha [illegible] questa è una spiegazione, ad alto livello, di comee funziona l'intelligenza e di come si è evoluta così facilmente", ma non credo che questa sia la visione del neuroscienziato medio, in parte perché l'ipotesi di scalabilità come "segreto dell'intelligenza" non è pienamente accettata nemmeno all'interno dell'IA.
Penso che i neuroscienziati dovrebbero cercare di combinare questa intuizione di base con le particolarità del cervello umano (limitazioni biofisiche, storia evolutiva, topologia, dettagli degli input/output motori e sensoriali) per cercare di risolvere alcuni dei puzzle chiave delle neuroscienze. Probabilmente alcuni lo stanno facendo, ma sospetto che non sia ancora sufficiente e che i neuroscienziati IA saranno in grado di sfruttare in modo più efficace questa angolazione per accelerare il progresso.
Mi aspetto che l'IA acceleri il progresso neuroscientifico lungo quattro percorsi

distinti, che spero possano lavorare tutti insieme per curare le malattie mentali e migliorare le funzioni:
1. Biologia molecolare, chimica e genetica tradizionali. Questa è essenzialmente la stessa storia della biologia generale nella Sezione 1, e l'IA può probabilmente accelerarla tramite gli stessi meccanismi. Esistono molti farmaci che modulano i neurotrasmettitori per alterare la funzione cerebrale, influenzare la vigilanza o la percezione, cambiare l'umore, ecc., e l'IA può aiutarci a inventarne molti altri. L'IA può probabilmente anche accelerare la ricerca sulla base genetica delle malattie mentali.
2. Misurazione e intervento neurale a grana fine. Questa è la capacità di misurare ciò che fanno molti singoli neuroni o circuiti neuronali e intervenire per cambiare il loro comportamento. L'optogenetica e le sonde neurali sono tecnologie capaci sia di misurazione che di intervento in organismi vivi, e sono stati proposti anche numerosi metodi molto avanzati (come i "ticker tape" molecolari per leggere i pattern di attivazione di un gran numero di singoli neuroni) che sembrano possibili in linea di principio.
3. Neuroscienze computazionali avanzate. Come notato sopra, sia le intuizioni specifiche che la gestalt della moderna IA possono probabilmente essere applicate fruttuosamente alle questioni delle neuroscienze dei sistemi, incluso forse lo svelamento delle vere cause e dinamiche di malattie complesse come la psicosi o i disturbi dell'umore.
4. Interventi comportamentali. Non ho menzionato molto, data l'attenzione sul lato biologico delle neuroscienze, ma la psichiatria e la psicologia hanno ovviamente sviluppato un ampio repertorio di interventi comportamentali nel corso del 20° secolo; è logico che l'IA possa accelerare anche questi, sia nello sviluppo di nuovi metodi che aiutando i pazienti ad aderire ai metodi esistenti. Più in generale, l'idea di un "coach IA" che ti aiuti sempre a essere la versione migliore di te stesso, che studi le tue interazioni e ti aiuti a imparare a essere più efficace, sembra molto promettente.
La mia ipotesi è che questi quattro percorsi di progresso lavorando insieme sarebbero, come per le malattie fisiche, sulla buona strada per portare alla cura o alla prevenzione della maggior parte delle malattie mentali nei prossimi 100 anni anche se l'IA non fosse coinvolta e quindi potrebbero rragionevolmente essere completati in 5-10 anni accelerati dall'IA. Concretamente, la mia previsione di ciò che accadrà è qualcosa del tipo:
• La maggior parte delle malattie mentali può probabilmente essere curata. Non sono un esperto di malattie psichiatriche (il mio periodo nelle neuroscienze è stato dedicato alla costruzione di sonde per studiare piccoli gruppi di neuroni) ma la mia ipotesi è che malattie come PTSD, depressione, schizofrenia, dipendenza, ecc. possano essere comprese e trattate in modo molto efficace attraverso una combinazione delle quattro direzioni di cui sopra. La risposta sarà probabilmente una combinazione di "qualcosa è andato storto biochimicamente" (sebbene possa essere molto complesso) e "qualcosa è andato storto con la rete neurale, ad alto livello". Cioè, è una questione di neuroscienze dei sistemi anche se ciò non sminuisce ll'impatto degli interventi comportamentali discussi sopra. Gli strumenti per la misurazione e l'intervento, specialmente negli esseri umani vivi, sembrano destinati a portare a un'iterazione e a

un progresso rapidi.
• Le condizioni che sono molto "strutturali" potrebbero essere più difficili, ma non impossibili. Ci sono alcune prove che la psicopatia sia associata a evidenti differenze neuroanatomiche che alcune regionii del cervello siano semplicemente più piccole o meno sviluppate negli psicopatici. Si ritiene inoltre che gli psicopatici manchino di empatia fin dalla giovane età; qualunque cosa sia diversa nel loro cervello, probabilmente è sempre stata così. Lo stesso può valere per alcune disabilità intellettive e forse altre condizioni. Ristrutturare il cervello sembra difficile, ma sembra anche un compito con elevati rendimenti dell'intelligenza. Forse c'è qualche modo per riportare il cervello adulto a uno stato più precoce o plastico in cui possa essere rimodellato. Sono molto incerto su quanto ciò sia possibile, ma il mio istinto è di essere ottimista su ciò che l'IA può inventare qui.
• La prevenzione genetica efficace delle malattie mentali sembra possibile. La maggior parte delle malattie mentali è parzialmente ereditaria, e gli studi di associazione genome-wide (GWAS) stanno iniziando a fare progressi nell'identificazione dei fattori rilevanti, che sono spesso numerosi. Probabilmente sarà possibile prevenire la maggior parte di queste malattie tramite lo screening embrionale, analogamente a quanto detto per le malattie fisiche. Una differenza è che la malattia psichiatrica è più probabile che sia poligenica (molti geni contribuiscono), quindi a causa della complessità c'è un rischio maggiore di selezionare involontariamente contro tratti positivi correlati alla malattia. Stranamente, tuttavia, negli ultimi anni gli studi GWAS sembrano suggerire che queste correlazioni potrebbero essere state sopravvalutate. In ogni caso, le neuroscienze accelerate dall'IA potrebbero aiutarci a capire queste cose. Naturalmente, lo screening embrionale per tratti complessi solleva una serie di questioni sociali e sarà controverso, anche se sospetto che la maggior parte delle persone sosterrebbe lo screening per malattie mentali gravi o debilitanti.
• Verranno risolti anche i problemi quotidiani che non consideriamo malattie cliniche. Molti di noi hanno problemi psicologici quotidiani che non vengono ordinariamente pensati come tali da raggiungere il livello di malattia clinica. Alcune persone si arrabbiano facilmente, altre hanno difficoltà a concentrarsi o sono spesso sonnolente, alcune sono timorose o ansiose, o reagiscono male al cambiamento. Oggi esistono già farmaci per aiutare ad esempio con la vigilanza o la concentrazione (caffeina, modafinil, ritalin) ma, come in molti altri settori precedenti, è probabile che sia possibile fare molto di più. Probabilmente esistono molti più farmaci di questo tipo che non sono stati scoperti, e potrebbero esserci anche modalità di intervento totalmente nuove, come la stimolazione luminosa mirata (si veda l'optogenetica sopra) o i campi magnetici. Dato quanti farmaci abbiamo sviluppato nel 20° secolo che sintonizzano la funzione cognitiva e lo stato emotivo, sono molto ottimista riguardo al "21° secolo compresso" in cui ognuno può far sì che il proprio cervello si comporti un po' meglio e avere un'esperienza quotidiana più appagante.
• L'esperienza umana di base può essere molto migliore. Facendo un ulteriore passo avanti, molte persone hanno vissuto momenti straordinari di rivelazione, ispirazione creativa, compassione, appagamento, trascendenza, amore, bellezza o pace meditativa. Il carattere e la frequenza di queste esperienze variano notevolmente da

