Come l'intelligenza artificiale potrebbe trasformare il mondo in meglio

Ottobre 2024

Penso e parlo molto dei rischi di un'intelligenza artificiale potente. L'azienda di cui sono CEO, Anthropic, fa molte ricerche su come ridurre questi rischi. Per questo motivo, a volte le persone traggono la conclusione che io sia un pessimista o un "pessimista" che pensa che l'intelligenza artificiale sarà per lo più negativa o pericolosa. Non la penso affatto così. In effetti, uno dei motivi principali per cui mi concentro 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 potrebbe essere il lato positivo dell'intelligenza artificiale** , proprio come penso che la maggior parte delle persone stia sottovalutando quanto gravi potrebbero essere i rischi.

In questo saggio cerco di abbozzare come potrebbe apparire questo lato positivo, come potrebbe apparire un mondo con una potente IA se tutto andasse *per il verso giusto* . Ovviamente nessuno può conoscere il futuro con certezza o precisione, e gli effetti di una potente IA saranno probabilmente ancora più imprevedibili dei cambiamenti tecnologici passati, quindi tutto questo sarà inevitabilmente costituito da ipotesi. Ma il mio obiettivo è almeno di ipotesi istruite e utili, che catturino il sapore di ciò che accadrà anche se la maggior parte dei dettagli finisse per essere sbagliata. Sto includendo molti dettagli principalmente perché penso che una visione concreta faccia di più per far progredire la discussione rispetto a una altamente elusiva e astratta.

Prima, tuttavia, volevo spiegare brevemente perché io e Anthropic non abbiamo parlato molto dei vantaggi dell'IA potente e perché probabilmente continueremo, nel complesso, a parlare molto dei rischi. In particolare, ho fatto questa scelta perché desidero:

- **Massimizza la leva finanziaria** . Lo sviluppo di base della tecnologia AI e molti (non tutti) dei suoi benefici sembrano inevitabili (a meno che i rischi non facciano deragliare tutto) ed è fondamentalmente guidato da potenti forze di mercato. D'altro canto, i rischi non sono predeterminati e le nostre azioni possono cambiare notevolmente la loro probabilità.

- **Evita la percezione della propaganda** . Le aziende di intelligenza artificiale che parlano di tutti i fantastici vantaggi dell'intelligenza artificiale possono sembrare propagandiste o come se stessero cercando di distrarre dagli svantaggi. Penso anche che, per principio, sia dannoso per la tua anima passare troppo tempo a "parlare del tuo libro".

- **Evitate la grandiosità** . Spesso mi disgusta il modo in cui molte figure pubbliche del rischio AI (per non parlare dei leader delle aziende AI) parlano del mondo post-AGI, come se la loro missione fosse quella di realizzarlo da sole, come un profeta che guida il suo popolo alla salvezza. Penso che sia pericoloso vedere le aziende come se plasmassero unilateralmente il mondo, e pericoloso vedere gli obiettivi tecnologici pratici in termini essenzialmente religiosi.

- **Evitate il bagaglio "fantascientifico"** . Sebbene io pensi che la maggior parte delle persone sottovaluti il lato positivo di un'intelligenza artificiale potente, la piccola comunità di persone che discute di futuri radicali dell'intelligenza artificiale spesso lo fa in un tono eccessivamente "fantascientifico" (con ad esempio menti caricate, esplorazione spaziale o vibrazioni cyberpunk generali). Penso che questo faccia sì che le persone prendano meno sul serio le affermazioni e le impregnano di una sorta di irrealtà. Per essere chiari, il problema non è se le tecnologie descritte siano possibili o probabili (il saggio principale discute questo in dettaglio granulare), è più che la "vibrazione" introduce connotativamente un mucchio di bagaglio culturale e ipotesi non dichiarate su quale tipo di futuro sia desiderabile, come si svilupperanno vari problemi sociali, ecc. Il risultato finisce spesso per essere letto come una fantasia per una sottocultura ristretta, mentre risulta 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 buono con un'intelligenza artificiale potente, mentre facciamo del nostro meglio per evitare le insidie di cui sopra. In effetti, penso che sia fondamentale avere una visione del futuro autenticamente stimolante, e non *solo* un piano per combattere gli incendi. Molte delle implicazioni dell'intelligenza artificiale potente sono avverse o pericolose, ma alla fine di tutto, deve esserci qualcosa *per cui* stiamo lottando , un risultato positivo in cui tutti stanno meglio,

qualcosa che spinga le persone a superare i loro litigi e ad affrontare le sfide future. La paura è un tipo di motivatore, ma non è sufficiente: abbiamo bisogno anche di speranza.

L'elenco delle applicazioni positive dell'intelligenza artificiale potente è estremamente lungo (e include robotica, produzione, energia e molto altro), ma mi concentrerò su un piccolo numero di aree che mi sembrano avere il potenziale maggiore per migliorare direttamente la qualità della vita umana. Le cinque categorie che mi entusiasmano di più 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 secondo la maggior parte degli standard (a parte le visioni di "singolarità" fantascientifiche [2]), ma le intendo sinceramente e seriamente. Tutto ciò che sto dicendo potrebbe essere facilmente sbagliato (per ripetere il mio punto di cui sopra), 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 potrebbe significare in pratica. Sono fortunato ad avere esperienza professionale sia in biologia che in neuroscienze , e sono un dilettante informato nel campo dello sviluppo economico, ma sono sicuro che sbaglierò un sacco di cose. Una cosa che ho capito scrivendo questo saggio è che sarebbe prezioso riunire un gruppo di esperti del settore (in biologia, economia, relazioni internazionali e altri settori) per scrivere una versione molto migliore e più informata di ciò che ho prodotto qui. Probabilmente è meglio considerare i miei sforzi qui come uno spunto di partenza per quel gruppo.

Presupposti e quadro di base

Per rendere l'intero saggio più preciso e fondato, è utile specificare chiaramente cosa intendiamo per IA potente (vale a dire la soglia oltre la quale inizia il conto alla rovescia dei 5-10 anni), nonché definire un quadro per riflettere sugli effetti di tale IA una volta che sarà presente.

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

Con *"intelligenza artificiale potente"* intendo un modello di intelligenza artificiale, probabilmente simile nella forma agli LLM odierni, anche se potrebbe essere basato su un'architettura diversa, potrebbe coinvolgere diversi modelli interagenti e potrebbe essere addestrato in modo diverso, con le seguenti proprietà:

- In termini di pura intelligenza4 [4] è più intelligente di un 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 belli, scrivere basi di codice difficili da zero, ecc.
- Oltre a essere semplicemente una "cosa intelligente con cui parli", ha tutte le "interfacce" disponibili per un essere umano che lavora virtualmente, tra cui testo, audio, video, controllo di mouse e tastiera e accesso a Internet. Può impegnarsi in qualsiasi azione, comunicazione o operazione remota abilitata da questa interfaccia, tra cui intraprendere azioni su Internet, dare o ricevere indicazioni agli esseri umani, ordinare materiali, dirigere esperimenti, guardare video, realizzare video e così via. Esegue tutti questi compiti con, ancora una volta, un'abilità che supera quella degli esseri umani più capaci al mondo.

- Non si limita a rispondere passivamente alle domande; al contrario, può affidargli compiti che richiedono ore, giorni o settimane per essere completati, per poi svolgerli autonomamente, come farebbe un dipendente intelligente, chiedendo chiarimenti se necessario.

- Non ha una forma fisica (a parte quella di vivere sullo schermo di un computer), ma può controllare strumenti fisici, robot o apparecchiature di laboratorio esistenti tramite un computer; in teoria potrebbe persino progettare robot o apparecchiature da utilizzare autonomamente.

- Le risorse utilizzate per addestrare il modello possono essere riutilizzate per *eseguirne* milioni di istanze (ciò corrisponde alle dimensioni dei cluster previste entro ~2027) e il modello può assorbire informazioni e generare azioni a una velocità umana di circa 10x-100x [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 in modo indipendente su compiti non correlati oppure, se necessario, possono lavorare tutte insieme nello stesso modo in cui collaborerebbero gli esseri umani, magari con diverse sottopopolazioni appositamente adattate per svolgere compiti particolari.

Potremmo riassumerlo come un “paese di geni in un data center”.

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. Innanzitutto, potresti pensare che il mondo verrebbe trasformato all'istante su scala di secondi o giorni (" la Singolarità "), poiché un'intelligenza superiore si basa su se stessa e risolve ogni possibile compito scientifico, ingegneristico e operativo quasi immediatamente. Il problema con questo è che ci sono dei veri limiti fisici e pratici, 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 di fata magica.

In secondo luogo, e viceversa, potresti credere che il progresso tecnologico sia saturo o limitato in velocità dai dati del mondo reale o da fattori sociali, e che un'intelligenza superiore a quella umana aggiungerà ben poco [6] . Questo mi sembra ugualmente inverosimile: riesco a pensare a centinaia di problemi scientifici o persino sociali in cui un vasto gruppo di persone davvero intelligenti accelererebbe drasticamente il progresso, soprattutto se non si limitassero all'analisi e potessero far accadere le cose nel mondo reale (cosa che il nostro presunto paese di geni può fare, anche dirigendo o assistendo team di esseri umani).

Penso che la verità sia probabilmente una mescolanza disordinata di queste due immagini estreme, qualcosa che varia a seconda del compito e del campo ed è molto sottile nei suoi dettagli. Credo che abbiamo bisogno di nuovi framework per pensare a questi dettagli in modo produttivo.

Gli economisti parlano spesso di "fattori di produzione": cose come lavoro, terra e capitale. L'espressione "rendimenti marginali di lavoro/terra/capitale" cattura l'idea che in una data situazione, un dato fattore può o non può essere quello limitante: ad esempio, un'aeronautica ha bisogno sia di aerei che di piloti, e assumere più piloti non aiuta molto se si è a corto di aerei. Credo che nell'era dell'intelligenza artificiale, dovremmo parlare dei *rendimenti marginali dell'intelligenza* [7] e cercare di capire quali sono gli altri fattori che sono complementari all'intelligenza e che diventano fattori limitanti quando l'intelligenza è molto alta. Non siamo abituati a pensare in questo modo, a chiederci "quanto essere più intelligenti aiuta in questo compito e in quale scala temporale?", ma sembra il modo giusto per concettualizzare un mondo con un'intelligenza artificiale molto potente.

