

Migliorare il processo decisionale (soprattutto nelle istituzioni importanti)

Sintesi

Aiutare i governi e altre importanti istituzioni a migliorare il loro processo decisionale in decisioni complesse e ad alto rischio, soprattutto in relazione a rischi catastrofici globali, potrebbe essere uno dei problemi più importanti su cui lavorare. Ma ci sono molti dubbi su quanto sia fattibile lavorare su questo problema e su quali siano le soluzioni migliori da implementare.

La nostra visione d'insieme

Consigliato

Pensiamo che lavorare su questo tema possa essere uno dei modi migliori per migliorare le prospettive per il futuro a lungo termine, anche se non siamo così convinti in quest'area come in altre, in parte perché non siamo sicuri di cosa sia più prezioso al suo interno.

Portata

Riteniamo che questo tipo di lavoro possa avere un grande impatto positivo. I miglioramenti potrebbero portare a un'allocazione più efficace delle risorse da parte di fondazioni e governi, a progressi più rapidi su alcuni dei problemi più urgenti del mondo, a una riduzione dei rischi derivanti dalle tecnologie emergenti o a una riduzione dei rischi di conflitto. E se questo lavoro potesse ridurre la probabilità di una catastrofe esistenziale - cosa che non sembra da escludere - potrebbe essere uno dei modi migliori per migliorare le prospettive del futuro a lungo termine.

Trascuratezza

Alcune parti di questo problema sembrano estremamente trascurate. Per i tipi di interventi che ci entusiasmano di più, pensiamo che ci siano circa 100-1000 persone che ci lavorano a tempo pieno, a seconda di come li si conta. Ma non è chiaro quanto questo renda il problema trascurato nel suo complesso. Molti, molti più ricercatori e consulenti lavorano per migliorare il processo decisionale in generale (ad esempio, aiutando le aziende ad assumere meglio). E molti degli attori esistenti hanno interessi acquisiti nel modo in cui le istituzioni prendono le decisioni, il che può indurli a resistere a certe riforme.

Risolvibilità

Fare progressi nel migliorare il processo decisionale in situazioni di grande rischio sembra moderatamente fattibile. Esistono tecniche che hanno dimostrato di poter migliorare il processo decisionale e i risultati ottenuti in passato suggeriscono che un maggiore finanziamento diretto ai migliori ricercatori in questo settore potrebbe essere molto fruttuoso. Tuttavia, alcune di queste tecniche potrebbero presto incontrare un muro nella loro utilità, e non è chiaro quanto sarà facile far sì che le pratiche decisionali migliorate vengano implementate nelle istituzioni cruciali.

Profondità del profilo: Poco profondo

Qual è il problema?

La nostra capacità di risolvere i problemi del mondo si basa molto sulla nostra capacità di comprenderli e di prendere decisioni di alta qualità. Dobbiamo essere in grado di identificare i problemi su cui lavorare, di capire quali fattori contribuiscono a questi problemi, di prevedere quali delle nostre azioni avranno i risultati desiderati e di rispondere al feedback e cambiare idea.

Molti dei problemi più importanti del mondo sono incredibilmente

complicati e richiedono la comprensione di sistemi complessi e interconnessi, la capacità di fare previsioni ragionevoli sull'esito di azioni diverse e la capacità di bilanciare considerazioni contrastanti e di riunire parti diverse per risolverli. Ciò significa che ci sono molte possibilità di commettere errori di valutazione.

Inoltre, in molte aree ci possono essere dubbi sostanziali persino sulla bontà o meno di un'azione: ad esempio, lavorare su modelli di grandi dimensioni e all'avanguardia per comprenderli meglio e allineare i loro obiettivi con i valori umani aumenta o diminuisce il rischio dell'IA? Un modo migliore per rispondere a queste domande sarebbe estremamente prezioso.

Prendere decisioni informate è difficile. Persino gli esperti di politica a volte fanno peggio di semplici previsioni attualistiche, quando stimano le probabilità di eventi fino a cinque anni nel futuro.

Le organizzazioni più adatte a risolvere i problemi più importanti del mondo - come i governi - sono spesso altamente burocratizzate, il che significa che i decisori devono affrontare molti vincoli e incentivi concorrenti, che non sempre sono allineati con un migliore processo decisionale.