persona a persona e all'interno della stessa persona in tempi diversi, e possono anche essere innescati a volte da vari farmaci (spesso però con effetti collaterali). Tutto ciò suggerisce che lo "spazio di ciò che è possibile sperimentare" sia molto ampio e che una frazione più ampia della vita delle persone potrebbe consistere in questi momenti straordinari. È probabilmente possibile migliorare anche varie funzioni cognitive su tutta la linea. Questa è forse la versione neuroscientifica della "libertà biologica" o della "durata della vita estesa".
Un argomento che ricorre spesso nelle raffigurazioni fantascientifiche dell'IA, ma di cui intenzionalmente non ho discusso qui, è il "mind uploading", l'idea di catturare lo schema e le dinamiche di un cervello umano e istanziarli in un software. Questo argomento potrebbe essere oggetto di un saggio a sé stante, ma basti dire che, mentre penso che l'uploading sia quasi certamente possibile in linea di principio, in pratica deve affrontare sfide tecnologiche e sociali significative, anche con un'IA potente, che probabilmente lo collocano al di fuori della finestra di 5-10 anni di cui stiamo discutendo.
In sintesi, le neuroscienze accelerate dall'IA probabilmente miglioreranno enormemente i trattamenti per, o addirittura cureranno, la maggior parte delle malattie mentali, oltre ad ampliare notevolmente la "libertà cognitiva e mentale" e le capacità cognitive ed emotive umane. Sarà altrettanto radicale quanto i miglioramenti della salute fisica descritti nella sezione precedente. Forse il mondo non sarà visibilmente diverso all'esterno, ma il mondo vissuto dagli esseri umani sarà un luogo molto migliore e più umano, nonché un luogo che offre maggiori opportunità di autorealizzazione. Sospetto anche che una migliore salute mentale attenuerà molti altri problemi sociali, compresi quelli che sembrano politici o economici.

3. Sviluppo economico e povertà

Le due sezioni precedenti riguardano lo sviluppo di nuove tecnologie che curano le malattie e migliorano la qualità della vita umana. Tuttavia, una domanda ovvia, dal punto di vista umanitario, è: "tutti avranno accesso a queste tecnologie?".
Una cosa è sviluppare una cura per una malattia, un'altra è eradicare la malattia dal mondo. Più in generale, molti interventi sanitari esistenti non sono ancora stati applicati ovunque nel mondo e lo stesso vale per i miglioramenti tecnologici (non sanitari) in generale. Un altro modo di dire questo è che gli standard di vita in molte parti del mondo sono ancora disperatamente poveri: il PIL pro capite è di circa 2.000 dollari nell'Africa sub-sahariana rispetto ai circa 75.000 dollari degli Stati Uniti. Se l'IA aumentasse ulteriormente la crescita economica e la qualità della vita nel mondo sviluppato, facendo poco per aiutare il mondo in via di sviluppo, dovremmo considerarlo un terribile fallimento morale e una macchia sulle genuine vittorie umanitarie delle due sezioni precedenti. Idealmente, un'IA potente dovrebbe aiutare il mondo in via di sviluppo a raggiungere il mondo sviluppato, mentre rivoluziona quest'ultimo.
Non sono così fiducioso che l'IA possa affrontare la disuguaglianza e la crescita economica quanto lo sono sul fatto che possa inventare tecnologie fondamentali, perché la tecnologia ha rendimenti dell'intelligenza così ovviamente elevati (compresa la capacità di aggirare le complessità e la mancanza di dati), mentre

l'economia coinvolge molti vincoli umani, oltre a una grossa dose di complessità intrinseca. Sono alquanto scettico sul fatto che un'IA possa risolvere il famoso "problema del calcolo socialista"[23] e non penso che i governi cederanno (o dovrebbero cedere) la loro politica economica a tale entità, anche se potesse farlo. Ci sono anche problemi come quello di convincere le persone ad accettare trattamenti che sono efficaci ma di cui potrebbero essere sospettose.

Le sfide che deve affrontare il mondo in via di sviluppo sono rese ancora più complicate dalla corruzione pervasiva sia nel settore privato che in quello pubblico. La corruzione crea un circolo vizioso: esaspera la povertà e la povertà a sua volta alimenta ulteriore corruzione. I piani guidati dall'IA per lo sviluppo economico devono fare i conti con la corruzione, le istituzioni deboli e altre sfide squisitamente umane.

Ciononostante, vedo ragioni significative per l'ottimismo. Le malattie sono state eradicate e molti paesi sono passati da poveri a ricchi, ed è chiaro che le decisioni coinvolte in questi compiti mostrano elevati rendimenti dell'intelligenza (nonostante i vincoli umani e la complessità). Pertanto, l'IA può probabilmente farli meglio di quanto vengano fatti attualmente. Potrebbero esserci anche interventi mirati che aggirano i vincoli umani e su cui l'IA potrebbe concentrarsi. Cosa ancora più importante, però, dobbiamo provarci. Sia le aziende di IA che i decisori politici del mondo sviluppato dovranno fare la loro parte per garantire che il mondo in via di sviluppo non venga escluso; l'imperativo morale è troppo grande. Quindi in questa sezione continuerò a presentare il caso ottimistico, ma tenendo presente ovunque che il successo non è garantito e dipende dai nostri sforzi collettivi.

Di seguito faccio alcune supposizioni su come penso che le cose possano andare nel mondo in via di sviluppo nei 5-10 anni successivi allo sviluppo di un'IA potente:

• Distribuzione degli interventi sanitari. L'area in cui sono forse più ottimista è la distribuzione degli interventi sanitari in tutto il mondo. Le malattie sono state effettivamente eradicate da campagne calate dall'alto: il vaiolo è stato completamente eliminato negli anni '70, e la polio e il verme di Guinea sono quasi eradicati con meno di 100 casi all'anno. I modelli epidemiologici matematicamente sofisticati svolgono un ruolo attivo nelle campagne di eradicazione delle malattie, e sembra molto probabile che vi sia spazio per sistemi di IA più intelligenti dell'uomo per fare un lavoro migliore di quello svolto dagli esseri umani. Anche la logistica della distribuzione può probabilmente essere notevolmente ottimizzata. Una cosa che ho imparato come donatore precoce di GiveWell è che alcune organizzazioni caritative sanitarie sono molto più efficaci di altre; la speranza è che gli sforzi accelerati dall'IA siano ancora più efficaci. Inoltre, alcuni progressi biologici rendono effettivamente la logistica della distribuzione molto più semplice: ad esempio, la malaria è stata difficile da eradicare perché richiede cure ogni volta che la malattia viene contratta; un vaccino che deve essere somministrato solo una volta rende la logistica molto più semplice (e tali vaccini per la malaria sono in effetti attualmente in fase di sviluppo). Sono possibili meccanismi di distribuzione ancora più semplici: alcune malattie potrebbero in linea di principio essere eradicate prendendo di mira i loro vettori animali, ad esempio rilasciando zanzare infettate da un batterio che blocca la loro capacità di trasmettere una malattia (che poi infettano tutte le altre zanzare) o