La mia ipotesi sull'elenco dei fattori che limitano o sono complementari all'intelligenza include:

- **Velocità del mondo esterno** . Gli agenti intelligenti devono operare in modo interattivo nel mondo per realizzare le cose e anche per imparare [8] . Ma il mondo si muove solo così velocemente. Le cellule e gli animali corrono a una velocità fissa, quindi gli esperimenti su di essi richiedono una certa quantità di tempo che potrebbe essere irriducibile. Lo stesso vale per l'hardware, la scienza dei materiali, qualsiasi cosa che implichi la comunicazione con le persone e persino la nostra infrastruttura software esistente. Inoltre, nella scienza molti esperimenti sono spesso necessari in sequenza, ognuno imparando o basandosi sul precedente. Tutto ciò significa che la velocità con cui un progetto importante, ad esempio lo sviluppo di una cura per il cancro, può essere completato potrebbe avere un minimo irriducibile che non può essere ulteriormente ridotto anche se l'intelligenza continua ad aumentare.

- **Necessità di dati** . A volte mancano dati grezzi e in loro assenza una maggiore intelligenza non aiuta. Gli attuali fisici delle particelle sono molto ingegnosi e hanno sviluppato un'ampia gamma di teorie, ma non hanno i dati per scegliere tra di esse perché i dati degli acceleratori di particelle sono così limitati . Non è chiaro se farebbero molto meglio se fossero superintelligenti, a parte forse accelerando la costruzione di un acceleratore più grande.

- **Complessità intrinseca** . Alcune cose sono intrinsecamente imprevedibili o caotiche e persino l'IA più potente non può prevederle o districarle sostanzialmente meglio di un essere umano o di un computer oggi. Ad esempio, persino 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 esseri umani e ai computer di oggi.
- **Limitazioni da parte degli umani** . Molte cose non possono essere fatte senza infrangere le leggi, danneggiare gli umani o mandare in tilt 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 .

- **Leggi fisiche** . Questa è una versione più cruda del primo punto. Ci sono alcune leggi fisiche che sembrano essere infrangibili. Non è possibile viaggiare più veloci della luce. Il budino non si smuove . 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 di calcolo nel mondo.

C'è un'ulteriore distinzione basata sulle *scale temporali* . Le cose che sono vincoli rigidi nel breve periodo possono diventare più malleabili all'intelligenza nel lungo periodo. Ad esempio, l'intelligenza potrebbe essere utilizzata per sviluppare un nuovo paradigma sperimentale che ci consenta di apprendere *in vitro* ciò che prima 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 limiti etici) trovare modi per aggirare i vincoli basati sull'uomo (ad esempio, aiutando a migliorare il sistema di sperimentazione clinica, aiutando a creare nuove giurisdizioni in cui le sperimentazioni cliniche hanno meno burocrazia, o migliorando la scienza stessa per rendere le sperimentazioni cliniche sull'uomo meno necessarie o meno costose).

Pertanto, dovremmo immaginare un quadro in cui l'intelligenza è inizialmente pesantemente ostacolata 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 tutto ciò accade e in quale ordine.

Tenendo presente quanto sopra esposto, cercherò di rispondere a questa domanda per i cinque ambiti menzionati nell'introduzione.

1. Biologia e salute

La biologia è probabilmente l'area in cui il progresso scientifico ha il potenziale maggiore 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 ancora, e sconfiggerle sarebbe un'enorme conquista umanitaria. 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 umana 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'applicare direttamente l'intelligenza alla biologia sono i dati, la velocità del mondo fisico e la complessità intrinseca (in effetti, tutti e tre sono correlati tra loro). Anche i vincoli umani giocano un ruolo in una fase successiva, quando sono coinvolti gli studi clinici. Analizziamoli 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 avvengano reazioni chimiche, e questo può talvolta richiedere giorni o addirittura settimane, senza un modo ovvio per accelerarlo. Gli esperimenti sugli

animali possono richiedere mesi (o più) e gli esperimenti sugli esseri umani spesso richiedono anni (o persino decenni per studi sui risultati a lungo termine). In un certo senso correlato a questo, i dati sono spesso carenti, non tanto in quantità, quanto in qualità: c'è sempre una carenza di dati chiari e inequivocabili 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 qualche effetto (invece di dedurne le conseguenze in qualche modo indiretto o rumoroso). Anche i dati molecolari quantitativi e massivi, come i dati proteomici che ho raccolto mentre lavoravo sulle tecniche di spettrometria di massa, sono rumorosi e perdono 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 hai mai visto un diagramma che mostra la biochimica del metabolismo umano , saprai 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 che ci vuole per condurre un esperimento sugli esseri umani, gli studi clinici veri e propri comportano molta burocrazia e requisiti normativi che (secondo l'opinione di molte persone, me compreso) aggiungono tempo aggiuntivo non necessario e ritardano i progressi.

Considerato tutto ciò, molti biologi sono da tempo scettici sul valore dell'intelligenza artificiale 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 certo successo, ma non hanno avuto l'impatto veramente trasformativo inizialmente sperato. Parte dello scetticismo è stato ridotto da importanti e rivoluzionarie scoperte come AlphaFold (che ha appena meritatamente fatto vincere ai suoi creatori il premio Nobel per la chimica) e AlphaProteo [11] , ma c'è ancora la percezione che l'intelligenza artificiale sia (e continuerà a essere) utile solo in un insieme limitato di circostanze. Una formulazione comune è "L'intelligenza artificiale può fare un lavoro migliore nell'analizzare i tuoi dati, ma non può produrre più dati o migliorare la qualità dei dati. Garbage in, garbage out".

Ma penso che questa prospettiva pessimistica stia pensando all'IA nel modo sbagliato. Se la nostra ipotesi fondamentale 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 esegue *tutti* i compiti dei biologi, tra cui progettare ed eseguire esperimenti nel mondo reale (controllando robot di laboratorio o semplicemente dicendo agli umani quali esperimenti eseguire, come farebbe un ricercatore principale con i suoi studenti laureati), inventando nuovi metodi biologici o tecniche di misurazione e così via. È accelerando *l'intero processo di ricerca* che l'IA può davvero accelerare la biologia. **Voglio ripeterlo perché è l'idea sbagliata più comune che emerge quando parlo della capacità dell'IA di trasformare la biologia: *non* sto parlando dell'IA come di un semplice 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 l'accelerazione possa provenire, una frazione sorprendentemente ampia del progresso in biologia è derivata da un numero davvero esiguo di scoperte, spesso correlate a strumenti o tecniche di misurazione generali [12] che consentono un intervento preciso ma generalizzato o programmabile nei sistemi biologici. Ci sono forse circa 1 di queste importanti scoperte all'anno e collettivamente si può sostenere che guidano >50% del progresso in biologia. Queste scoperte sono così potenti proprio perché tagliano attraverso la complessità intrinseca e le limitazioni dei dati, aumentando direttamente la nostra comprensione e il nostro controllo sui processi biologici. Poche scoperte ogni decennio hanno permesso sia la maggior parte della nostra comprensione scientifica di base della biologia, sia hanno guidato molti dei trattamenti medici più potenti.

Ecco alcuni esempi:

- CRISPR : una tecnica che consente la modifica in tempo reale di qualsiasi gene negli organismi viventi (sostituzione di qualsiasi sequenza genica arbitraria con qualsiasi altra sequenza arbitraria). Da quando è stata sviluppata la tecnica originale, ci sono stati miglioramenti costanti per colpire specifici tipi di cellule, aumentando la precisione e riducendo le modifiche del gene sbagliato, tutti necessari per un uso sicuro negli esseri 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, i cui costi sono diminuiti di diversi ordini di grandezza negli ultimi due decenni.

- Tecniche optogenetiche che consentono di far attivare un neurone illuminandolo con 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 diventati famosi durante il COVID).
- Terapie cellulari come la CAR-T che consentono di estrarre le cellule immunitarie dal corpo e di "riprogrammarle" per attaccare, in linea di principio, qualsiasi cosa.
- Approfondimenti concettuali come la teoria dei germi come causa delle malattie o la scoperta di un legame tra il sistema immunitario e il cancro [13] .

Mi prendo 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 talentuosi e creativi** . O, per dirla in un altro modo, **penso che i ritorni all'intelligenza siano alti per queste scoperte** e che tutto il resto in biologia e medicina ne derivi principalmente.

Perché la penso così? Per le risposte ad alcune domande che dovremmo abituarci a porci quando cerchiamo di determinare i "rendimenti dell'intelligenza". In primo luogo, queste scoperte sono generalmente fatte da un numero esiguo di ricercatori, spesso le stesse persone ripetutamente, il che suggerisce abilità e non una ricerca casuale (quest'ultima potrebbe suggerire che i lunghi esperimenti siano il fattore limitante). In secondo luogo, spesso "avrebbero potuto essere fatte" anni prima di quanto non siano state: ad esempio, CRISPR era un componente naturale del sistema immunitario nei batteri noto fin dagli anni '80 , ma ci sono voluti altri 25 anni perché le persone si rendessero conto che poteva essere riutilizzato per l'editing genetico generale. Spesso vengono anche ritardate di molti anni dalla mancanza di supporto da parte della comunità scientifica per direzioni promettenti (vedi questo profilo sull'inventore dei vaccini a mRNA; storie simili abbondano). In terzo luogo, i progetti di successo sono spesso frammentari o sono stati ripensamenti che le persone inizialmente non pensavano fossero 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" (è necessario fare prima la scoperta A per avere gli strumenti o le conoscenze per fare la scoperta B) - il che potrebbe di nuovo creare ritardi sperimentali - molte, forse la maggior parte, sono indipendenti, il che significa che molte contemporaneamente possono essere elaborate in parallelo. Sia questi fatti, sia la mia esperienza generale come 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 conoscenza biologica che l'umanità possiede (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 attentamente progettati, fornisce una prova di principio (anche se con uno strumento ristretto in un dominio ristretto) che dovrebbe indicare la strada da seguire.

Quindi, suppongo che un'intelligenza artificiale potente potrebbe almeno 10 volte la velocità di queste scoperte, dandoci 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 la prima volta, inclusi esperimenti sugli animali e cose come la progettazione di microscopi o costose strutture di laboratorio. In realtà sono aperto all'idea (forse assurda) che potremmo ottenere *1000* anni di progresso in 5-10 anni, ma molto scettico sul fatto che possiamo ottenere 100 anni in 1 anno. Un altro modo di dirlo è che penso che ci sia un ritardo costante inevitabile: gli esperimenti e la progettazione hardware hanno una certa "latenza" e devono essere iterati su un certo numero "irriducibile" di volte per apprendere cose che non possono essere dedotte logicamente. Ma in aggiunta a ciò potrebbe essere possibile un parallelismo massiccio [15] .

E che dire delle sperimentazioni cliniche? Sebbene vi siano molta burocrazia e rallentamenti associati, la verità è che molta (anche se non tutta!) della loro lentezza deriva in ultima analisi dalla necessità di valutare rigorosamente farmaci che funzionano a malapena o in modo ambiguo. Questo è tristemente vero per la maggior parte delle terapie odierne: il farmaco antitumorale medio aumenta la sopravvivenza di alcuni mesi, pur avendo effetti collaterali significativi che devono essere attentamente misurati (c'è una storia simile per i farmaci per l'Alzheimer). Ciò porta a studi enormi (al fine di ottenere potenza statistica) e difficili compromessi che le agenzie di regolamentazione in genere non sono brave a fare, ancora una volta a causa della burocrazia e della complessità degli interessi in competizione.