Ci sono anche grandi sfide epistemiche - sfide che hanno a che fare con il modo di interpretare le informazioni e i ragionamenti - coinvolte nel comprendere e affrontare i problemi del mondo. Ad esempio, come si fa a dare una probabilità a un evento senza precedenti? Come si modificano le proprie convinzioni nel tempo - ad esempio, sul livello di rischio in un campo come l'intelligenza artificiale - quando persone informate non sono d'accordo sulle implicazioni degli sviluppi in questo campo? Come si fa a decidere tra, o a bilanciare, le opinioni di persone diverse con informazioni diverse (o parzialmente sovrapposte)? Quanta fiducia dovremmo avere in diversi tipi di argomentazioni?

E quali sono gli strumenti decisionali, le regole, le strutture e le euristiche che possono impostare al meglio le organizzazioni per aumentare sistematicamente le possibilità di ottenere buoni risultati per il futuro a lungo termine, che riteniamo estremamente importante? Come possiamo prendere decisioni sensate, visto che è difficile prevedere gli eventi anche a 5 o 10 anni di distanza?

E ci sono processi migliori per l'attuazione di tali decisioni, in cui le stesse conclusioni possono portare a risultati migliori?

Pensiamo che migliorare il ragionamento interno e la competenza decisionale di istituzioni chiave - alcune agenzie governative, aziende potenti e che lavorano con tecnologie rischiose - sia particolarmente cruciale. Se abbiamo ragione nel ritenere che i rischi che affrontiamo come società sono sostanziali e che queste istituzioni possono avere un ruolo di primo piano nella loro gestione, l'ampia scala del loro impatto rende anche piccoli miglioramenti di grande effetto.

La sfida di migliorare il processo decisionale abbraccia quindi diverse aree problematiche. Una migliore epistemica e un migliore processo decisionale sembrano importanti per ridurre la possibilità di conflitti tra grandi potenze, per comprendere i rischi dell'IA e capire cosa fare al riguardo, per progettare politiche che aiutino a prevenire pandemie catastrofiche, per coordinare la lotta al cambiamento climatico e, in pratica, per ogni altra area di culto e ad alta posta in gioco in cui ci sono più attori con informazioni incomplete che cercano di coordinarsi per fare progressi - cioè la maggior parte delle questioni che riteniamo più urgenti.

Quali sono le parti del problema su cui sembra più promettente lavorare?

"Migliorare il processo decisionale" è un termine ampio. Purtroppo, le persone con cui abbiamo parlato non sono molto d'accordo su cosa sia meglio concentrarsi in modo più specifico. A causa di questa mancanza di chiarezza e di consenso, nutriamo un certo scetticismo sul fatto che questo problema, così come lo abbiamo definito, sia altrettanto urgente di problemi meglio definiti come la sicurezza dell'intelligenza artificiale e il rischio nucleare.

Tuttavia, alcune idee in questo spazio sono abbastanza interessanti da farci pensare che siano promettenti, anche se non siamo così convinti di consigliare di lavorarci come per altri interventi.

Ecco alcune aree (parzialmente sovrapposte) su cui potrebbe valere la pena di lavorare:

Previsioni

Una nuova scienza delle previsioni si sta sviluppando per migliorare la capacità delle persone di fare previsioni più ponderate sugli eventi futuri, sviluppando strumenti e tecniche e modi per combinarli in previsioni aggregate. Questo campo cerca di rispondere a domande come, ad esempio, qual è la probabilità che la guerra in Ucraina finisca nel 2023? (Al momento in cui scriviamo è ancora in corso), oppure qual è la probabilità che un'azienda sia in grado di costruire un sistema di IA che possa fungere da "assistente esecutivo IA" commercialmente valido entro il 2030?

Le definizioni sono scivolose e molte cose vengono chiamate "previsioni". Ad esempio, le aziende prevedono continuamente la domanda e l'offerta, la spesa e altri fattori importanti per i loro profitti. Ciò che sembra speciale in alcuni tipi di previsione è l'attenzione a trovare i modi migliori per ottenere previsioni informate e complete per eventi davvero importanti ma complessi, senza precedenti o altrimenti difficili da prevedere con metodi consolidati.

Un approccio a questa categoria è la ricerca di "superprevisionari" da parte del dottor Phil Tetlock: persone che hanno dimostrato di essere insolitamente brave a prevedere gli eventi. Tetlock e i suoi collaboratori hanno esaminato cosa fanno in pratica questi superprevisionari per fare previsioni così buone e hanno sperimentato modi per migliorare le loro prestazioni. Due dei loro risultati: i migliori previsionari tendevano a prendere in considerazione molti tipi di informazioni e a combinarle in un'ipotesi complessiva, piuttosto che concentrarsi, ad esempio, su alcuni studi o estrapolare una tendenza; e un'aggregazione di previsioni di superprevisionari tendeva a essere ancora più accurata dei migliori singoli esperti di previsioni.