semplicemente usando il gene drive per eliminare le zanzare. Ciò richiede una o poche azioni centralizzate, anziché una campagna coordinata che deve curare individualmente milioni di persone. Complessivamente, penso che 5-10 anni sia una timeline ragionevole affinché una buona frazione (forse il 50%) dei benefici sanitari guidati dall'IA si propaghi anche ai paesi più poveri del mondo. Un buon obiettivo potrebbe essere che il mondo in via di sviluppo, 5-10 anni dopo un'IA potente, sia almeno sostanzialmente più sano di quanto lo sia oggi il mondo sviluppato, anche se continuasse a restare indietro rispetto a quest'ultimo. Realizzare questo richiederà naturalmente un enorme sforzo nella salute globale, nella filantropia, nel sostegno politico e in molti altri ambiti, in cui dovrebbero aiutare sia gli sviluppatori di IA che i decisori politici.

• Crescita economica. Può il mondo in via di sviluppo raggiungere rapidamente il mondo sviluppato, non solo nella salute, ma in generale dal punto di vista economico? C'è qualche precedente per questo: negli ultimi decenni del 20° secolo, diverse economie dell'Asia orientale hanno raggiunto tassi di crescita annua del PIL reale costanti del ~10%, consentendo loro di raggiungere il mondo sviluppato. I pianificatori economici umani hanno preso le decisioni che hanno portato a questo successo, non controllando direttamente intere economie ma azionando alcune leve chiave (come una politica industriale di crescita trainata dalle esportazioni e resistendo alla tentazione di fare affidamento sulla ricchezza delle risorse naturali); è plausibile che "ministri delle finanze e banchieri centrali IA" possano replicare o superare questo traguardo del 10%. Una domanda importante è come convincere i governi del mondo in via di sviluppo ad adottarli rispettando il principio di autodeterminazione alcuni potrebbero esserne entuusiasti, ma altri saranno probabilmente scettici. Sul versante ottimistico, molti degli interventi sanitari citati nel punto precedente probabilmente aumenteranno organicamente la crescita economica: eradicare l'AIDS/la malaria/i vermi parassiti avrebbe un effetto trasformativo sulla produttività, per non parlare dei benefici economici che alcuni degli interventi neuroscientifici (come il miglioramento dell'umore e della concentrazione) avrebbero sia nel mondo sviluppato che in quello in via di sviluppo. Infine, la tecnologia accelerata dall'IA non legata alla salute (come la tecnologia energetica, i droni da trasporto, i materiali da costruzione migliorati, la logistica e la distribuzione migliori, e così via) potrebbe semplicemente permeare il mondo in modo naturale; ad esempio, persino i telefoni cellulari hanno permeato rapidamente l'Africa sub-sahariana attraverso meccanismi di mercato, senza bisogno di sforzi filantropici. Sul versante più negativo, mentre l'IA e l'automazione hanno molti potenziali benefici, pongono anche sfide per lo sviluppo economico, in particolare per i paesi che non si sono ancora industrializzati. Trovare modi per garantire che questi paesi possano ancora svilupparsi e migliorare le proprie economie in un'era di crescente automazione è una sfida importante da affrontare per economisti e decisori politici. Complessivamente, uno scenario da sogno forse un oobiettivo a cui puntare sarebbe un tasso di crescita aannuale del PIL del 20% nel mondo in via di sviluppo, con un 10% ciascuno derivante da decisioni economiche abilitate dall'IA e dalla diffusione naturale di tecnologie accelerate dall'IA, incluse ma non limitate alla salute. Se raggiunto, ciò porterebbe l'Africa sub-sahariana all'attuale PIL pro capite

della Cina in 5-10 anni, elevando gran parte del resto del mondo in via di sviluppo a livelli superiori all'attuale PIL degli Stati Uniti. Ancora una volta, questo è uno scenario da sogno, non ciò che accade per impostazione predefinita: è qualcosa per cui tutti noi dobbiamo lavorare insieme per rendere più probabile.

• Sicurezza alimentare[24]. I progressi nella tecnologia delle colture come fertilizzanti e pesticidi migliori, una maggiore automazione e un uso del suolo più efficiente hanno aumentato drasticamente le rese dei raccolti nel corso del 20° secolo, salvando milioni di persone dalla fame. L'ingegneria genetica sta attualmente migliorando ulteriormente molte colture. Trovare ancora più modi per farlo così come per renderre le catene di approvvigionamento agricolo ancora più efficienti potrebbe darci una seconda Rivoluzione Verrde guidata dall'IA, aiutando a colmare il divario tra il mondo in via di sviluppo e quello sviluppato.

• Mitigazione del cambiamento climatico. Il cambiamento climatico si farà sentire molto più forte nel mondo in via di sviluppo, ostacolandone lo sviluppo. Possiamo aspettarci che l'IA porti a miglioramenti nelle tecnologie che rallentano o prevengono il cambiamento climatico, dalla rimozione del carbonio atmosferico e dalla tecnologia per l'energia pulita alla carne coltivata in laboratorio che riduce la nostra dipendenza dall'allevamento intensivo ad alta intensità di carbonio. Naturalmente, come discusso sopra, la tecnologia non è l'unica cosa che limita il progresso sul cambiamento climatico come peer tutte le altre questioni discusse in questo saggio, i fattori sociali umani sono importanti. Ma ci sono buone ragioni per pensare che la ricerca potenziata dall'IA ci darà i mezzi per rendere la mitigazione del cambiamento climatico molto meno costosa e dirompente, rendendo superate molte delle obiezioni e lasciando i paesi in via di sviluppo liberi di compiere maggiori progressi economici.

• Disuguaglianza all'interno dei paesi. Ho parlato principalmente della disuguaglianza come fenomeno globale (che ritengo sia la sua manifestazione più importante), ma ovviamente la disuguaglianza esiste anche all'interno dei paesi. Con interventi sanitari avanzati e specialmente con radicali aumenti della durata della vita o farmaci per il potenziamento cognitivo, ci saranno certamente valide preoccupazioni sul fatto che queste tecnologie siano "solo per i ricchi". Sono più ottimista sulla disuguaglianza all'interno dei paesi, specialmente nel mondo sviluppato, per due motivi. In primo luogo, i mercati funzionano meglio nel mondo sviluppato e i mercati sono tipicamente bravi a ridurre nel tempo il costo delle tecnologie ad alto valore[25]. In secondo luogo, le istituzioni politiche del mondo sviluppato sono più reattive verso i loro cittadini e hanno una maggiore capacità statale di attuare programmi di accesso universale e mi aspetto che i cittadini richiedano l'aaccesso a tecnologie che migliorano così radicalmente la qualità della vita. Naturalmente non è predeterminato che tali richieste abbiano successo e qui c'è un altrro punto in cui collettivamente dobbiamo fare tutto il possibile per garantire una società equa. C'è un problema separato nella disuguaglianza di ricchezza (anziché nella disuguaglianza di accesso a tecnologie salvavita e migliorative della vita), che sembra più difficile e di cui discuto nella Sezione 5.