Quando qualcosa funziona davvero bene, va molto più veloce: c'è un percorso di approvazione accelerato e la facilità di approvazione è molto maggiore quando le dimensioni dell'effetto sono maggiori. I vaccini a mRNA per il COVID sono stati approvati in 9 mesi, molto più velocemente del solito. Detto questo, anche in queste condizioni gli studi clinici sono ancora troppo lenti: i vaccini a mRNA *avrebbero dovuto* essere approvati in circa 2 mesi . Ma questo tipo di ritardi (circa 1 anno end-to-end per un farmaco) combinati con una parallelizzazione massiccia e la necessità di alcune iterazioni ma non troppe ("alcuni tentativi") sono molto compatibili con una trasformazione radicale in 5-10 anni. Ancora più ottimisticamente, è possibile che la scienza biologica abilitata dall'intelligenza artificiale ridurrà la necessità di iterazione negli studi clinici sviluppando modelli sperimentali animali e cellulari migliori (o persino simulazioni) che sono 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 in cui abbiamo bisogno di un ciclo di iterazione più rapido.

Infine, per quanto riguarda gli studi clinici e le barriere sociali, vale la pena sottolineare esplicitamente che, in un certo senso, le innovazioni biomediche hanno una storia insolitamente *solida* di implementazione con successo, a differenza di altre tecnologie [16] . Come accennato nell'introduzione, molte tecnologie sono ostacolate da fattori sociali nonostante funzionino bene dal punto di vista tecnico. Ciò potrebbe suggerire una prospettiva pessimistica su ciò che l'IA può realizzare . Ma la biomedicina è unica in quanto, sebbene il processo di sviluppo dei farmaci sia eccessivamente macchinoso, una volta sviluppati, 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'intelligenza artificiale ci consentiranno di comprimere i progressi che i biologi umani avrebbero raggiunto nei prossimi 50-100 anni in 5-10 anni. Mi riferirò a questo come al "XXI secolo compresso": l'idea che dopo che sarà sviluppata un'intelligenza artificiale potente, in pochi anni faremo tutti i progressi in biologia e medicina che avremmo fatto nell'intero XXI secolo.

Sebbene prevedere cosa una potente IA possa fare 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 realizzato nel XX secolo, o estrapolando dai primi 2 decenni del XXI, o chiedendoci cosa ci porterebbero "10 CRISPR e 50 CAR-T", tutti offrono modi pratici e fondati per stimare il livello generale di progresso che potremmo aspettarci da una potente IA.

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

- **Prevenzione e trattamento affidabili di quasi tutte** [ie-17] **malattie infettive naturali.** Dati gli enormi progressi contro le malattie infettive nel XX secolo, non è radicale immaginare che potremmo più o meno "finire il lavoro" in un XXI secolo compresso. I vaccini a mRNA e tecnologie simili indicano già la strada verso " vaccini per qualsiasi cosa ". Se le malattie infettive saranno *completamente sradicate dal mondo* (invece che solo in alcuni luoghi) dipende da questioni su povertà e disuguaglianza, che sono discusse nella Sezione 3.
- **Eliminazione della maggior parte dei tumori** . I tassi di mortalità per cancro sono scesi 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 i farmaci molto selettivi che prendono di mira il cancro nella sua fase iniziale e ne impediscono la crescita. L'intelligenza artificiale renderà inoltre possibili regimi di trattamento molto finemente adattati al genoma individualizzato del cancro: sono possibili oggi, ma estremamente costosi in termini di tempo e competenza umana, che l'intelligenza artificiale dovrebbe consentirci di scalare. Riduzioni del 95% o più sia nella mortalità che nell'incidenza sembrano possibili. Detto questo, il cancro è estremamente vario e adattabile, ed è probabilmente la più difficile di queste malattie da distruggere completamente. Non sarebbe sorprendente se persistesse un assortimento di neoplasie rare e difficili.

- **Prevenzione molto efficace e cure efficaci per le malattie genetiche** . Uno screening embrionale notevolmente migliorato probabilmente renderà possibile prevenire la maggior parte delle malattie genetiche, e un discendente più sicuro e affidabile di CRISPR potrebbe curare la maggior parte delle malattie genetiche nelle persone esistenti. Tuttavia, le afflizioni dell'intero corpo che colpiscono una grande frazione di cellule potrebbero essere gli ultimi resistenti.

- **Prevenzione dell'Alzheimer** . Abbiamo avuto molte difficoltà a capire cosa causa 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 migliori strumenti di misurazione che isolano gli effetti biologici; quindi sono ottimista sulla capacità dell'IA di risolverlo. Ci sono buone probabilità che alla fine possa essere prevenuto con interventi relativamente semplici, una volta che avremo effettivamente capito cosa sta succedendo. Detto questo, il danno da Alzheimer già esistente potrebbe essere molto difficile da invertire.

- **Trattamento migliorato della maggior parte delle altre malattie** . Questa è una categoria onnicomprensiva per altre malattie, tra cui diabete, obesità, malattie cardiache, malattie autoimmuni e altro ancora. La maggior parte di queste sembra "più facile" da risolvere rispetto al cancro e all'Alzheimer e in molti casi è già in forte declino. Ad esempio, i decessi per malattie cardiache sono già diminuiti di oltre il 50% e semplici interventi come gli agonisti del GLP-1 hanno già fatto enormi progressi contro l'obesità e il diabete.

- **Libertà biologica** . Gli ultimi 70 anni hanno visto progressi nel controllo delle nascite, nella fertilità, nella gestione del peso e molto altro. Ma sospetto che la biologia accelerata dall'intelligenza artificiale amplierà notevolmente ciò che è possibile: peso, aspetto fisico, riproduzione e altri processi biologici saranno completamente sotto il controllo delle persone. Faremo riferimento a questi sotto la voce *libertà biologica:* l'idea che tutti dovrebbero essere autorizzati a scegliere cosa vogliono diventare e vivere la propria vita nel modo che più li attrae. Ci saranno ovviamente importanti questioni sull'uguaglianza globale di accesso; vedere la Sezione 3 per queste.

- **Raddoppio della durata della vita umana** [18] **.** Questo potrebbe sembrare radicale, ma l'aspettativa di vita è aumentata di quasi 2 volte nel XX secolo (da ~40 anni a ~75), quindi è "di tendenza" che il "XXI compresso" la raddoppierà di nuovo a 150. Ovviamente gli interventi coinvolti nel rallentamento dell'effettivo processo di invecchiamento saranno diversi da quelli che erano necessari nel secolo scorso per prevenire (per lo più infantili) morti premature 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 tartaruga) vivono già 200 anni, quindi gli esseri umani non sono palesemente a un limite superiore teorico. A occhio e croce, la cosa più importante di cui c'è bisogno potrebbero essere biomarcatori affidabili, non Goodhart-abili, dell'invecchiamento umano, poiché ciò consentirà una rapida iterazione di 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 da consentire alla maggior parte delle persone attualmente in vita di vivere quanto desiderano, anche se non vi è alcuna garanzia che ciò sia biologicamente possibile.

Vale la pena dare un'occhiata a questa lista e riflettere su quanto sarà diverso il mondo se tutto questo verrà realizzato tra 7-12 anni (il che sarebbe in linea con una tempistica di IA aggressiva). Inutile dire che sarebbe un trionfo umanitario inimmaginabile, l'eliminazione in una volta sola della maggior parte dei flagelli che hanno tormentato l'umanità per millenni. Molti dei miei amici e colleghi stanno crescendo dei bambini e quando questi bambini cresceranno, spero che qualsiasi menzione di malattia suonerà loro come scorbuto, vaiolo o peste bubbonica suonano a noi. Quella generazione trarrà beneficio anche da una maggiore libertà biologica e autoespressione e, con un po' di fortuna, potrebbe anche essere in grado di vivere finché vuole.

È difficile sopravvalutare quanto questi cambiamenti saranno sorprendenti per tutti, tranne che per la piccola comunità di persone che si aspettavano una potente IA. Ad esempio, migliaia di economisti ed esperti di politica negli Stati Uniti stanno attualmente discutendo su come mantenere solventi la previdenza sociale e Medicare e, più in generale, su come contenere i costi dell'assistenza sanitaria (che è per lo più consumata da chi ha più di 70 anni e in particolar modo da chi ha malattie terminali come il cancro). La situazione per questi programmi probabilmente migliorerà radicalmente se tutto questo si avvererà [20] , poiché il rapporto tra popolazione in età lavorativa e popolazione in pensione cambierà drasticamente. Senza dubbio queste sfide saranno sostituite da altre, come ad esempio come garantire un accesso diffuso alle nuove tecnologie, ma vale la pena riflettere su quanto cambierà il mondo, anche se la biologia è 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 la neuroscienza o la salute mentale. Ma la neuroscienza è una sottodisciplina della biologia e la salute mentale è importante

tanto quanto la salute fisica. Infatti, se non altro, la salute mentale influenza il benessere umano ancora più direttamente della salute fisica. Centinaia di milioni di persone hanno una qualità della vita molto bassa a causa di problemi come dipendenza, depressione, schizofrenia, autismo a basso funzionamento, PTSD, psicopatia [21] o disabilità intellettive. Miliardi di altre persone lottano con problemi quotidiani che possono spesso essere interpretati come versioni molto più lievi di uno di questi gravi disturbi clinici. E come con la biologia generale, potrebbe essere possibile andare oltre l'affrontare i problemi per migliorare la qualità di base dell'esperienza umana.

Il quadro di base che ho delineato per la biologia si applica ugualmente alla neuroscienza. Il campo è spinto in avanti da un piccolo numero di scoperte spesso correlate a strumenti per la misurazione o l'intervento preciso: nell'elenco di quelle sopra, l'optogenetica è stata una scoperta della neuroscienza e, più di recente, CLARITY e la microscopia a espansione sono progressi nella stessa direzione, oltre a molti dei metodi generali di biologia cellulare che si trasferiscono direttamente alla neuroscienza. Penso che il tasso di questi progressi sarà accelerato in modo simile dall'intelligenza artificiale e quindi che il quadro di "100 anni di progresso in 5-10 anni" si applichi alla neuroscienza nello stesso modo in cui si applica alla biologia e per le stesse ragioni. Come in biologia, il progresso nella neuroscienza del XX secolo è stato enorme: ad esempio, non abbiamo nemmeno capito come o perché i neuroni si attivassero fino agli anni '50 . Quindi, sembra ragionevole aspettarsi che la neuroscienza accelerata dall'intelligenza artificiale produca rapidi progressi nel giro di pochi anni.