Lo sviluppo e l'implementazione di una migliore previsione degli eventi importanti e delle pietre miliari del progresso tecnologico sarebbe estremamente utile per prendere decisioni più informate e ad alto rischio, e questa nuova scienza sta iniziando a fare progressi. Ad esempio, l'Intelligence Advanced Research Projects Activity (IARPA) degli Stati Uniti si è interessata al settore nel 2021, quando ha organizzato un torneo di previsione. La squadra vincitrice è stata quella di Tetlock, il Good Judgement Project, che si è trasformato in un'azienda commerciale di previsioni, la Good Judgement Inc.

Ecco un esempio diverso di utilizzo delle previsioni, tratto dalla nostra

esperienza: una delle questioni più importanti su cui ci proponiamo di aiutare le persone a lavorare è la riduzione delle minacce esistenziali derivanti dall'IA trasformativa. Ma la maggior parte del lavoro in quest'area sarà probabilmente più utile prima che i sistemi di IA raggiungano livelli trasformativi di potenza e intelligenza. Per questo motivo, il momento in cui avverrà questa transizione (che potrebbe essere graduale) è un'importante considerazione strategica. Se pensassimo che l'IA trasformativa sia dietro l'angolo, sarebbe più sensato cercare di lavorare con i professionisti a metà carriera per spingerli a passare a lavorare su questo tema, dato che hanno già molta esperienza. Ma se queste IA non arriveranno prima di qualche decennio, c'è tempo per aiutare una nuova generazione di persone interessate a entrare nel settore e a prepararsi.

Data questa situazione, teniamo d'occhio le previsioni dei ricercatori nel campo della sicurezza dell'IA e del ML; ad esempio, le previsioni di Ajeya Cotra basate su leggi di scala empiriche e ancora biologiche (come i requisiti computazionali del cervello umano), i sondaggi degli esperti e le previsioni di un gruppo di "superforecasting" chiamato Samotsvety. Ognuna di queste previsioni utilizza una metodologia diversa e presenta molti punti interrogativi, e non è chiaro quale sia la migliore; ma sono la migliore risorsa che abbiamo a disposizione per questa decisione. Finora le previsioni suggeriscono che c'è una discreta probabilità che ci sia abbastanza tempo per aiutare i più giovani a entrare nelle carriere d'élite e iniziare a contribuire. Tuttavia, maggiori e migliori previsioni in questo settore, così come maggiori indicazioni sulle migliori pratiche per combinare le diverse previsioni, sarebbero estremamente preziose e rilevanti per le decisioni.

La politica è un'altra area in cui le previsioni sembrano importanti. Le previsioni sugli effetti delle diverse politiche proposte sono l'ingrediente principale che ne determina il valore, ma spesso queste previsioni sono molto aleatorie. Naturalmente, questo non è l'unico ostacolo che si frappone a una migliore definizione delle politiche: per esempio, a volte le buone politiche si scontrano con ostacoli politici o necessitano di ulteriori ricerche. Ma prevedere gli effetti complessi di diverse politiche è una delle sfide fondamentali della governance e della società civile.

Quindi, migliorare la nostra capacità di prevedere eventi importanti e complessi sembra decisamente utile. Tuttavia, i maggiori progressi sono stati fatti finora su previsioni relativamente ben definite nel futuro prossimo. Non è ancora dimostrato che possiamo fare previsioni sostanzialmente migliori della probabilità su questioni o eventi più complessi o amorfi con un orizzonte temporale superiore a qualche anno, come ad esempio "quando l'IA trasformerà l'economia, se la trasformerà?".

Inoltre, non sappiamo molto sulla misura in cui metodi di previsione migliori potrebbero migliorare il processo decisionale nella pratica, soprattutto in ambienti istituzionali di grandi dimensioni. Quali sono i modi migliori e più realistici in cui le agenzie, le organizzazioni non profit e le aziende più importanti possono utilizzare sistematicamente le previsioni per migliorare il loro processo decisionale? Quali proposte saranno più interessanti per la leadership e più facili da attuare?

Sembra anche possibile che questa nuova scienza si scontri con un muro. Forse abbiamo imparato la maggior parte di ciò che impareremo, o forse l'adozione delle tecniche è troppo intrattabile per valere la pena di provarci. Non lo sappiamo ancora.

Quindi: non sappiamo fino a che punto il lavoro sul miglioramento delle previsioni sarà in grado di aiutare a risolvere i grandi problemi globali - ma l'area è abbastanza piccola e nuova che sembra possibile fare una differenza sostanziale al margine, e lo sviluppo di questo campo potrebbe rivelarsi molto prezioso.