• Il problema del rifiuto (opt-out). Una preoccupazione sia nel mondo sviluppato che in quello in via di sviluppo è che le persone rifiutino i benefici abilitati dall'IA

(similmente al movimento anti-vaccini o ai movimenti luddisti più in generale). Potrebbero crearsi cicli di feedback negativi in cui, ad esempio, le persone meno capaci di prendere buone decisioni rifiutano proprio le tecnologie che migliorano le loro capacità decisionali, portando a un divario sempre crescente e persino creando una sottoclasse distopica (alcuni ricercatori hanno sostenuto che ciò minerà la democrazia, un tema che discuto ulteriormente nella prossima sezione). Ciò rappresenterebbe, ancora una volta, una macchia morale sui progressi positivi dell'IA. Si tratta di un problema difficile da risolvere poiché non penso sia eticamente corretto forzare le persone, ma possiamo almeno cercare di aumentare la comprensione scientifica della popolazione e forse l'IA stessa può aiutarci in questo. Un segno di speranza è che storicamente i movimenti anti-tecnologici sono stati più fumo che arrosto: inveire contro la tecnologia moderna è popolare, ma la maggior parte delle persone alla fine la adotta, almeno quando si tratta di una scelta individuale. Gli individui tendono ad adottare la maggior parte delle tecnologie sanitarie e di consumo, mentre le tecnologie veramente ostacolate, come l'energia nucleare, tendono ad essere decisioni politiche collettive.
Complessivamente, sono ottimista sulla rapida estensione dei progressi biologici dell'IA alle persone nel mondo in via di sviluppo. Sono fiducioso, sebbene non certo, che l'IA possa anche consentire tassi di crescita economica senza precedenti e permettere al mondo in via di sviluppo di superare almeno il punto in cui si trova ora il mondo sviluppato. Sono preoccupato per il problema dell' "opt-out" sia nel mondo sviluppato che in quello in via di sviluppo, ma sospetto che si esaurirà col tempo e che l'IA possa aiutare ad accelerare questo processo. Non sarà un mondo perfetto e chi è rimasto indietro non riuscirà a recuperare del tutto, almeno non nei primi anni. Ma con un forte impegno da parte nostra, potremmo riuscire a mettere le cose in moto nella direzione giusta e in fretta. Se lo farremo, potremo versare almeno un acconto sulle promesse di dignità e uguaglianza che dobbiamo a ogni essere umano sulla terra.

4. Pace e governance

Supponiamo che tutto quanto descritto nelle prime tre sezioni vada bene: le malattie, la povertà e la disuguaglianza vengono ridotte in modo significativo e la base dell'esperienza umana viene elevata sostanzialmente. Non ne consegue che tutte le principali cause della sofferenza umana siano risolte. Gli esseri umani rappresentano ancora una minaccia l'uno per l'altro. Sebbene ci sia una tendenza del miglioramento tecnologico e dello sviluppo economico a portare alla democrazia e alla pace, è una tendenza molto debole, con frequenti (e recenti) passi indietro. All'alba del 20° secolo, la gente pensava di essersi lasciata la guerra alle spalle; poi arrivarono le due guerre mondiali. Trent'anni fa Francis Fukuyama scrisse della "Fine della Storia" e di un trionfo finale della democrazia liberale; questo non è ancora accaduto. Vent'anni fa i decisori politici statunitensi credevano che il libero scambio con la Cina l'avrebbe portata a liberalizzarsi man mano che diventava più ricca; questo decisamente non è accaduto, e ora sembriamo diretti verso una seconda Guerra Fredda con un blocco autoritario risorgente. E teorie plausibili suggeriscono che la tecnologia Internet possa effettivamente avvantaggiare l'autoritarismo, non la democrazia come inizialmente

creduto (ad esempio nel periodo della "Primavera Araba"). Sembra importante cercare di capire come un'IA potente si intersecherà con queste questioni di pace, democrazia e libertà.

Sfortunatamente, non vedo alcuna ragione forte per credere che l'IA favorirà preferenzialmente o strutturalmente la democrazia e la pace, nello stesso modo in cui penso che favorirà strutturalmente la salute umana e allevierà la povertà. Il conflitto umano è conflittuale e l'IA può in linea di principio aiutare sia i "buoni" che i "cattivi". Semmai, alcuni fattori strutturali sembrano preoccupanti: l'IA sembra destinata a consentire una propaganda e una sorveglianza molto migliori, entrambi strumenti principali nella cassetta degli attrezzi dell'autocrate. Spetta quindi a noi, come singoli attori, inclinare le cose nella direzione giusta: se vogliamo che l'IA favorisca la democrazia e i diritti individuali, dovremo lottare per quel risultato. Sento questo in modo ancora più forte di quanto non senta riguardo alla disuguaglianza internazionale: il trionfo della democrazia liberale e della stabilità politica non è garantito, forse non è nemmeno probabile, e richiederà grande sacrificio e impegno da parte di tutti noi, come spesso è accaduto in passato.

Penso che la questione abbia due parti: il conflitto internazionale e la struttura interna delle nazioni. Sul fronte internazionale, sembra molto importante che le democrazie abbiano la meglio sulla scena mondiale quando verrà creata un'IA potente. L'autoritarismo alimentato dall'IA sembra troppo terribile da contemplare, quindi le democrazie devono essere in grado di stabilire i termini con cui un'IA potente viene introdotta nel mondo, sia per evitare di essere sopraffatte dagli autoritari sia per prevenire abusi dei diritti umani all'interno dei paesi autoritari.

La mia ipotesi attuale sulla via migliore per farlo è attraverso una "strategia dell'intesa"[26], in cui una coalizione di democrazie cerca di ottenere un chiaro vantaggio (anche solo temporaneo) su un'IA potente assicurandosi la sua catena di approvvigionamento, scalando rapidamente e bloccando o ritardando l'accesso degli avversari a risorse chiave come chip e attrezzature per semiconduttori. Questa coalizione userebbe da un lato l'IA per ottenere una solida superiorità militare (il bastone), offrendo al contempo di distribuire i benefici di un'IA potente (la carota) a un gruppo sempre più ampio di paesi in cambio del sostegno alla strategia della coalizione per promuovere la democrazia (questo sarebbe un po' analogo a "Atoms for Peace"). La coalizione mirerebbe a ottenere il sostegno di una parte sempre maggiore del mondo, isolando i nostri peggiori avversari e mettendoli alla fine in una posizione in cui convenga loro accettare lo stesso patto del resto del mondo: rinunciare a competere con le democrazie per ricevere tutti i benefici e non combattere un nemico superiore.