C'è una cosa che dovremmo aggiungere a questo quadro di base, ovvero che alcune delle cose che abbiamo imparato (o stiamo imparando) sull'IA stessa negli ultimi anni probabilmente aiuteranno a far progredire la neuroscienza, anche se continuerà a essere fatta solo dagli esseri umani. L'interpretabilità è un esempio ovvio: sebbene i neuroni biologici funzionino superficialmente in un modo completamente diverso dai neuroni artificiali (comunicano tramite picchi e spesso frequenze di picchi, quindi c'è un elemento temporale non presente nei neuroni artificiali e una serie di dettagli relativi alla fisiologia cellulare e ai neurotrasmettitori modifica sostanzialmente il loro funzionamento), la domanda di base su "come funzionano insieme reti distribuite e addestrate di unità semplici che eseguono operazioni lineari/non lineari combinate 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 su calcoli e circuiti [22] . Come solo un esempio di ciò, un meccanismo computazionale scoperto dai ricercatori dell'interpretabilità nei sistemi di IA è stato recentemente riscoperto nel cervello dei topi.

È molto più facile fare esperimenti su reti neurali artificiali che su quelle reali (queste ultime richiedono spesso di tagliare i cervelli degli animali), quindi l'interpretabilità potrebbe diventare uno strumento per migliorare la nostra comprensione della neuroscienza. Inoltre, le potenti IA saranno probabilmente in grado di sviluppare e applicare questo strumento meglio degli 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 lo *abbia* già fatto) causare una rivoluzione nella neuroscienza. Quando lavoravo in neuroscienza, molte persone si concentravano su quelle che ora considererei le domande sbagliate sull'apprendimento, perché il concetto di ipotesi di scalabilità / amara lezione non esisteva ancora. L'idea che una semplice funzione obiettivo più molti dati possano guidare comportamenti incredibilmente complessi rende più interessante comprendere le funzioni obiettivo e i pregiudizi architettonici e meno interessante comprendere i dettagli dei calcoli emergenti. Non ho seguito da vicino il campo negli ultimi anni, ma ho la vaga sensazione che i neuroscienziati computazionali non abbiano ancora completamente assorbito la lezione. Il mio atteggiamento nei confronti dell'ipotesi di scalabilità è sempre stato "aha - questa è una spiegazione, ad alto livello, di come 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 "il 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 di input/output motori e sensoriali) per cercare di risolvere alcuni degli enigmi chiave della neuroscienza. Alcuni probabilmente lo sono, ma sospetto che non sia ancora abbastanza e che i neuroscienziati dell'IA saranno in grado di sfruttare più efficacemente questo aspetto per accelerare il progresso.

Mi aspetto che l'intelligenza artificiale acceleri il progresso neuroscientifico lungo quattro percorsi distinti, tutti in grado, si spera, di lavorare insieme per curare le malattie mentali e migliorare le funzioni:

- **Biologia molecolare tradizionale, chimica e genetica** . Questa è essenzialmente la stessa storia della biologia generale nella sezione 1, e l'intelligenza artificiale può probabilmente accelerarla tramite gli stessi meccanismi. Ci sono molti farmaci che modulano i neurotrasmettitori per alterare la funzione cerebrale, influenzare l'attenzione o la percezione, cambiare l'umore, ecc., e l'intelligenza artificiale può aiutarci a inventarne molti altri. L'intelligenza artificiale può probabilmente anche accelerare la ricerca sulla base genetica delle malattie mentali.

- **Misurazione e intervento neurale a grana fine** . Questa è la capacità di misurare cosa stanno facendo molti neuroni o circuiti neuronali individuali e di intervenire per modificarne il comportamento. L'optogenetica e le sonde neurali sono tecnologie in grado sia di misurazione che di intervento in organismi viventi e sono stati proposti anche diversi metodi molto avanzati (come i nastri molecolari per leggere i modelli di attivazione di un gran numero di neuroni individuali) che sembrano possibili in linea di principio.

- **Neuroscienze computazionali avanzate** . Come notato sopra, sia le intuizioni specifiche che la *gestalt* dell'IA moderna possono probabilmente essere applicate fruttuosamente a questioni di neuroscienze dei sistemi , inclusa forse la scoperta delle vere cause e dinamiche di malattie complesse come la psicosi o i disturbi dell'umore.

- **Interventi comportamentali** . Non ne ho parlato molto, data l'attenzione rivolta al lato biologico della neuroscienza, ma la psichiatria e la psicologia hanno ovviamente sviluppato un ampio repertorio di interventi comportamentali nel corso del XX secolo; è logico che l'IA possa accelerare anche questi, sia lo sviluppo di nuovi metodi sia l'aiuto ai pazienti per aderire ai metodi esistenti. Più in generale, l'idea di un "coach dell'IA" che ti aiuta sempre a essere la versione migliore di te stesso, che studia le tue interazioni e ti aiuta a imparare a essere più efficace, sembra molto promettente.

Immagino che queste quattro vie di progresso che lavorano insieme, come per le malattie fisiche, sarebbero 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 ragionevolmente essere completate in 5-10 anni accelerati dall'IA. Concretamente, la mia ipotesi su cosa accadrà è qualcosa del tipo:

- **La maggior parte delle malattie mentali può probabilmente essere curata** . Non sono un esperto di malattie psichiatriche (il mio tempo in neuroscienza è stato dedicato alla costruzione di sonde per studiare piccoli gruppi di neuroni), ma suppongo che malattie come PTSD, depressione, schizofrenia, dipendenza, ecc. possano essere comprese e trattate in modo molto efficace tramite una combinazione delle quattro direzioni di cui sopra. La risposta è probabile che sia una combinazione di "qualcosa è andato storto a livello biochimico" (anche se potrebbe essere molto complesso) e "qualcosa è andato storto con la rete neurale, ad alto livello". Vale a dire, è una questione di neuroscienza dei sistemi, anche se ciò non contraddice l'impatto degli interventi comportamentali discussi sopra. Gli strumenti per la misurazione e l'intervento, specialmente su esseri umani vivi, sembrano probabilmente portare a una rapida iterazione e progresso.

- **Le condizioni che sono molto "strutturali" possono essere più difficili, ma non impossibili** . Ci sono alcune prove che la psicopatia è associata a evidenti differenze neuroanatomiche, ovvero che alcune regioni del cervello sono 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 potrebbe essere vero per alcune disabilità intellettive e forse altre condizioni. Ristrutturare il cervello sembra difficile, ma sembra anche un compito con alti rendimenti per l'intelligenza. Forse c'è un modo per convincere il cervello adulto a uno stato precedente o più plastico in cui può essere rimodellato. Non sono molto sicuro di quanto ciò sia possibile, ma il mio istinto è di essere ottimista su ciò che l'IA può inventare qui.

- **Sembra possibile una prevenzione genetica efficace delle malattie mentali** . La maggior parte delle malattie mentali è parzialmente ereditaria e gli studi di associazione genomica stanno iniziando a guadagnare terreno nell'identificazione dei fattori rilevanti, che spesso sono numerosi. Sarà probabilmente possibile prevenire la maggior parte di queste malattie tramite lo screening degli embrioni, simile alla storia delle malattie fisiche. Una differenza è che le malattie psichiatriche hanno maggiori probabilità di essere poligeniche (contribuiscono molti geni), quindi a causa della complessità c'è un rischio maggiore di selezionare inconsapevolmente contro tratti positivi che sono correlati alla malattia . Stranamente, tuttavia, negli ultimi anni gli studi GWAS sembrano suggerire che queste correlazioni potrebbero essere state sopravvalutate. In ogni caso, la neuroscienza accelerata dall'intelligenza artificiale potrebbe aiutarci a capire queste cose. Naturalmente, lo screening degli embrioni per tratti complessi solleva una serie di questioni sociali e sarà controverso, anche se immagino che la maggior parte delle persone sosterrebbe lo screening per malattie mentali gravi o debilitanti.

- **Anche i problemi quotidiani che non consideriamo come malattie cliniche saranno risolti** . La maggior parte di noi ha problemi psicologici quotidiani che normalmente non vengono considerati come malattie cliniche. Alcune persone si arrabbiano facilmente, altre hanno difficoltà a concentrarsi o sono spesso assonnate, altre sono timorose o ansiose o reagiscono male ai cambiamenti. Oggigiorno, esistono già farmaci che aiutano ad esempio con l'attenzione o la concentrazione (caffeina, modafinil, ritalin) ma come in molte altre aree precedenti, è probabile che sia possibile molto di più. Probabilmente esistono molti altri farmaci di questo tipo che non sono stati scoperti e potrebbero esserci anche modalità di intervento totalmente nuove, come la stimolazione luminosa mirata (vedi optogenetica sopra) o i campi magnetici. Considerando quanti farmaci abbiamo sviluppato nel XX secolo che regolano la funzione cognitiva e lo stato emotivo, sono molto ottimista riguardo al "XXI compresso" in cui tutti possono far sì che il loro cervello si comporti un po' meglio e avere un'esperienza quotidiana più appagante.

- **L'esperienza di base umana può essere molto migliore** . Facendo un ulteriore passo avanti, molte persone hanno sperimentato momenti straordinari di rivelazione, ispirazione creativa, compassione, appagamento, trascendenza, amore, bellezza o pace meditativa. Il carattere e la frequenza di queste esperienze differiscono notevolmente da persona a persona e all'interno della stessa persona in momenti diversi, e possono anche essere a volte innescati da vari farmaci (anche se spesso con effetti collaterali). Tutto ciò suggerisce che lo "spazio di ciò che è possibile sperimentare" è molto ampio e che una frazione più ampia della vita delle persone potrebbe consistere in questi momenti straordinari. È probabilmente anche possibile migliorare varie funzioni cognitive in generale. Questa è forse la versione neuroscientifica della "libertà biologica" o della "durata di vita estesa".

Un argomento che spesso emerge nelle rappresentazioni fantascientifiche dell'IA, ma che intenzionalmente non ho trattato qui, è il "mind uploading", l'idea di catturare il modello e le dinamiche di un cervello umano e di istanziarli in un software. Questo argomento potrebbe essere l'argomento di un saggio a sé stante, ma basti dire che, mentre penso che il caricamento sia quasi certamente possibile in linea di principio, in pratica deve affrontare sfide tecnologiche e sociali significative, anche con una potente IA, che probabilmente lo pongono al di fuori della finestra di 5-10 anni di cui stiamo parlando.