Aggregatori e mercati di previsione

I mercati di previsione utilizzano i meccanismi di mercato per aggregare le opinioni di molte persone sul verificarsi di un evento o di un altro. Ad esempio, i mercati di previsione Manifold Markets e Polymarket permettono agli utenti di scommettere o fare previsioni su domande come "chi vincerà le elezioni presidenziali statunitensi nel 2024?" o anche ["Assisteremo a un'esplosione nucleare entro la metà del 2023?"](<https://polymarket.com/event/will-a-nuclear-weapon-detonate-by-june-30>).

Gli utenti che prevedono correttamente gli eventi possono guadagnare denaro, e possono ottenere guadagni maggiori in proporzione a quanto inaspettato era il risultato che avevano previsto.

L'idea è che i mercati di previsione forniscano un forte incentivo per gli individui a fare previsioni accurate sul futuro. In aggregato, le tendenze di queste previsioni dovrebbero fornire informazioni utili sulla probabilità degli esiti futuri in questione, sempre che il mercato sia ben progettato e funzioni correttamente.

Esistono anche altri aggregatori di previsioni, come Metaculus, che utilizzano la reputazione come incentivo piuttosto che il denaro (anche se Metaculus ospita anche tornei di previsioni con premi in denaro) e che sono specializzati nell'aggregazione e nell'analisi delle previsioni in base ai precedenti per migliorare le previsioni collettive.

In teoria, i mercati delle previsioni sono molto accurati quando c'è un alto volume di utenti, così come il mercato azionario è molto bravo a valutare le azioni delle società. Nello scenario ideale, potrebbero essere in grado di fornire un quadro autorevole e condiviso delle conoscenze aggregate del mondo su molti argomenti. Immaginate di non essere più in disaccordo su ciò che "pensano gli esperti" al pranzo del Ringraziamento - e immaginate che questi disaccordi evaporino nella sala riunioni della Casa Bianca.

Ma i mercati di previsione o altri tipi di aggregatori non esistono ancora su larga scala per molti argomenti, quindi la promessa è ancora in gran parte teorica. Inoltre, non sarebbero una soluzione perfetta: come i mercati azionari, potremmo vederli muoversi ma non sapere perché; e non ci direbbero molto su argomenti per i quali le persone non hanno molte ragioni per pensare una cosa o l'altra. Allo stesso modo, aggiungono un'altra serie di complessi incentivi al processo decisionale politico: molti mercati di previsione possono anche incentivare le cose che prevedono. Inoltre, anche se venissero creati mercati di previsione ampi e accurati, gli attori istituzionali più importanti ne terrebbero conto?

Si tratta di sfide importanti, ma le argomentazioni teoriche a favore dei mercati di previsione e di altri aggregatori sono convincenti e saremmo entusiasti di vedere i mercati guadagnare una trazione sufficiente per testare a fondo i loro potenziali benefici.

Scienza del processo decisionale

Un'altra categoria è quella dello sviluppo di strumenti e tecniche per prendere decisioni migliori - ad esempio le "tecniche analitiche strutturate" (SAT) come il metodo Delphi, che aiuta a costruire il consenso nei gruppi utilizzando più iterazioni di domande per raccogliere dati dai diversi membri del gruppo, e la creazione di tecniche di crowdsourcing come SWARM ("Smartly-assembled Wiki-style Argument Marshalling"). (È possibile consultare gli usi del metodo Delphi sul sito web della RAND Corporation, che lo sviluppò per la prima volta negli anni Cinquanta).

Questi strumenti potrebbero aiutare le persone a evitare fallimenti cognitivi come il bias di conferma, ad aggregare le opinioni o a raggiungere il consenso, o a fermare le "cascate di informazioni".

Molte di queste tecniche sono già in uso. Ma sembra probabile che si possano sviluppare strumenti migliori e sarebbe davvero sorprendente se non ci fossero più contesti istituzionali in cui introdurre le tecniche giuste.

Migliorare la governance aziendale delle principali istituzioni private

La governance aziendale potrebbe essere particolarmente importante con lo sviluppo di tecnologie trasformative e potenzialmente pericolose nei settori dell'IA e della bioingegneria.

Il processo decisionale risponde agli incentivi, che spesso possono essere influenzati

dalla struttura e dalla governance di un'azienda. Ad esempio, le aziende a scopo di lucro sono tenute in ultima analisi a massimizzare il profitto degli azionisti, anche se ciò comporta ampi effetti negativi su altri ("esternalità"), mentre le aziende non profit non lo sono.