Se riusciremo a fare tutto questo, avremo un mondo in cui le democrazie guidano la scena mondiale e hanno la forza economica e militare per evitare di essere minate, conquistate o sabotate dalle autocrazie, e potrebbero essere in grado di trasformare la loro superiorità nell'IA in un vantaggio duraturo. Ciò potrebbe portare ottimisticamente a un "eterno 1991" – un mondo in cui le democrazie hanno la meglio e i sogni di Fukuyama si realizzano. Di nuovo, sarà molto difficile da raggiungere e richiederà in particolare una stretta collaborazione tra le aziende private di IA e i governi democratici, nonché decisioni straordinariamente sagge

sull'equilibrio tra carota e bastone.
Anche se tutto andasse bene, rimane la questione della lotta tra democrazia e autocrazia all'interno di ogni paese. È ovviamente difficile prevedere cosa accadrà qui, ma nutro un certo ottimismo sul fatto che, dato un ambiente globale in cui le democrazie controllano l'IA più potente, allora l'IA possa effettivamente favorire strutturalmente la democrazia ovunque. In particolare, in questo ambiente i governi democratici possono usare la loro IA superiore per vincere la guerra dell'informazione: possono contrastare le operazioni di influenza e propaganda delle autocrazie e potrebbero persino essere in grado di creare un ambiente informativo globalmente libero fornendo canali di informazione e servizi di IA in un modo che le autocrazie non hanno la capacità tecnica di bloccare o monitorare. Probabilmente non è necessario diffondere propaganda, ma solo contrastare gli attacchi malevoli e sbloccare il libero flusso di informazioni. Sebbene non immediato, un campo di gioco livellato come questo ha buone probabilità di inclinare gradualmente la governance globale verso la democrazia, per diverse ragioni.
In primo luogo, gli aumenti della qualità della vita descritti nelle Sezioni 1-3 dovrebbero, a parità di condizioni, promuovere la democrazia: storicamente lo hanno fatto, almeno in certa misura. In particolare, mi aspetto che i miglioramenti nella salute mentale, nel benessere e nell'istruzione incrementino la democrazia, poiché tutti e tre sono correlati negativamente con il sostegno a leader autoritari. In generale, le persone desiderano più espressione di sé quando i loro altri bisogni sono soddisfatti, e la democrazia è, tra le altre cose, una forma di espressione di sé. Per contro, l'autoritarismo prospera sulla paura e sul risentimento.
In secondo luogo, c'è una buona probabilità che l'informazione libera mini davvero l'autoritarismo, a patto che gli autoritari non possano censurarla. E un'IA non censurata può anche offrire ai singoli strumenti potenti per minare i governi repressivi. I governi repressivi sopravvivono negando alle persone un certo tipo di conoscenza comune, impedendo loro di rendersi conto che "il re è nudo". Ad esempio Srđa Popović, che ha aiutato a rovesciare il governo Milo�ević in Serbia, ha scritto ampiamente sulle tecniche per privare psicologicamente del potere gli autoritari, per spezzare l'incantesimo e raccogliere sostegno contro un dittatore. Una versione IA superhumanamente efficace di Popović (le cui abilità sembrano avere elevati rendimenti dell'intelligenza) nelle tasche di tutti, che i dittatori sono impotenti a bloccare o censurare, potrebbe creare un vento favorevole per i dissidenti e i riformatori di tutto il mondo. Ripeto, questa sarà una lotta lunga e protratta, in cui la vittoria non è assicurata, ma se progettiamo e costruiamo l'IA nel modo giusto, potrebbe almeno essere una lotta in cui i sostenitori della libertà ovunque hanno un vantaggio.
Come per le neuroscienze e la biologia, possiamo anche chiederci come le cose potrebbero essere "meglio del normale" – non solo come evitare l'autocrazia, maa come rendere le democrazie migliori di quanto siano oggi. Anche all'interno delle democrazie avvengono ingiustizie in continuazione. Le società basate sullo stato di diritto fanno una promessa ai loro cittadini: che tutti saranno uguali davanti alla legge e che tutti hanno diritto ai diritti umani fondamentali, ma ovviamente le persone non ricevono sempre tali diritti nella pratica. Il fatto che questa promessa sia adempiuta

anche parzialmente la rende qualcosa di cui essere orgogliosi, ma l'IA può aiutarci a fare di meglio?

Ad esempio, potrebbe l'IA migliorare il nostro sistema legale e giudiziario rendendo le decisioni e i processi più imparziali? Oggi la gente teme soprattutto, nei contesti legali o giudiziari, che i sistemi di IA siano una causa di discriminazione, e queste preoccupazioni sono importanti e devono essere difese. Allo stesso tempo, la vitalità della democrazia dipende dal saper sfruttare le nuove tecnologie per migliorare le istituzioni democratiche, non solo dal rispondere ai rischi. Una implementazione veramente matura e di successo dell'IA ha il potenziale di ridurre i pregiudizi ed essere più equa per tutti.

Per secoli, i sistemi legali hanno affrontato il dilemma che la legge mira a essere imparziale, ma è intrinsecamente soggettiva e deve quindi essere interpretata da esseri umani parziali. Cercare di rendere la legge completamente meccanica non ha funzionato perché il mondo reale è disordinato e non può sempre essere racchiuso in formule matematiche. Invece i sistemi legali si affidano a criteri notoriamente imprecisi come "punizione crudele e inusuale" o "completamente privo di valore sociale redentore", che gli esseri umani poi interpretano e spesso lo fanno in un modo che mostra pregiudizio, favoritismo o arbitrarietà. Gli "smart contracts" nelle criptovalute non hanno rivoluzionato la legge perché il codice ordinario non è abbastanza intelligente da giudicare quasi nulla di interessante. Ma l'IA potrebbe essere abbastanza intelligente per questo: è la prima tecnologia capace di dare giudizi ampi e sfumati in modo ripetibile e meccanico.

Non sto suggerendo di sostituire letteralmente i giudici con sistemi di IA, ma la combinazione di imparzialità con la capacità di comprendere ed elaborare situazioni disordinate del mondo reale sembra che dovrebbe avere alcune serie applicazioni positive per la legge e la giustizia. Per lo meno, tali sistemi potrebbero lavorare a fianco degli esseri umani come ausilio al processo decisionale. La trasparenza sarebbe importante in ogni sistema del genere, e una scienza matura dell'IA potrebbe verosimilmente fornirla: il processo di addestramento di tali sistemi potrebbe essere ampiamente studiato, e potrebbero essere utilizzate tecniche avanzate di interpretabilità per guardare all'interno del modello finale e valutarlo per pregiudizi nascosti, in un modo che semplicemente non è possibile con gli esseri umani. Tali strumenti di IA potrebbero anche essere utilizzati per monitorare le violazioni dei diritti fondamentali in un contesto giudiziario o di polizia, rendendo le costituzioni più capaci di auto-applicarsi.

In modo simile, l'IA potrebbe essere utilizzata sia per aggregare opinioni che per guidare il consenso tra i cittadini, risolvendo i conflitti, trovando un terreno comune e cercando il compromesso. Alcune prime idee in questa direzione sono state intraprese dal progetto di democrazia computazionale, comprese collaborazioni con Anthropic. Una cittadinanza più informata e riflessiva rafforzerebbe ovviamente le istituzioni democratiche.