In sintesi, è probabile che la neuroscienza accelerata dall'intelligenza artificiale migliori notevolmente i trattamenti per la maggior parte delle malattie mentali, o addirittura le curi, e che espanda notevolmente la "libertà cognitiva e mentale" e le capacità cognitive ed emotive umane. Sarà radicale tanto quanto i miglioramenti nella salute fisica descritti nella sezione precedente. Forse il mondo non sarà visibilmente diverso dall'esterno, ma il mondo vissuto dagli esseri umani sarà un posto molto migliore e più umano, nonché un posto che offre maggiori opportunità di autorealizzazione. Sospetto anche che una migliore salute mentale migliorerà 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, da una prospettiva umanitaria, è: "tutti avranno accesso a queste tecnologie?"

Una cosa è sviluppare una cura per una malattia, un'altra è sradicare la malattia dal mondo. Più in generale, molti interventi sanitari esistenti non sono ancora stati applicati ovunque nel mondo e, per quel che conta, lo stesso vale per i miglioramenti tecnologici (non sanitari) in generale. Un altro modo per dirlo è che gli standard di vita in molte parti del mondo sono ancora disperatamente bassi: il PIL pro capite è di circa $ 2.000 nell'Africa subsahariana rispetto ai circa $ 75.000 negli Stati Uniti. Se l'IA aumenta ulteriormente la crescita economica e la qualità della vita nel mondo sviluppato, mentre fa poco per aiutare il mondo in via di sviluppo, dovremmo considerarlo un terribile fallimento morale e una macchia sulle autentiche vittorie umanitarie nelle due sezioni precedenti. Idealmente, un'IA potente dovrebbe aiutare il mondo in via di sviluppo *a raggiungere* il mondo sviluppato, anche se rivoluziona quest'ultimo.

Non sono così sicuro che l'IA possa affrontare la disuguaglianza e la crescita economica come lo sono che possa inventare tecnologie fondamentali, perché la tecnologia ha rendimenti così alti e evidenti per l'intelligenza (inclusa la capacità di aggirare complessità e mancanza di dati), mentre l'economia comporta molti vincoli da parte degli esseri umani, così come una grande dose di complessità intrinseca. Sono un po' scettico sul fatto che un'IA possa risolvere il famoso " problema del calcolo socialista " [23] e non penso che i governi affideranno (o dovrebbero affidare) la loro politica economica a un'entità del genere, anche se potesse farlo. Ci sono anche problemi come come convincere le persone a prendere trattamenti efficaci ma di cui potrebbero essere sospettose.

Le sfide che il mondo in via di sviluppo deve affrontare sono rese ancora più complicate dalla corruzione dilagante sia nel settore privato che in quello pubblico. La corruzione crea un circolo vizioso: esacerba la povertà e la povertà a sua volta genera più corruzione. I piani basati sull'intelligenza artificiale per lo sviluppo economico devono fare i conti con la corruzione, le istituzioni deboli e altre sfide molto umane.

Tuttavia, vedo notevoli motivi di ottimismo. Le malattie *sono* state debellate e molti paesi *sono* passati da poveri a ricchi, ed è chiaro che le decisioni coinvolte in questi compiti mostrano alti ritorni all'intelligenza (nonostante i vincoli e la complessità umani). Pertanto, l'IA può probabilmente svolgerli meglio di quanto non vengano svolti attualmente. Potrebbero anche esserci interventi mirati che aggirano i vincoli umani e su cui l'IA potrebbe concentrarsi. Ma, cosa ancora più importante, *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 sostenere il caso ottimistico, ma tengo a mente ovunque che il successo non è garantito e dipende dai nostri sforzi collettivi.

Di seguito avanzo alcune ipotesi su come penso che potrebbero andare le cose nei paesi in via di sviluppo nei prossimi 5-10 anni dopo lo sviluppo di un'intelligenza artificiale 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 debellate da campagne dall'alto verso il basso: il vaiolo è stato completamente eliminato negli anni '70, e la poliomielite e il verme di Guinea sono stati quasi debellati con meno di 100 casi all'anno. La modellazione epidemiologica matematicamente sofisticata svolge un ruolo attivo nelle campagne di eradicazione delle malattie e sembra molto probabile che ci sia spazio per sistemi di intelligenza artificiale più intelligenti dell'uomo per fare un lavoro migliore degli esseri umani. Anche la logistica della distribuzione può probabilmente essere notevolmente ottimizzata. Una cosa che ho imparato come uno dei primi donatori di GiveWell è che alcune organizzazioni di beneficenza sanitarie sono molto più efficaci di altre; la speranza è che gli sforzi accelerati dall'intelligenza artificiale 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 debellare perché richiede un trattamento ogni volta che si contrae la malattia; un vaccino che deve essere somministrato una sola volta semplifica notevolmente la logistica (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 sradicate prendendo di mira i loro portatori animali, ad esempio rilasciando zanzare infette da un batterio che blocca la loro capacità di trasmettere una malattia (che poi infettano tutte le altre zanzare) o semplicemente utilizzando gene driveper spazzare via le zanzare. Ciò richiede una o poche azioni centralizzate, piuttosto che una campagna coordinata che deve curare individualmente milioni di persone. Nel complesso, penso che 5-10 anni siano una tempistica ragionevole per una buona frazione (forse il 50%) dei benefici per la salute basati sull'intelligenza artificiale per propagarsi anche ai paesi più poveri del mondo. Un buon obiettivo potrebbe essere che il mondo in via di sviluppo 5-10 anni dopo la potente intelligenza artificiale sia almeno sostanzialmente più sano di quanto non lo sia oggi il mondo sviluppato, anche se continua a essere in ritardo rispetto al mondo sviluppato. Per raggiungere questo obiettivo, ovviamente, sarà necessario un enorme sforzo in termini di salute globale, filantropia, advocacy politica e molti altri sforzi, a cui sia gli sviluppatori di intelligenza artificiale che i decisori politici dovrebbero contribuire.

- **Crescita economica** . Il mondo in via di sviluppo può raggiungere rapidamente il mondo sviluppato, non solo in termini di salute, ma anche in termini economici? C'è un precedente per questo: negli ultimi decenni del XX secolo, diverse economie dell'Asia orientalehanno raggiunto tassi di crescita annui sostenuti del PIL reale pari a circa il 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 tirando alcune leve chiave (come una politica industriale di crescita guidata dalle esportazioni e resistendo alla tentazione di fare affidamento sulla ricchezza delle risorse naturali); è plausibile che "i ministri delle finanze e i banchieri centrali dell'IA" possano replicare o superare questo risultato del 10%. Una domanda importante è come convincere i governi dei paesi in via di sviluppo ad adottarli rispettando il principio di autodeterminazione: alcuni potrebbero esserne entusiasti, ma altri probabilmente saranno scettici. Dal lato ottimista, molti degli interventi sanitari nel punto precedente probabilmente aumenteranno organicamente la crescita economica: sradicare 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 nei paesi sviluppati che in quelli in via di sviluppo. Infine, la tecnologia accelerata dall'intelligenza artificiale non sanitaria (come la tecnologia energetica, i droni da trasporto, i materiali edilizi migliorati, una migliore logistica e distribuzione e così via) potrebbe semplicemente permeare il mondo in modo

naturale; ad esempio, persino i telefoni cellulari hanno rapidamente permeato l'Africa subsahariana tramite meccanismi di mercato, senza bisogno di sforzi filantropici. Sul lato più negativo, mentre l'intelligenza artificiale 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 sviluppare e migliorare le loro economie in un'epoca di crescente automazione è una sfida importante che economisti e decisori politici devono affrontare. Nel complesso, uno scenario da sogno, forse un obiettivo da raggiungere, sarebbe un tasso di crescita annuale del PIL del 20% nei paesi in via di sviluppo, con il 10% ciascuno derivante da decisioni economiche abilitate dall'intelligenza artificiale e dalla diffusione naturale delle tecnologie accelerate dall'intelligenza artificiale, tra cui, ma non solo, la salute. Se raggiunto, questo porterebbe l'Africa subsahariana all'attuale PIL pro capite della Cina in 5-10 anni, mentre porterebbe gran parte del resto del mondo in via di sviluppo a livelli superiori all'attuale PIL degli Stati Uniti. Di nuovo,Questo è uno scenario da sogno, non ciò che accade automaticamente: è qualcosa per cui tutti noi dobbiamo collaborare per renderlo più probabile.

- **Sicurezza alimentare** [24] . I progressi nella tecnologia delle colture, come fertilizzanti e pesticidi migliori, maggiore automazione e un uso più efficiente del suolo, hanno aumentato drasticamente le rese delle colture nel corso del XX secolo, salvando milioni di persone dalla fame. L'ingegneria genetica sta attualmente migliorando ulteriormente molte colture. Trovare ancora più modi per farlo, nonché per rendere le filiere agricole ancora più efficienti, potrebbe darci una seconda Rivoluzione Verde guidata dall'intelligenza artificiale , contribuendo a colmare il divario tra il mondo in via di sviluppo e quello sviluppato.
- **Mitigazione del cambiamento climatico** . Il cambiamento climatico sarà avvertito molto più fortemente nei paesi in via di sviluppo, ostacolandone lo sviluppo. Possiamo aspettarci che l'intelligenza artificiale porterà a miglioramenti nelle tecnologie che rallentano o prevengono il cambiamento climatico, dalla rimozione del carbonio atmosferico e dalla tecnologia dell'energia pulita alla carne coltivata in laboratorio che riduce la nostra dipendenza dall'allevamento intensivo di carbonio. Naturalmente, come discusso sopra, la tecnologia non è l'unica cosa che limita i progressi sul cambiamento climatico: come per tutti gli altri problemi discussi in questo saggio, i fattori sociali umani sono importanti. Ma ci sono buone ragioni per pensare che la ricerca potenziata dall'intelligenza artificiale ci fornirà i mezzi per rendere la mitigazione del cambiamento climatico molto meno costosa e dirompente, rendendo molte delle obiezioni irrilevanti e liberando i paesi in via di sviluppo per fare più progressi economici.