Ad esempio, il laboratorio di IA OpenAI ha impostato la sua struttura di governance con l'obiettivo esplicito di non far prevalere i profitti sulla missione codificata nel suo statuto (che i sistemi di IA generale vadano a beneficio di tutta l'umanità), ponendo la sua entità commerciale sotto il controllo del suo consiglio di amministrazione non profit.³

Non siamo sicuri che sia abbastanza per ridurre la possibilità di esternalità negative dalle operazioni di OpenAI e prendiamo sul serio le preoccupazioni sollevate da altri riguardo alle azioni dell'azienda che potrebbero portare a dinamiche di corsa agli armamenti, ma ci sembra più accettabile di una configurazione commerciale tradizionale per modellare gli incentivi aziendali in una buona direzione e saremmo entusiasti di vedere altre innovazioni e l'interesse delle aziende in questo spazio. (Anche se non siamo sicuri che il miglioramento dei meccanismi di governance sia più promettente della difesa e dell'educazione sui rischi potenziali - entrambi ci sembrano utili).

Migliorare il funzionamento dei governi

I governi controllano vaste risorse e sono talvolta gli unici in grado di lavorare su questioni di portata globale. Ma le loro decisioni sembrano sostanzialmente influenzate dalle strutture e dai processi che li compongono. Se potessimo migliorare questi processi, soprattutto in situazioni ad alta posta in gioco o in rapido movimento, sarebbe molto utile.

Ad esempio, durante la pandemia di COVID-19, negli Stati Uniti non sono stati disponibili test rapidi affidabili per molto tempo rispetto ad altri Paesi simili, rallentando probabilmente la capacità degli Stati Uniti di gestire la pandemia. Procedure decisionali istituzionali migliori, soprattutto in caso di emergenza, avrebbero potuto migliorare la situazione?

Affrontare questo problema richiederebbe una conoscenza dettagliata del funzionamento di istituzioni complesse e burocratiche come il governo statunitense e dei suoi quadri normativi. Dov'è stato esattamente il collo di bottiglia nel caso dei test rapidi per il COVID? Quali sono i colli di bottiglia più importanti da affrontare prima della prossima pandemia?

Una seconda area è la riforma del voto. Spesso eleggiamo i nostri leader con il metodo del "primo turno", ma molti sostengono che questo porti a risultati distorti. Metodi di voto diversi potrebbero portare a governi più funzionali e rappresentativi in generale, il che potrebbe essere utile per una serie di questioni. Per saperne di più, consultate il nostro podcast con Aaron Hamlin.

C'è anche l'importanza della sicurezza del voto per evitare elezioni contestate, di cui abbiamo parlato nella nostra intervista con Bruce Schneier.

Infine, ci sono interventi volti a far sì che i meccanismi di governance tengano sistematicamente conto degli interessi delle generazioni future. In questo momento le persone del futuro non hanno potere politico commerciale o formale, quindi dobbiamo aspettarci che i loro interessi siano poco tutelati. Alcuni tentativi di combattere questo fenomeno includono il Commissario per le generazioni future in Galles e la legge sul benessere delle generazioni future nel Regno Unito.

In definitiva, non siamo sicuri di quali siano le leve migliori su cui premere. Sembra chiaro che ci sono problemi nella capacità della società di ragionare su problemi grandi e complessi e di agire in modo appropriato, soprattutto quando le decisioni devono essere prese rapidamente, quando ci sono motivazioni contrastanti, quando è necessaria la cooperazione e quando le questioni riguardano soprattutto il futuro. E siamo incuriositi da alcuni dei progressi che abbiamo visto.

Ma alcuni dei migliori interventi potrebbero anche essere specifici per determinate

aree problematiche, come ad esempio lavorare su una migliore previsione dello sviluppo dell'IA.

E la trattabilità rimane un problema anche quando vengono sviluppati metodi migliori. Sebbene esistano vari esempi di istituzioni che adottano tecniche per migliorare la qualità del processo decisionale, sembra difficile convincere gli altri ad adottare le migliori pratiche, soprattutto se i benefici sono difficili da dimostrare, se non si conoscono abbastanza bene le organizzazioni o se non hanno una motivazione adeguata per adottare tecniche migliori.

Quali sono i principali fattori che impediscono di migliorare il processo decisionale?

Non è chiaro quali siano gli interventi migliori - e alcuni non sono molto considerati.

Come abbiamo detto sopra, non siamo sicuri di cosa sia meglio in questo settore.

Uno dei motivi per cui dubitiamo di alcuni tipi di lavoro che potrebbero rientrare in questa categoria è che sembra che dovrebbero essere gestiti dai meccanismi di mercato.