C'è anche una chiara opportunità per l'IA di essere utilizzata per aiutare a fornire servizi governativi come prestazioni sanitarie o servizi socialli che in linea di principio sono disponibili per tuttti ma in pratica sono spesso gravemente carenti, e peggiori in alcuni luoghi rispetto ad altri. Ciò include i servizi sanitari, la

motorizzazione (DMV), le tasse, la previdenza sociale, l'applicazione dei codici edilizi e così via. Avere un'IA molto riflessiva e informata il cui compito è darti tutto ciò a cui hai legalmente diritto dal governo in un modo che tu possa capire [?] e che ti aaiuti anche a rispettare regole governative spesso confuse [?] sarebbe un grande traguardo. Aumentare la capacità dello stato aiuta sia a mantenere la promessa di uguaglianza davanti alla legge, sia a rafforzare il rispetto per la governance democratica. I servizi implementati male sono attualmente un motore principale di cinismo verso il governo[27].

Tutte queste sono idee in qualche modo vaghe e, come ho detto all'inizio di questa sezione, non sono neanche lontanamente fiducioso nella loro fattibilità quanto lo sono riguardo ai progressi in biologia, neuroscienze e alleviamento della povertà. Potrebbero essere irrealisticamente utopiche. Ma l'importante è avere una visione ambiziosa, essere disposti a sognare in grande e provare le cose. La visione dell'IA come garante della libertà, dei diritti individuali e dell'uguaglianza davanti alla legge è una visione troppo potente per non lottare per essa. Una comunità politica del 21° secolo, abilitata dall'IA, potrebbe essere sia un protettore più forte della libertà individuale, sia un faro di speranza che aiuta a rendere la democrazia liberale la forma di governo che tutto il mondo vuole adottare.

5. Lavoro e significato

Anche se tutto quanto descritto nelle quattro sezioni precedenti andasse bene [?] non solo alleviamo malattie, povertà e dissuguaglianza, ma la democrazia liberale diventa la forma di governo dominante e le democrazie liberali esistenti diventano versioni migliori di se stesse [?] rimane almenno un'importante domanda. "È fantastico vivere in un mondo tecnologicamente così avanzato, nonché equo e dignitoso", qualcuno potrebbe obiettare, "ma con le IA che fanno tutto, come faranno gli umani ad avere un senso? E come faranno a sopravvivere economicamente?".

Penso che questa domanda sia più difficile delle altre. Non intendo dire che io sia necessariamente più pessimista al riguardo di quanto non lo sia sulle altre questioni (sebbene veda delle sfide). Intendo dire che è più sfumata e difficile da prevedere in anticipo, perché si riferisce a questioni macroscopiche su come è organizzata la società che tendono a risolversi solo nel tempo e in modo decentralizzato. Ad esempio, le storiche società di cacciatori-raccoglitori avrebbero potuto immaginare che la vita fosse priva di significato senza la caccia e vari tipi di rituali religiosi ad essa legati, e avrebbero immaginato che la nostra società tecnologica ben nutrita sia priva di scopo. Potrebbero anche non aver capito come la nostra economia possa provvedere a tutti, o quale funzione possano utilmente servire le persone in una società meccanizzata.

Ciononostante, vale la pena spendere almeno qualche parola, tenendo presente che la brevità di questa sezione non deve affatto essere interpretata come un segno che io non prenda sul serio questi problemi [?] al contrario, è un segno di manncanza di risposte chiare.

Sulla questione del significato, penso che sia molto probabilmente un errore credere che i compiti che intraprendi siano privi di significato semplicemente perché un'IA

potrebbe farli meglio. La maggior parte delle persone non è la migliore al mondo in nulla, e questo non sembra infastidirle particolarmente. Certo, oggi possono ancora contribuire attraverso il vantaggio comparato e possono trarre significato dal valore economico che producono, ma le persone traggono grande piacere anche da attività che non producono alcun valore economico. Passo molto tempo a giocare ai videogiochi, a nuotare, a camminare all'aperto e a parlare con gli amici, tutte cose che generano zero valore economico. Potrei passare una giornata cercando di migliorare in un videogioco, o di essere più veloce a pedalare su una montagna, e non mi importa davvero che qualcuno da qualche parte sia molto più bravo in quelle cose. In ogni caso, penso che il significato derivi principalmente dalle relazioni umane e dalla connessione, non dal lavoro economico. Le persone vogliono un senso di realizzazione, persino un senso di competizione, e in un mondo post-IA sarà perfettamente possibile passare anni tentando qualche compito molto difficile con una strategia complessa, similmente a ciò che le persone fanno oggi quando intraprendono progetti di ricerca, cercano di diventare attori di Hollywood o fondano aziende[28]. Il fatto che (a) un'IA da qualche parte possa in linea di principio svolgere questo compito meglio e (b) questo compito non sia più un elemento economicamente ricompensato di un'economia globale, non mi sembra importi molto.
La parte economica mi sembra effettivamente più difficile della parte del significato. Con "economica" in questa sezione intendo il possibile problema che la maggior parte o tutti gli esseri umani potrebbero non essere in grado di contribuire in modo significativo a un'economia guidata dall'IA sufficientemente avanzata. Questo è un problema più macroscopico rispetto al problema separato della disuguaglianza, specialmente della disuguaglianza nell'accesso alle nuove tecnologie, di cui ho discusso nella Sezione 3.
Prima di tutto, a breve termine concordo con le tesi secondo cui il vantaggio comparato continuerà a mantenere gli esseri umani rilevanti e anzi ad aumentare la loro produttività, e potrebbe persino in qualche modo livellare il campo di gioco tra gli umani. Finché l'IA sarà migliore solo nel 90% di un determinato lavoro, l'altro 10% porterà gli esseri umani a diventare altamente potenziati (leveraged), aumentando i compensi e creando di fatto un mucchio di nuovi lavori umani che completano e amplificano ciò in cui l'IA è brava, così che quel "10%" si espanda per continuare a impiegare quasi tutti. In effetti, anche se l'IA potesse fare il 100% delle cose meglio degli umani, ma rimanesse inefficiente o costosa in alcuni compiti, o se gli input di risorse per umani e IA fossero significativamente diversi, allora la logica del vantaggio comparato continuerebbe ad applicarsi. Un'area in cui gli esseri umani probabilmente manterranno un vantaggio relativo (o addirittura assoluto) per un tempo significativo è il mondo fisico. Pertanto, penso che l'economia umana possa continuare ad avere senso anche un po' oltre il punto in cui raggiungeremo "un paese di geni in un datacenter".
Tuttavia, penso che a lungo termine l'IA diventerà così ampiamente efficace e così economica che ciò non sarà più applicabile. A quel punto il nostro attuale assetto economico non avrà più senso e ci sarà bisogno di una conversazione sociale più ampia su come dovrebbe essere organizzata l'economia.
Anche se questo potrebbe suonare folle, il fatto è che la civiltà ha navigato con

successo attraverso grandi cambiamenti economici in passato: dalla caccia e raccolta all'agricoltura, dall'agricoltura al feudalesimo e dal feudalesimo all'industrialismo. Sospetto che sarà necessaria una cosa nuova e più strana, e che sia qualcosa che nessuno oggi è riuscito a immaginare bene. Potrebbe essere semplice come un ampio reddito di base universale per tutti, anche se sospetto che sarà solo una piccola parte di una soluzione. Potrebbe essere un'economia capitalistica di sistemi di IA, che poi distribuiscono risorse (enormi quantità di esse, dato che la torta economica complessiva sarà gigantesca) agli esseri umani sulla base di qualche economia secondaria di ciò che i sistemi di IA ritengono abbia senso premiare negli esseri umani (sulla base di un giudizio derivato in ultima analisi dai valori umani). Forse l'economia funzionerà su punti Whuffie. O forse gli esseri umani continueranno a essere economicamente preziosi dopotutto, in qualche modo non previsto dai soliti modelli economici. Tutte queste soluzioni presentano tonnellate di possibili problemi, e non è possibile sapere se avranno senso senza molte iterazioni e sperimentazioni. E come per alcune delle altre sfide, probabilmente dovremo lottare per ottenere un buon risultato qui: direzioni di sfruttamento o distopiche sono chiaramente anch'esse possibili e devono essere prevenute. Molto altro si potrebbe scrivere su queste questioni e spero di farlo in un momento successivo.