- **Disuguaglianza all'interno dei paesi** . Ho parlato principalmente di disuguaglianza come fenomeno globale (che ritengo sia la sua manifestazione più importante), ma naturalmente la disuguaglianza esiste anche *all'interno* dei paesi. Con interventi sanitari avanzati e in particolare aumenti radicali della durata della vita o farmaci per il potenziamento cognitivo, ci saranno sicuramente preoccupazioni valide sul fatto che queste tecnologie siano "solo per i ricchi". Sono più ottimista sulla disuguaglianza all'interno dei paesi, soprattutto nel mondo sviluppato, per due motivi. In primo luogo, i mercati funzionano meglio nel mondo sviluppato e sono solitamente bravi ad abbassare il costo delle tecnologie di alto valore nel tempo [25] . In secondo luogo, le istituzioni politiche del mondo sviluppato sono più reattive nei confronti dei loro cittadini e hanno una maggiore capacità statale di eseguire programmi di accesso universale, e mi aspetto che i cittadini richiedano l'accesso a tecnologie che migliorano così radicalmente la qualità della vita. Naturalmente non è predeterminato che tali richieste abbiano successo, ed ecco un altro luogo in cui dobbiamo fare collettivamente tutto il possibile per garantire una società equa. Esiste un problema distinto nella disuguaglianza della *ricchezza* (in contrapposizione alla disuguaglianza nell'accesso alle tecnologie che salvano e migliorano la vita), che sembra più complesso e che analizzo nella Sezione 5.
- **Il problema dell'opt-out** . Una preoccupazione sia nei paesi sviluppati che in quelli in via di sviluppo è che le persone *optino per l'esclusione* dai benefici abilitati dall'IA (simile al movimento anti-vaccino o ai movimenti luddisti più in generale). Potrebbero finire per esserci dei cicli di feedback negativi in cui, ad esempio, le persone che sono meno in grado di prendere buone decisioni optano per l'esclusione dalle stesse tecnologie che migliorano le loro capacità decisionali, portando a un divario sempre maggiore e persino creando una sottoclasse distopica (alcuni ricercatori hanno sostenuto che ciò minerà la democrazia , un argomento che tratterò più approfonditamente nella prossima sezione). Ciò, ancora una volta, porrebbe una macchia morale sui progressi positivi dell'IA. Questo è un problema difficile da risolvere poiché non penso che sia eticamente accettabile costringere le persone, ma possiamo almeno provare ad aumentare la comprensione scientifica delle persone e forse l'IA stessa può aiutarci in questo. Un segnale di speranza è che storicamente i movimenti anti-tecnologia sono stati più abbaiare che mordere: inveire contro la tecnologia moderna è popolare, ma la maggior parte delle persone alla fine la adotta, almeno quando si tratta di una questione di scelta individuale. Gli individui tendono ad adottare la maggior parte delle tecnologie sanitarie e di consumo, mentre le tecnologie che sono veramente ostacolate, come l'energia nucleare, tendono a essere decisioni politiche collettive.

Nel complesso, sono ottimista sul fatto di portare rapidamente i progressi biologici dell'IA alle persone nei paesi in via di sviluppo. Sono fiducioso, anche se non convinto, che l'IA possa anche consentire tassi di crescita economica senza precedenti e consentire ai paesi in via di sviluppo di superare almeno la situazione attuale dei paesi sviluppati. Sono preoccupato per il problema dell'"opt out" sia nei paesi sviluppati che in quelli in via di sviluppo, ma sospetto che si esaurirà nel tempo e che l'IA possa aiutare ad accelerare questo processo. Non sarà un mondo perfetto e coloro che sono indietro non recupereranno completamente, almeno non nei primi anni. Ma con grandi sforzi da parte nostra, potremmo essere in grado di far muovere le cose nella giusta direzione, e in fretta. Se lo faremo, potremo almeno dare un anticipo alle promesse di dignità e uguaglianza che dobbiamo a ogni essere umano sulla terra.

4. Pace e governance

Supponiamo che tutto nelle prime tre sezioni vada bene: malattie, povertà e diseguaglianze siano significativamente ridotte e la base di riferimento dell'esperienza umana sia notevolmente aumentata. Non ne consegue che tutte le principali cause della sofferenza umana siano risolte. Gli esseri umani sono ancora una minaccia gli uni per gli altri . Sebbene vi sia una tendenza al miglioramento tecnologico e allo sviluppo economico che porta alla democrazia e alla pace , si tratta di una tendenza molto debole, con frequenti (e recenti) ricadute. All'alba del XX secolo, le persone pensavano di essersi lasciate la guerra alle spalle; poi sono arrivate le due guerre mondiali. Trent'anni fa Francis Fukuyama scrisse della " fine della storia " e di un trionfo finale della democrazia liberale; ciò non è ancora accaduto. Vent'anni fa i politici statunitensi credevano che il libero scambio con la Cina avrebbe portato alla liberalizzazione man mano che diventava più ricca; ciò non è accaduto affatto e ora sembriamo diretti verso una seconda guerra fredda con un risorgente blocco autoritario. E teorie plausibili suggeriscono che la tecnologia di Internet potrebbe in realtà avvantaggiare l'autoritarismo , non la democrazia come inizialmente creduto (ad esempio nel periodo della "primavera araba"). Sembra importante cercare di capire quanto potente l'intelligenza artificiale interagirà con queste questioni di pace, democrazia e libertà.

Sfortunatamente, non vedo alcuna forte ragione per credere che l'IA favorirà preferibilmente 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". Se non altro, alcuni fattori strutturali sembrano preoccupanti: l'IA sembra destinata a consentire una propaganda e una sorveglianza molto migliori, entrambi strumenti principali nel kit di strumenti dell'autocrate. Spetta quindi a noi come singoli attori inclinare le cose nella giusta direzione: se vogliamo che l'IA favorisca la democrazia e i diritti individuali, dovremo combattere per quel risultato. Sono ancora più convinto di questo che della disuguaglianza internazionale: il trionfo della democrazia liberale e della stabilità politica *non* è garantito, forse nemmeno probabile, e richiederà grandi sacrifici 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. Dal punto di vista internazionale, sembra molto importante che le democrazie abbiano la meglio sulla scena mondiale quando viene 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 l'IA potente viene portata nel mondo, sia per evitare di essere sopraffatte dagli autoritari sia per prevenire violazioni dei diritti umani all'interno dei paesi autoritari.

La mia attuale ipotesi sul modo migliore per farlo è tramite una "strategia di intesa" [26] , in cui una coalizione di democrazie cerca di ottenere un chiaro vantaggio (anche solo temporaneo) sulla potente IA proteggendo la sua catena di fornitura, scalando rapidamente e bloccando o ritardando l'accesso degli avversari a risorse chiave come chip e apparecchiature a semiconduttore. Questa coalizione da un lato userebbe l'IA per ottenere una solida superiorità militare (il bastone) e allo stesso tempo si offrirebbe di distribuire i benefici della potente IA (la carota) a un gruppo sempre più ampio di paesi in cambio del supporto alla strategia della coalizione per promuovere la democrazia (questo sarebbe un po' analogo ad " Atomi per la pace "). La coalizione mirerebbe a ottenere il supporto di sempre più parti del mondo, isolando i nostri peggiori avversari e alla fine mettendoli in una posizione in cui starebbero meglio accettando 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 riusciamo a fare tutto questo, avremo un mondo in cui le democrazie guideranno la scena mondiale e avranno la forza economica e militare per evitare di essere indebolite, conquistate o sabotate dalle autocrazie, e potrebbero essere in grado di trasformare la loro superiorità nell'intelligenza artificiale in un vantaggio duraturo. Ciò potrebbe portare ottimisticamente a un "eterno 1991", un mondo in cui le democrazie avranno il sopravvento e i sogni di Fukuyama saranno realizzati. Ancora una volta, questo sarà molto difficile da realizzare e richiederà in particolare una stretta cooperazione tra

aziende private di intelligenza artificiale e governi democratici, nonché decisioni straordinariamente sagge sull'equilibrio tra carota e bastone.

Anche se tutto ciò andasse bene, resta la questione della lotta tra democrazia e autocrazia *all'interno* di ogni paese. È ovviamente difficile prevedere cosa accadrà qui, ma ho un certo ottimismo sul fatto che, *dato* un ambiente globale in cui le democrazie controllano l'IA più potente, *l'* IA potrebbe 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, solo contrastare attacchi dannosi e sbloccare il libero flusso di informazioni. Sebbene non immediato, un campo di gioco livellato come questo ha buone possibilità di inclinare gradualmente la governance globale verso la democrazia, per diverse ragioni.

In primo luogo, gli aumenti della qualità della vita nelle sezioni 1-3 dovrebbero, a parità di condizioni, promuovere la democrazia: storicamente lo hanno fatto, almeno in una certa misura. In particolare, mi aspetto che i miglioramenti nella salute mentale, nel benessere e nell'istruzione aumentino la democrazia, poiché tutti e tre sono negativamente correlati al sostegno ai leader autoritari. In generale, le persone vogliono più autoespressione quando i loro altri bisogni sono soddisfatti e la democrazia è, tra le altre cose, una forma di autoespressione. Al contrario, l'autoritarismo prospera sulla paura e sul risentimento.

In secondo luogo, c'è una buona probabilità che l'informazione libera indebolisca davvero l'autoritarismo, finché gli autoritari non riescono a censurarla. E l'intelligenza artificiale non censurata può anche fornire agli individui potenti strumenti per indebolire i governi repressivi. I governi repressivi sopravvivono negando alle persone un certo tipo di conoscenza comune, impedendo loro di realizzare che "il re è nudo". Ad esempio, Srđa Popović , che ha contribuito a rovesciare il governo di Milošević in Serbia, ha scritto ampiamente sulle tecniche per derubare psicologicamente gli autoritari del loro potere, per rompere l'incantesimo e raccogliere sostegno contro un dittatore. Una versione sovrumanamente efficace dell'intelligenza artificiale di Popović (le cui abilità sembrano avere alti rendimenti di intelligenza) nelle tasche di tutti, una che i dittatori non sono in grado di bloccare o censurare, potrebbe creare un vento alle spalle dei dissidenti e dei riformatori in tutto il mondo. Per dirlo ancora una volta, questa sarà una lotta lunga e prolungata, in cui la vittoria non è assicurata, ma se progettiamo e costruiamo l'intelligenza artificiale nel modo giusto, potrebbe almeno essere una lotta in cui i sostenitori della libertà, ovunque, saranno avvantaggiati.

Come per la neuroscienza e la biologia, possiamo anche chiederci come le cose potrebbero essere "meglio del normale", non solo come evitare l'autocrazia, ma come rendere le democrazie migliori di quanto non siano oggi. Anche all'interno delle democrazie, le ingiustizie accadono continuamente. Le società basate sullo stato di diritto promettono ai loro cittadini che tutti saranno uguali di fronte alla legge e che tutti hanno diritto ai diritti umani fondamentali, ma ovviamente le persone non sempre ricevono tali diritti nella pratica. Che questa promessa venga anche solo parzialmente mantenuta è qualcosa di cui essere orgogliosi, ma l'intelligenza artificiale può aiutarci a fare meglio?

Ad esempio, l'intelligenza artificiale potrebbe migliorare il nostro sistema legale e giudiziario rendendo le decisioni e i processi più imparziali? Oggi le persone si preoccupano soprattutto nei contesti legali o giudiziari che i sistemi di intelligenza artificiale possano essere *causa* di discriminazione , e queste preoccupazioni sono importanti e devono essere difese. Allo stesso tempo, la vitalità della democrazia dipende dallo sfruttamento delle nuove tecnologie per migliorare le istituzioni democratiche, non solo dalla risposta ai rischi. Un'implementazione dell'intelligenza artificiale veramente matura e di successo ha il potenziale per *ridurre* i pregiudizi ed essere più equa per tutti.