Un'euristica che usiamo spesso per dare priorità ai problemi globali è: il nostro sistema economico incentiverà le persone a risolvere questo problema da sole? Se è così, è meno ovvio che le persone debbano lavorare su questo problema per cercare di migliorare il mondo. Questo è un modo per stimare la futura trascuratezza di un problema. Se c'è da guadagnare risolvendo un problema, ed è possibile risolverlo, ha buone probabilità di essere risolto.

Quindi ci si potrebbe chiedere: molte aziende non dovrebbero avere interesse a migliorare il processo decisionale, anche per quanto riguarda eventi importanti e complessi nel futuro a medio e lungo termine? Prevedere gli sviluppi futuri del mondo, aggregare le opinioni degli esperti e, in generale, essere in grado di prendere meglio le decisioni ad alto rischio sembra incredibilmente utile in generale, non solo per scopi altruistici.

Questa è un'ottima domanda, ma si applica più ad alcuni interventi che ad altri. Per esempio, lavorare per modellare gli incentivi dei governi in modo che tengano maggiormente conto degli interessi delle generazioni future sembra molto improbabile che possa essere affrontato dal mercato. Lo stesso vale per una migliore governance aziendale.

Migliori metodi di previsione e mercati di previsione sembrano più validi dal punto di vista commerciale. In effetti, il fatto che non siano stati sviluppati maggiormente sotto la pressione competitiva del mercato potrebbe essere considerato una prova a sfavore della loro efficacia. I mercati di previsione, in particolare, sono stati teorizzati per la prima volta 20 anni fa: forse se fossero stati utili sarebbero già diventati mainstream? Anche se questo potrebbe essere in parte dovuto al fatto che negli Stati Uniti sono spesso considerati una forma di gioco d'azzardo illegale, e la nostra ipotesi è che siano ancora validi e vengano sottovalutati.

Gli strumenti della scienza delle decisioni sono probabilmente i meno trascurati tra le sotto-aree discusse in precedenza e, in pratica, molte industrie utilizzano questi metodi e i think tank e le società di consulenza li sviluppano e li formano.

Detto questo, il problema generale sembra ancora irrisolto e anche i metodi che sembrano dover ricevere un sostegno commerciale non hanno ancora preso piede. Inoltre, la nostra capacità di prendere decisioni sembra più inadeguata nelle situazioni

più rischiose, che spesso implicano la riflessione su incognite sconosciute, piccole probabilità di esiti catastrofici e una profonda complessità - il che suggerisce che è necessaria una maggiore innovazione. Inoltre, sebbene l'uso di diverse tecniche decisionali sia diffuso in tutta la società, siamo molto lontani dal fatto che tutte le istituzioni più importanti utilizzino i migliori strumenti a loro disposizione.

Quindi, anche se c'è un ampio interesse a migliorare il processo decisionale, sarebbe sorprendente per noi se non ci fosse un ampio margine di miglioramento rispetto allo status quo. Quindi, in definitiva, siamo entusiasti di vedere ulteriori lavori in questo campo, nonostante questi dubbi.

Forse è meglio lavorare più direttamente su un problema urgente.

Supponiamo che pensiate che allo stato attuale il cambiamento climatico sia il problema più importante del mondo. Potreste pensare che una parte importante del motivo per cui non riusciamo ad affrontare il cambiamento climatico in modo efficace è che le persone hanno una propensione a lavorare su problemi concreti e a breve termine piuttosto che su quelli che hanno maggiori probabilità di influire sulle generazioni future, e che questo è un problema sistematico del nostro processo decisionale come società.

Si potrebbe quindi pensare di fare ricerca su come superare questo pregiudizio, con la speranza di rendere le istituzioni importanti più propense ad affrontare il cambiamento climatico.

Tuttavia, se pensate che la minaccia del cambiamento climatico sia particolarmente urgente rispetto ad altri problemi, questo potrebbe non essere il modo migliore per dare il proprio contributo. Anche se scoprite una tecnica utile per ridurre la tendenza a lavorare su problemi immediati, potrebbe essere molto difficile implementarla in modo da ridurre direttamente l'impatto del cambiamento climatico.

In questo caso, è probabile che sia meglio concentrare i propri sforzi sul cambiamento climatico in modo più diretto, ad esempio lavorando per un think tank che fa ricerca sui modi più efficaci per ridurre le emissioni di carbonio o sviluppando tecnologie verdi. In generale, gli interventi più diretti sembrano avere maggiori probabilità di spostare l'ago della bilancia su problemi particolari, perché sono più focalizzati sui colli di bottiglia più urgenti in quell'area.