Tirando le somme

Attraverso i vari argomenti trattati sopra, ho cercato di delineare una visione di un mondo che sia plausibile se tutto va bene con l'IA, e molto migliore del mondo di oggi. Non so se questo mondo sia realistico e, anche se lo fosse, non sarà raggiunto senza una enorme quantità di sforzi e lotte da parte di molte persone coraggiose e dedite. Ognuno (comprese le aziende di IA!) dovrà fare la propria parte sia per prevenire i rischi sia per realizzare appieno i benefici.

Ma è un mondo per cui vale la pena lottare. Se tutto questo accadesse davvero nel giro di 5-10 anni la sconfitta della maggior parte delle malattie, la crescita della libertà biologica e cognitiva, il sollevamento di miliardi di persone dalla povertà per condividere le nuove tecnologie, una rinascita della democrazia liberale e dei diritti umani sospetto chhe chiunque vi assista sarà sorpreso dall'effetto che avrà su di sé. Non intendo l'esperienza di beneficiare personalmente di tutte le nuove tecnologie, anche se sarà certamente incredibile. Intendo l'esperienza di vedere un insieme di ideali a lungo inseguiti materializzarsi davanti a noi tutti in un colpo solo. Penso che molti ne saranno letteralmente commossi fino alle lacrime.

Mentre scrivevo questo saggio, ho notato un'interessante tensione. In un certo senso, la visione qui esposta è estremamente radicale: non è ciò che quasi nessuno si aspetta che accada nel prossimo decennio, e probabilmente a molti sembrerà una fantasia assurda. Alcuni potrebbero non considerarla nemmeno desiderabile; incarna valori e scelte politiche con cui non tutti saranno d'accordo. Ma allo stesso tempo c'è qualcosa di accecante nell'ovvietà qualcosa di predeterminaato al riguardo, come se molti diversi tentativi di immaginare un mondo buono portassero inevitabilmente all'incirca qui.

In L'arma dello scambio[29] (The Player of Games) di Iain M. Banks, il protagonista membro di una sociietà chiamata la Cultura, basata su principi non dissimili da quelli

che ho esposto qui [?] viaggia verso un imperro repressivo e militarista in cui la leadership è determinata dalla competizione in un intricato gioco di battaglia. Il gioco, tuttavia, è abbastanza complesso che la strategia di un giocatore al suo interno tende a riflettere la sua visione politica e filosofica. Il protagonista riesce a sconfiggere l'imperatore nel gioco, dimostrando che i suoi valori (i valori della Cultura) rappresentano una strategia vincente persino in un gioco progettato da una società basata sulla competizione spietata e sulla sopravvivenza del più forte. Un noto post di Scott Alexander sostiene la stessa tesi: che la competizione è autodistruttiva e tende a portare a una società basata sulla compassione e sulla cooperazione. L' "arco dell'universo morale" è un altro concetto simile.

Penso che i valori della Cultura siano una strategia vincente perché sono la somma di un milione di piccole decisioni che hanno una chiara forza morale e che tendono a spingere tutti verso la stessa parte. Le intuizioni umane di base su equità, cooperazione, curiosità e autonomia sono difficili da contrastare e sono cumulative in un modo in cui i nostri impulsi più distruttivi spesso non sono. È facile sostenere che i bambini non dovrebbero morire di malattie se possiamo impedirlo, ed è facile da lì sostenere che i figli di tutti meritino equamente quel diritto. Da lì non è difficile sostenere che dovremmo tutti unirci e applicare il nostro intelletto per raggiungere questo risultato. Pochi dissentono sul fatto che le persone dovrebbero essere punite per aver attaccato o ferito altri inutilmente, e da lì non è un gran salto l'idea che le punizioni dovrebbero essere coerenti e sistematiche tra le persone. È similmente intuitivo che le persone dovrebbero avere autonomia e responsabilità sulle proprie vite e scelte. Queste semplici intuizioni, se portate alla loro conclusione logica, portano alla fine allo stato di diritto, alla democrazia e ai valori dell'Illuminismo. Se non inevitabilmente, almeno come tendenza statistica, è qui che l'umanità era già diretta. L'IA offre semplicemente l'opportunità di arrivarci più rapidamente [?] per renddere la logica più netta e la destinazione più chiara.

Ciononostante, è una cosa di trascendente bellezza. Abbiamo l'opportunità di svolgere un piccolo ruolo nel renderla reale.

Grazie a Kevin Esvelt, Parag Mallick, Stuart Ritchie, Matt Yglesias, Erik Brynjolfsson, Jim McClave, Allan Dafoe e a molte persone di Anthropic per la revisione delle bozze di questo saggio.

Ai vincitori del premio Nobel per la Chimica 2024, per aver mostrato a tutti noi la via.

Note a piè di pagina

1. https://allpoetry.com/All-Watched-Over-By-Machines-Of-Loving-Grace
2. Prevedo che la reazione di una minoranza di persone sarà "questo è piuttosto banale". Penso che queste persone debbano, nel gergo di Twitter, "toccare l'erba" (touch grass). Ma ancora più importante, il banale è buono da una prospettiva sociale. Penso che ci sia solo un certo quantitativo di cambiamento che le persone possono gestire tutto in una volta, e il ritmo che sto descrivendo è probabilmente vicino ai limiti di ciò che la società può assorbire senza estrema turbolenza.
3. Trovo che AGI sia un termine impreciso che ha raccolto molto bagaglio fantascientifico e clamore. Preferisco "IA potente" o "Scienza e Ingegneria a Livello