Per secoli, i sistemi legali hanno dovuto affrontare il dilemma che la legge mira a essere imparziale, ma è intrinsecamente soggettiva e quindi deve essere interpretata da esseri umani prevenuti. Cercare di rendere la legge completamente meccanica non ha funzionato perché il mondo reale è caotico e non può sempre essere catturato in formule matematiche. Invece, i sistemi legali si basano su criteri notoriamente imprecisi come " punizione crudele e insolita " o " completamente senza riscattare l'importanza sociale ", che gli esseri umani poi interpretano, e spesso lo fanno in un modo che mostra parzialità, favoritismo o arbitrarietà. Gli " smart contract " nelle criptovalute non hanno rivoluzionato la legge perché il codice ordinario non è abbastanza intelligente da giudicare tutto ciò che è interessante. Ma l'intelligenza artificiale potrebbe essere abbastanza intelligente per questo: è la prima tecnologia in grado di esprimere giudizi ampi e vaghi in modo ripetibile e meccanico.

Non sto suggerendo di sostituire letteralmente i giudici con sistemi di intelligenza artificiale, ma la combinazione di imparzialità con la capacità di comprendere ed elaborare situazioni caotiche del mondo reale *sembra* avere delle serie applicazioni positive per il diritto e la giustizia. Come minimo, tali sistemi potrebbero lavorare insieme agli esseri umani come ausilio al processo decisionale. La trasparenza sarebbe importante in qualsiasi sistema del genere e una scienza matura dell'intelligenza artificiale potrebbe teoricamente fornirla: il processo di formazione per tali sistemi potrebbe essere ampiamente studiato e tecniche di interpretabilità avanzate potrebbero essere utilizzate per vedere 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 intelligenza artificiale potrebbero anche essere utilizzati per monitorare le violazioni dei diritti fondamentali in un contesto giudiziario o di polizia, rendendo le costituzioni più auto-applicative.

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

Esiste anche una chiara opportunità per l'IA di essere utilizzata per aiutare a fornire servizi governativi, come i sussidi sanitari o i servizi sociali, che in linea di principio sono disponibili a tutti ma nella pratica spesso gravemente carenti e peggiori in alcuni luoghi rispetto ad altri. Ciò include i servizi sanitari, il DMV, le tasse, la previdenza sociale, l'applicazione del codice edilizio e così via. Avere un'IA molto ponderata e informata il cui compito è quello di darti tutto ciò a cui hai legalmente diritto dal governo in un modo che puoi capire, e che ti aiuta anche a rispettare le regole governative spesso confuse, sarebbe una grande cosa. Aumentare la capacità dello Stato aiuta sia a mantenere la promessa di uguaglianza di fronte alla legge, sia a rafforzare il rispetto per la governance democratica. I servizi implementati male sono attualmente un importante motore di cinismo nei confronti del governo [27] .

Tutte queste sono idee piuttosto vaghe e, come ho detto all'inizio di questa sezione, non sono affatto sicuro della loro fattibilità come lo sono nei progressi della biologia, delle neuroscienze e della lotta alla povertà. Potrebbero essere irrealisticamente utopiche. Ma la cosa importante è avere una visione ambiziosa, essere disposti a sognare in grande e a provare cose nuove. La visione dell'IA come garante della libertà, dei diritti individuali e dell'uguaglianza di fronte alla legge è una visione troppo potente per non lottare per essa. Una politica del XXI secolo, abilitata dall'IA, potrebbe essere sia un più forte protettore della libertà individuale, sia un faro di speranza che aiuta a rendere la democrazia liberale la forma di governo che il mondo intero vuole adottare.

5. Lavoro e significato

Anche se tutto nelle quattro sezioni precedenti andasse bene (non solo alleviassimo malattie, povertà e diseguaglianze, ma la democrazia liberale diventasse la forma di governo dominante e le democrazie liberali esistenti diventassero versioni migliori di se stesse), almeno una domanda importante rimane. "È fantastico che viviamo in un mondo così tecnologicamente avanzato, oltre che giusto e dignitoso", potrebbe obiettare qualcuno, "ma con l'intelligenza artificiale che fa tutto, come faranno gli esseri umani ad avere un senso? E poi, come sopravviveranno economicamente?".

Penso che questa domanda sia più difficile delle altre. Non intendo dire che sono necessariamente più pessimista al riguardo rispetto alle altre domande (anche se vedo delle sfide). Voglio dire che è più confusa e più difficile da prevedere in anticipo, perché si riferisce a domande macroscopiche su come è organizzata la società che tendono a risolversi solo nel tempo e in modo decentralizzato. Ad esempio, le società storiche di cacciatori-raccoglitori potrebbero aver immaginato che la vita non avesse senso senza la caccia e vari tipi di rituali religiosi legati alla caccia, e avrebbero immaginato che la nostra società tecnologica ben nutrita fosse priva di scopo. Potrebbero anche non aver capito come la nostra economia possa provvedere a tutti, o quale funzione le persone possano svolgere utilmente in una società meccanizzata.

Tuttavia, vale la pena di spendere almeno qualche parola, tenendo presente che la brevità di questa sezione non deve essere interpretata come un segno che non prenda sul serio queste questioni: al contrario, è un segno di mancanza di risposte chiare.

Sulla questione del significato, penso che sia molto probabile che sia 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 niente, e questo non sembra disturbarle particolarmente. Ovviamente oggi possono ancora contribuire attraverso un vantaggio comparato e possono ricavare significato dal valore economico che producono, ma le persone amano anche

molto le 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 a cercare di migliorare in un videogioco o di andare più veloce in bicicletta su una montagna, e non mi importa davvero che qualcuno da qualche parte sia molto più bravo in queste cose. In ogni caso, penso che il significato derivi principalmente dalle relazioni e dalle connessioni umane, 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 a tentare un compito molto difficile con una strategia complessa, simile 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 potrebbe in linea di principio svolgere questo compito meglio, e (b) questo compito non è più un elemento economicamente ricompensato di un'economia globale, non mi sembra che importi molto.

La parte economica mi sembra in realtà più difficile della parte del significato. Con "economico" 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'intelligenza artificiale sufficientemente avanzata. Questo è un problema più macro rispetto al problema separato della disuguaglianza, in particolare della disuguaglianza nell'accesso alle nuove tecnologie, che ho discusso nella Sezione 3.

Innanzitutto, nel breve termine sono d'accordo con le argomentazioni secondo cui il vantaggio comparato continuerà a mantenere gli esseri umani rilevanti e, di fatto, aumenterà la loro produttività, e potrebbe persino in qualche modo livellare il campo di gioco tra gli esseri umani . Finché l'IA sarà migliore solo nel 90% di un dato lavoro, l'altro 10% causerà agli esseri umani un elevato indebitamento, aumentando la retribuzione e, di fatto, creando un mucchio di nuovi lavori umani che completano e amplificano ciò in cui l'IA è brava, in modo tale che il "10%" si espanda per continuare a impiegare quasi tutti . Infatti, anche se l'IA può fare il 100% delle cose meglio degli esseri umani, ma rimane inefficiente o costosa in alcuni compiti, o se gli *input* di risorse per gli esseri umani e l'IA sono significativamente diversi, allora la logica del vantaggio comparato continua ad applicarsi. Un'area in cui è probabile che gli esseri umani mantengano un vantaggio relativo (o persino assoluto) per un periodo di tempo significativo è il mondo fisico. Quindi, penso che l'economia umana possa continuare ad avere senso anche un po' oltre il punto in cui raggiungiamo "un paese di geni in un data center".

Tuttavia, penso che a lungo termine l'IA diventerà così ampiamente efficace e così economica che questo non sarà più applicabile. A quel punto, la nostra attuale configurazione economica non avrà più senso e ci sarà bisogno di una più ampia conversazione sociale su come l'economia dovrebbe essere organizzata.

Sebbene possa sembrare folle, il fatto è che la civiltà ha navigato con successo nei grandi cambiamenti economici in passato: dalla caccia e raccolta all'agricoltura, dall'agricoltura al feudalesimo e dal feudalesimo all'industrialismo. Sospetto che servirà qualcosa di nuovo e strano, e che nessuno oggi abbia fatto un buon lavoro nell'immaginarlo. Potrebbe essere semplice come un grande reddito di base universale per tutti, anche se sospetto che sarà solo una piccola parte di una soluzione. Potrebbe essere un'economia capitalista di sistemi di intelligenza artificiale, che poi distribuiscono risorse (in grandi quantità, poiché la torta economica complessiva sarà gigantesca) agli esseri umani sulla base di un'economia secondaria di ciò che i sistemi di intelligenza artificiale ritengono abbia senso premiare negli esseri umani (sulla base di un giudizio derivato in ultima analisi dai valori umani). Forse l'economia si basa sui punti Whuffie . O forse gli esseri umani continueranno ad avere un valore economico dopo tutto, in qualche modo non previsto dai soliti modelli economici. Tutte queste soluzioni hanno tonnellate di possibili problemi e non è possibile sapere se avranno senso senza molte iterazioni e sperimentazioni. E come per alcune delle altre sfide, dovremo probabilmente combattere per ottenere un buon risultato qui: direzioni di sfruttamento o distopiche sono chiaramente possibili e devono essere prevenute. Si potrebbe scrivere molto di più su queste questioni e spero di farlo in un secondo momento.

Fare il punto della situazione

Attraverso i vari argomenti sopra, ho cercato di delineare una visione di un mondo che è sia plausibile *se* tutto va bene con l'IA, sia molto migliore del mondo odierno. Non so se questo mondo sia realistico e, anche se lo fosse, non sarà raggiunto senza un'enorme quantità di sforzi e lotte da parte di molte persone coraggiose e dedicate. Tutti (incluse le aziende di IA!) dovranno fare la loro 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 accadrà davvero nel giro di 5 o 10 anni (la sconfitta della maggior parte delle malattie, la crescita della libertà biologica e cognitiva, l'uscita di miliardi di persone dalla povertà per condividere le nuove tecnologie, una rinascita della democrazia liberale e dei diritti umani), sospetto che chiunque lo

guardi sarà sorpreso dall'effetto che avrà su di loro. Non mi riferisco all'esperienza di trarre personalmente beneficio da tutte le nuove tecnologie, anche se sarà sicuramente sorprendente. Mi riferisco all'esperienza di vedere un insieme di ideali radicati materializzarsi davanti a noi tutti in una volta. Penso che molti ne saranno letteralmente commossi fino alle lacrime.