Detto questo, se non è possibile implementare soluzioni a un problema senza migliorare il ragionamento e i processi decisionali coinvolti, può rivelarsi un passaggio necessario.

Il vantaggio principale di interventi di ampio respiro come il miglioramento del processo decisionale è che possono essere applicati a un'ampia gamma di problemi. Lo svantaggio corrispondente è che potrebbe essere più difficile indirizzare gli interventi verso un problema specifico. Quindi, se si ritiene che un problema specifico sia significativamente più urgente di altri e si ha l'opportunità di lavorare su quel problema in modo più diretto, è probabile che sia più efficace farlo.

Vale la pena notare che in alcuni casi questa distinzione non è particolarmente significativa. Lo sviluppo di buoni meccanismi di corporate governance che consentano alle aziende che sviluppano tecnologie ad alto rischio di cooperare meglio tra loro è un modo per migliorare il processo decisionale. Ma potrebbe anche essere uno dei modi migliori per ridurre direttamente i rischi catastrofici dell'IA.

Spesso è difficile introdurre il cambiamento

nella pratica

Forse il problema principale di questo settore è che non è chiaro quanto sia facile attuare nella pratica strategie decisionali migliori, soprattutto nelle organizzazioni burocratiche e dove gli incentivi non sono orientati all'accuratezza.

Spesso è difficile convincere i gruppi a implementare pratiche che sono costose per loro nel breve termine - che richiedono tempo, risorse e a volte sfidano chi ha attualmente degli interessi in quel campo - mentre promettono solo benefici astratti o a lungo termine. Ci possono essere anche altri vincoli: come già detto, negli Stati Uniti la gestione di mercati di previsione con denaro reale è solitamente considerata illegale.

Tuttavia, i problemi di implementazione possono essere superati se si riesce a dimostrare ai decisori che le tecniche li aiuteranno a raggiungere gli obiettivi che stanno loro a cuore.

Se lavorate o siete in grado di influenzare in altro modo l'assetto di un'istituzione, potreste anche contribuire a modificare le pratiche decisionali creando incentivi istituzionali che favoriscano la deliberazione e la ricerca della verità.

Tuttavia, riconoscendo le difficoltà di ottenere un cambiamento nella pratica, sembra particolarmente importante per chi pensa a questo problema sviluppare una comprensione approfondita di come operano i gruppi e le istituzioni importanti e dei tipi di incentivi e barriere che devono affrontare. Sembra plausibile che il superamento delle barriere burocratiche che impediscono un migliore processo decisionale sia ancora più importante dello sviluppo di tecniche migliori.

Potrebbero esserci modi migliori per migliorare la nostra capacità di risolvere i problemi del mondo.

Uno dei principali argomenti a favore del lavoro in questo settore è che, se si riesce a migliorare le capacità decisionali delle persone che lavorano su problemi importanti, allora aumenta l'efficacia di tutto ciò che queste fanno per risolvere quei problemi.

Ma si potrebbe pensare che ci siano modi migliori per aumentare la velocità o l'efficacia del lavoro sui problemi più importanti del mondo.

Per esempio, forse il più grande collo di bottiglia nella risoluzione dei problemi del mondo non è la scarsa capacità decisionale, ma semplicemente la mancanza di informazioni: è possibile che le persone non stiano lavorando ai problemi più importanti del mondo perché mancano informazioni cruciali su quali siano questi problemi. In questo caso, l'area più importante potrebbe essere la ricerca sulle priorità globali, ovvero la ricerca per identificare le questioni in cui le persone possono fare la massima differenza.

Inoltre, molto lavoro sulla costruzione di un altruismo efficace potrebbe rientrare in questa categoria: fornire alle persone migliori informazioni sull'efficacia di diverse cause, interventi e carriere, nonché diffondere valori positivi come l'attenzione alle generazioni future.

Cosa si può fare in questo ambito?

Il lavoro in questo campo rientra in diverse categorie generali:

- Testare in modo più rigoroso le tecniche esistenti che sembrano promettenti.

- Fare più ricerca di base per individuare nuove tecniche.
- Promuovere l'adozione delle migliori tecniche collaudate nelle aree ad alto impatto.
- Orientare maggiori finanziamenti verso i modi migliori per fare tutto ciò.

Tutte queste strategie sembrano promettenti e potrebbero portare a progressi immediati (sappiamo già abbastanza per iniziare a cercare di implementare tecniche migliori, ma prove più solide renderanno più facile l'adozione, ad esempio). Ciò significa che la scelta dell'area su cui concentrarsi dipenderà in larga misura dal vostro personale e dalle particolari opportunità a vostra disposizione.