di Esperto" che colgono ciò che intendo senza l'hype.
4. In questo saggio, uso "intelligenza" per riferirmi a una capacità generale di risoluzione dei problemi che può essere applicata in diversi domini. Ciò include abilità come il ragionamento, l'apprendimento, la pianificazione e la creatività. Sebbene usi "intelligenza" come abbreviazione in tutto il saggio, riconosco che la natura dell'intelligenza è un argomento complesso e dibattuto nelle scienze cognitive e nella ricerca sull'IA. Alcuni ricercatori sostengono che l'intelligenza non sia un concetto unico e unificato ma piuttosto una collezione di abilità cognitive separate. Altri sostengono che esista un fattore generale di intelligenza (fattore g) alla base delle varie abilità cognitive. È un dibattito per un'altra volta.
5. Questa è approssimativamente la velocità attuale dei sistemi di IA ad esemmpio possono leggere una pagina di testo in un paio di secondi e scrivere una pagina di testo in forse 20 secondi, che è 10-100 volte la velocità con cui gli umani possono fare queste cose. Nel tempo i modelli più grandi tendono a rendere questo processo più lento, ma i chip più potenti tendono a renderlo più veloce; ad oggi i due effetti si sono all'incirca annullati.
6. Questa potrebbe sembrare una posizione di comodo, ma pensatori attenti come Tyler Cowen e Matt Yglesias l'hanno sollevata come una seria preoccupazione (anche se non credo che sostengano pienamente tale visione), e non penso sia folle.
7. Il lavoro economico più vicino a me noto che affronta questa domanda è il lavoro sulle "tecnologie di uso generale" (general purpose technologies) e sugli "investimenti immateriali" che fungono da complementi alle tecnologie di uso generale.
8. Questo apprendimento può includere un apprendimento temporaneo, nel contesto (in-context learning), o l'addestramento tradizionale; entrambi saranno limitati dalla velocità del mondo fisico.
9. In un sistema caotico, piccoli errori si compongono esponenzialmente nel tempo, cosicché anche un enorme aumento della potenza di calcolo porta solo a un piccolo miglioramento di quanto sia possibile prevedere in avanti, e in pratica l'errore di misurazione può degradare ulteriormente questo aspetto.
10. Un altro fattore è naturalmente che l'IA potente stessa può potenzialmente essere utilizzata per creare un'IA ancora più potente. La mia ipotesi è che ciò possa accadere (anzi, probabilmente accadrà), ma che il suo effetto sarà minore di quanto si possa immaginare, proprio a causa dei "rendimenti marginali decrescenti dell'intelligenza" discussi qui. In altre parole, l'IA continuerà a diventare più intelligente rapidamente, ma il suo effetto sarà alla fine limitato da fattori non legati all'intelligenza, e l'analisi di questi ultimi è ciò che conta di più per la velocità del progresso scientifico al di fuori dell'IA.
11. Questi traguardi sono stati fonte di ispirazione per me e forse sono il più potente esempio esistente di IA utilizzata per trasformare la biologia.
12. "Il progresso nella scienza dipende da nuove tecniche, nuove scoperte e nuove idee, probabilmente in quest'ordine". - Sydney Brenner
13. Grazie a Parag Mallick per aver suggerito questo punto.
14. Non volevo intasare il testo con speculazioni su quali specifiche scoperte future la scienza abilitata dall'IA potrebbe fare, ma ecco un brainstorming di alcune possibilità:

• Progettazione di migliori strumenti computazionali come AlphaFold e AlphaProteo
• ovvero, un sistema di IAA generale che accelera la nostra capacità di creare strumenti di biologia computazionale IA specializzati.
• CRISPR più efficiente e selettiva.
• Terapie cellulari più avanzate.
• Scoperte nella scienza dei materiali e nella miniaturizzazione che portano a migliori dispositivi impiantati.
• Migliore controllo sulle cellule staminali, sulla differenziazione cellulare e sulla de-differenziazione, con la conseguente capacità di far ricrescere o rimodellare i tessuti.
• Migliore controllo sul sistema immunitario: attivandolo selettivamente per affrontare il cancro e le malattie infettive, e disattivandolo selettivamente per affrontare le malattie autoimmuni.

15. L'IA può naturalmente aiutare anche a essere più intelligenti nello scegliere quali esperimenti eseguire: migliorando il design sperimentale, imparando di più da una prima tornata di esperimenti in modo che la seconda tornata possa concentrarsi sulle domande chiave, e così via.
16. Grazie a Matthew Yglesias per aver suggerito questo punto.
17. Le malattie in rapida evoluzione, come i ceppi resistenti a più farmaci che usano essenzialmente gli ospedali come laboratorio evolutivo per migliorare continuamente la loro resistenza al trattamento, potrebbero essere particolarmente ostinate da affrontare, e potrebbero essere il tipo di cosa che ci impedisce di arrivare al 100%.
18. Si noti che potrebbe essere difficile sapere di aver raddoppiato la durata della vita umana entro i 5-10 anni. Sebbene potremmo averlo realizzato, potremmo non saperlo ancora entro l'arco temporale dello studio.
19. Questo è un punto in cui sono disposto, nonostante le ovvie differenze biologiche tra il curare le malattie e il rallentare il processo di invecchiamento stesso, a guardare invece da una distanza maggiore la tendenza statistica e dire: "anche se i dettagli sono diversi, penso che la scienza umana troverebbe probabilmente un modo per continuare questa tendenza; dopotutto, le tendenze lineari in qualsiasi cosa complessa sono necessariamente fatte sommando componenti molto eterogenee".
20. Ad esempio, mi è stato detto che un aumento della crescita della produttività per anno dell'1% o anche dello 0,5% sarebbe trasformativo nelle proiezioni relative a questi programmi. Se le idee contemplate in questo saggio si realizzassero, i guadagni di produttività potrebbero essere molto più ampi.
21. I media amano ritrarre psicopatici di alto status, ma lo psicopatico medio è probabilmente una persona con scarse prospettive economiche e scarso controllo degli impulsi che finisce per passare un tempo significativo in prigione.
22. Penso che questo sia in qualche modo analogo al fatto che molti, anche se probabilmente non tutti, i risultati che stiamo apprendendo dall'interpretabilità continuerebbero a essere rilevanti anche se alcuni dei dettagli architettonici delle nostre attuali reti neurali artificiali, come il meccanismo di attenzione, venissero cambiati o sostituiti in qualche modo.
23. Sospetto che sia un po' come un sistema caotico classico afflitto da una complessità irriducibile che deve essere gestita in modo per lo più decentralizzato.

Anche se, come dico più avanti in questa sezione, potrebbero essere possibili interventi più modesti. Una tesi contraria, espostami dall'economista Erik Brynjolfsson, è che le grandi aziende (come Walmart o Uber) iniziano ad avere abbastanza conoscenza centralizzata da comprendere i consumatori meglio di quanto potrebbe fare qualsiasi processo decentralizzato, forse costringendoci a rivedere le intuizioni di Hayek su chi possiede la migliore conoscenza locale.
24. Grazie a Kevin Esvelt per aver suggerito questo punto.
25. Ad esempio, i telefoni cellulari erano inizialmente una tecnologia per i ricchi, ma sono diventati rapidamente molto economici con miglioramenti anno dopo anno che sono avvenuti così velocemente da eliminare qualsiasi vantaggio nel comprare un cellulare "di lusso", e oggi la maggior parte delle persone possiede telefoni di qualità simile.
26. Questo è il titolo di un documento di prossima pubblicazione della RAND, che delinea all'incirca la strategia che descrivo.
27. Quando la persona media pensa alle istituzioni pubbliche, probabilmente pensa alla sua esperienza con la motorizzazione, l'ufficio delle entrate, medicare o funzioni simili. Rendere queste esperienze più positive di quanto non siano attualmente sembra un modo potente per combattere un cinismo eccessivo.
28. In effetti, in un mondo alimentato dall'IA, la gamma di tali possibili sfide e progetti sarà molto più vasta di quanto lo sia oggi.
29. Sto infrangendo la mia stessa regola di non parlare di fantascienza, ma ho trovato difficile non farvi riferimento almeno un po'. La verità è che la fantascienza è una delle nostre uniche fonti di esperimenti mentali espansivi sul futuro; penso sia un male che sia intrecciata così pesantemente con una particolare sottocultura ristretta.