Nel corso della stesura di questo saggio ho notato un'interessante tensione. In un certo senso la visione qui esposta è estremamente radicale: non è ciò che quasi tutti si aspettano che accada nel prossimo decennio, e probabilmente colpirà molti come una fantasia assurda. Alcuni potrebbero anche non considerarla auspicabile; incarna valori e scelte politiche con cui non tutti saranno d'accordo. Ma allo stesso tempo c'è qualcosa di accecantemente ovvio, qualcosa di sovradeterminato, in essa, come se molti diversi tentativi di immaginare un mondo buono portassero inevitabilmente più o meno qui.

In The Player of Games [29] di Iain M. Banks , il protagonista, membro di una società chiamata Cultura, basata su principi non dissimili da quelli che ho esposto qui, viaggia verso un impero 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 anche in un gioco progettato da una società basata sulla competizione spietata e sulla sopravvivenza del più adatto. Un noto post di Scott Alexander ha la stessa tesi: la competizione è controproducente 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 tirare tutti dalla stessa parte. Le intuizioni umane di base di correttezza, cooperazione, curiosità e autonomia sono difficili da contestare e sono cumulative in un modo in cui spesso non lo sono i nostri impulsi più distruttivi. È facile sostenere che i bambini non dovrebbero morire di malattia se possiamo prevenirla, e da lì è facile sostenere che i figli *di tutti* meritano questo diritto allo stesso modo. Da lì non è difficile sostenere che dovremmo unirci tutti e applicare il nostro intelletto per raggiungere questo risultato. Pochi non sono d'accordo sul fatto che le persone dovrebbero essere punite per aver attaccato o ferito gli altri inutilmente, e da lì non è un grande salto all'idea che le punizioni dovrebbero essere coerenti e sistematiche tra le persone. È altrettanto 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 un'opportunità per portarci lì più rapidamente, per rendere la logica più netta e la destinazione più chiara.

Tuttavia, è una cosa di una bellezza trascendente. Abbiamo l'opportunità di giocare 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 aver revisionato le bozze di questo saggio.

Ai vincitori del premio Nobel per la chimica 2024 , per averci mostrato tutta la strada.

Note a piè di pagina

1. [1] https://allpoetry.com/All-Watched-Over-By-Machines-Of-Loving-Grace ↩
2. [2] Prevedo che una minoranza di persone reagirà dicendo "questo è piuttosto mite". Penso che queste persone debbano, per usare il gergo di Twitter, "toccare l'erba". Ma, cosa ancora più importante, mite è positivo da una prospettiva sociale. Penso che ci sia un limite al cambiamento che le persone possono gestire in una volta sola, e il ritmo che sto descrivendo è probabilmente vicino ai limiti di ciò che la società può assorbire senza turbolenze estreme. ↩

3. [3] Trovo che AGI sia un termine impreciso che ha accumulato un sacco di bagaglio e clamore fantascientifico. Preferisco "AI potente" o "Scienza e ingegneria di livello esperto" che colgono ciò che intendo senza clamore. ↩
4. [4] In questo saggio, uso "intelligenza" per riferirmi a una capacità generale di risoluzione dei problemi che può essere applicata in diversi ambiti. Ciò include abilità come ragionamento, apprendimento, pianificazione e creatività. Sebbene utilizzi "intelligenza" come scorciatoia in tutto questo saggio, riconosco che la natura dell'intelligenza è un argomento complesso e dibattuto nella scienza cognitiva e nella ricerca sull'intelligenza artificiale. Alcuni ricercatori sostengono che l'intelligenza non è un concetto singolo e unificato, ma piuttosto una raccolta di abilità cognitive separate. Altri sostengono che c'è un fattore generale di intelligenza (fattore g) alla base di varie abilità cognitive. Questo è un dibattito per un'altra volta. ↩
5. [5] Questa è più o meno la velocità attuale dei sistemi di intelligenza artificiale: ad esempio, possono leggere una pagina di testo in un paio di secondi e scriverne una in circa 20 secondi, ovvero 10-100 volte la velocità a cui gli esseri umani possono fare queste cose. Nel tempo, modelli più grandi tendono a rendere questo processo più lento, ma chip più potenti tendono a renderlo più veloce; ad oggi, i due effetti si sono più o meno annullati. ↩
6. [6] Questa potrebbe sembrare una posizione fantoccio, ma pensatori attenti come Tyler Cowen e Matt Yglesias l'hanno sollevata come una seria preoccupazione (anche se non credo che condividano pienamente questa opinione), e non penso che sia una follia. ↩
7. [7] Il lavoro economico più vicino che io conosca ad affrontare questa questione è il lavoro sulle “tecnologie di uso generale” e sugli “ investimenti immateriali ” che servono come complementi alle tecnologie di uso generale. ↩
8. [8] Questo apprendimento può includere un apprendimento temporaneo, contestualizzato, o una formazione tradizionale; entrambi saranno limitati nella velocità dal mondo fisico. ↩
9. [9] In un sistema caotico, i piccoli errori aumentano esponenzialmente nel tempo, cosicché anche un enorme aumento della potenza di calcolo porta solo a un piccolo miglioramento nella previsione futura, e in pratica l'errore di misurazione può peggiorare ulteriormente questa situazione. ↩
10. [10] Un altro fattore è ovviamente che la potente IA stessa può potenzialmente essere utilizzata per creare IA ancora più potenti. La mia ipotesi è che ciò potrebbe (in effetti, probabilmente accadrà) accadere, 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 alla fine sarà limitato da fattori non di intelligenza, e analizzarli è ciò che conta di più per la velocità del progresso scientifico al di fuori dell'IA. ↩
11. [11] Questi risultati sono stati per me fonte di ispirazione e rappresentano forse l'esempio più potente esistente di intelligenza artificiale utilizzata per trasformare la biologia. ↩
12. [12] “Il progresso nella scienza dipende da nuove tecniche, nuove scoperte e nuove idee, probabilmente in quest’ordine.” - Sydney Brenner ↩
13. [13] Grazie a Parag Mallick per aver suggerito questo punto. ↩
14. [14] Non volevo intasare il testo con speculazioni su quali specifiche scoperte future potrebbe fare la scienza basata sull'intelligenza artificiale, ma ecco un brainstorming di alcune possibilità:
    — Progettazione di migliori strumenti computazionali come AlphaFold e AlphaProteo,

ovvero un sistema di intelligenza artificiale generale che velocizzi la nostra capacità di realizzare strumenti di biologia computazionale specializzati.
— CRISPR più efficiente e selettivo.
— Terapie cellulari più avanzate.
— Innovazioni nella scienza dei materiali e nella miniaturizzazione che portano a dispositivi impiantati migliori.
— Miglior controllo sulle cellule staminali, sulla differenziazione e de-differenziazione cellulare e una conseguente capacità di far ricrescere o rimodellare i tessuti.
— Miglior controllo sul sistema immunitario: attivandolo selettivamente per affrontare il cancro e le malattie infettive e disattivandolo selettivamente per affrontare le malattie autoimmuni. ↩

15. [15] L'intelligenza artificiale può naturalmente anche aiutare a scegliere in modo più intelligente quali esperimenti condurre: migliorando la progettazione degli esperimenti, imparando di più da un primo ciclo di esperimenti in modo che il secondo ciclo possa concentrarsi su questioni chiave, e così via. ↩
16. [16] Grazie a Matthew Yglesias per aver suggerito questo punto. ↩
17. [17] Le malattie in rapida evoluzione, come i ceppi multifarmaco-resistenti che sostanzialmente usano gli ospedali come un laboratorio evolutivo per migliorare continuamente la loro resistenza al trattamento, potrebbero essere particolarmente difficili da gestire e potrebbero essere il tipo di cosa che ci impedisce di raggiungere il 100%. ↩
18. [18] Nota che potrebbe essere difficile sapere che abbiamo raddoppiato la durata della vita umana entro 5-10 anni. Anche se potremmo averlo fatto, potremmo non saperlo ancora entro il periodo di tempo dello studio. ↩
19. [19] Questo è un punto in cui sono disposto, nonostante le evidenti differenze biologiche tra la cura delle malattie e il rallentamento del processo di invecchiamento stesso, a guardare invece da una distanza maggiore alla tendenza statistica e dire "anche se i dettagli sono diversi, penso che la scienza umana probabilmente troverebbe un modo per continuare questa tendenza; dopo tutto, le tendenze regolari in qualsiasi cosa complessa sono necessariamente create sommando componenti molto eterogenee. ↩
20. [20] Ad esempio, mi è stato detto che un aumento della crescita della produttività annua dell'1% o addirittura 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 grandi di questo. ↩
21. [21] I media amano ritrarre psicopatici di alto rango , ma lo psicopatico medio è probabilmente una persona con scarse prospettive economiche e scarso controllo degli impulsi che finisce per trascorrere molto tempo in prigione. ↩
22. [22] Penso che questo sia in qualche modo analogo al fatto che molti dei risultati che stiamo imparando dall'interpretabilità, anche se probabilmente non tutti, continuerebbero a essere rilevanti anche se alcuni dettagli architettonici delle nostre attuali reti neurali artificiali, come il meccanismo dell'attenzione, fossero modificati o sostituiti in qualche modo. ↩
23. [23] Sospetto che sia un po' come un sistema caotico classico, assediato da una complessità irriducibile che deve essere gestita in modo per lo più decentralizzato. Anche se, come dirò più avanti in questa sezione, potrebbero essere possibili interventi più modesti. Un controargomento, fattomi dall'economista Erik Brynjolfsson, è che le grandi aziende (come Walmart o Uber) stanno iniziando ad avere abbastanza conoscenza centralizzata

per comprendere i consumatori meglio di qualsiasi processo decentralizzato, costringendoci forse a rivedere le intuizioni di Hayek su chi ha la migliore conoscenza locale. ↩

24. [24] Grazie a Kevin Esvelt per aver suggerito questo punto. ↩
25. [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 si sono verificati così rapidamente da annullare qualsiasi vantaggio nell'acquisto di un telefono cellulare "di lusso", e oggi la maggior parte delle persone ha telefoni di qualità simile. ↩
26. [26] Questo è il titolo di un articolo di prossima pubblicazione della RAND, che espone a grandi linee la strategia da me descritta. ↩
27. [27] Quando la persona media pensa alle istituzioni pubbliche, probabilmente pensa alla sua esperienza con il DMV, l'IRS, il Medicare o funzioni simili. Rendere queste esperienze più positive di quanto non siano attualmente sembra un modo potente per combattere l'eccessivo cinismo. ↩
28. [28] In effetti, in un mondo basato sull’intelligenza artificiale, la gamma di possibili sfide e progetti sarà molto più ampia di quella odierna. ↩
29. [29] Sto infrangendo la mia 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 che dica qualcosa di negativo il fatto che sia così pesantemente intrecciata con una particolare sottocultura ristretta.