1. Testare in modo più rigoroso le tecniche esistenti che sembrano promettenti

Abbiamo esaminato alcune aree in cui si potrebbero avere diversi tipi di interventi. Pensiamo che testare in modo più rigoroso i risultati provvisori in tutte queste aree, soprattutto quelle meno sviluppate, potrebbe essere molto utile.

Qui parleremo solo di alcuni esempi di tecniche più formali, perché sono quelle che conosciamo meglio:



L'idea è quella di prendere tecniche che sembrano promettenti, ma che non sono ancora state testate in modo rigoroso, e cercare di ottenere prove più solide su come e se sono valide. Alcune tecniche o aree di ricerca che rientrano in questa categoria:

Formazione sulla calibrazione — un modo per migliorare i giudizi di probabilità - ha una quantità ragionevole di prove che ne suggeriscono l'efficacia. Tuttavia, la maggior parte dei corsi di calibrazione si concentra su domande banali: verificare se questo tipo di formazione migliora effettivamente il giudizio in scenari reali potrebbe essere promettente e potrebbe contribuire a far sì che queste tecniche vengano applicate più ampiamente.

- Il “Pastcasting” è un metodo per ottenere una pratica di previsione più realistica utilizzando domande di previsione reali, ma già risolte.
- Tecniche analitiche strutturate (SAT) — Ad esempio, la verifica delle ipotesi chiave e la messa in discussione delle opinioni di consenso. Queste tecniche sembrano basarsi su una comprensione della letteratura psicologica, ma poche sono state testate in modo rigoroso (ad esempio con un gruppo di controllo e analizzando l'impatto sull'accuratezza). Potrebbe essere utile selezionare alcune di queste tecniche che sembrano più promettenti e provare a verificare quali sono effettivamente efficaci nel migliorare i giudizi del mondo reale.⁴ Potrebbe essere particolarmente interessante e utile provare a mettere direttamente alcune di queste tecniche una contro l'altra e confrontare i loro livelli di successo.
- Metodi di aggregazione dei giudizi degli esperti, tra cui il modello classico di Roger Cooke per il giudizio strutturato degli esperti (che assegna un punteggio ai diversi giudizi in base alla loro accuratezza e informatività e poi utilizza questi punteggi per combinarli),⁵ mercati di previsione o il già citato metodo Delphi.
- Migliori indagini di esperti in aree importanti come l'IA e il biorischio, ad esempio sul modello del gruppo di esperti economici dell'IGM.

Alcuni ricercatori accademici e altri gruppi che lavorano in questo settore:

- Il professor Philip Tetlock dell'Università della Pennsylvania, che abbiamo citato in precedenza e che è stato il pioniere della ricerca sul miglioramento delle previsioni.
- L'istituto Forecasting Research progetta e testa tecniche di previsione con l'obiettivo di farle applicare dai politici e dalle organizzazioni non profit.
- Il professor Robin Hanson, economista della George Mason University, che ha svolto ricerche e sostenuto l'uso dei mercati di previsione, ha dichiarato: "Non è possibile che i mercati di previsione non siano in grado di fornire un'informazione adeguata".
- Il professor George Wright dell'Università Strathclyde di Glasgow, uno psicologo che svolge ricerche sull'applicazione del metodo Delphi e del pensiero di scenario per migliorare i giudizi.
- Stephen Coulthart dell'Università di Albany ha condotto una ricerca sull'efficacia dei diversi SAT nell'analisi dell'intelligence.⁶
- Esistono alcuni gruppi che si concentrano sulla ricerca decisionale più applicata e che possono disporre di fondi per condurre esperimenti di grandi dimensioni, in particolare nelle business school, tra cui: il Behavioural Insights Group della Harvard Kennedy School; il gruppo Operations, Information and Decisions della Wharton School dell'Università della Pennsylvania; il Decision Sciences Group della Fuqua School of Business della Duke; il Center for Decision Research della Booth School of Business dell'Università di Chicago; il Behavioural Science Group della Warwick Business School.

Potrebbero esserci anche organizzazioni non accademiche con finanziamenti e interesse a condurre test più rigorosi di tecniche decisionali note:

- L'Intelligence Advanced Research Projects Activity (IARPA) è probabilmente il più grande finanziatore della ricerca in questo settore, soprattutto per quanto riguarda il miglioramento delle decisioni di alto livello.
- Open Philanthropy, che finanzia anche 80,000 Hours, a volte concede sovvenzioni per testare e sviluppare migliori tecniche decisionali, in particolare per aiutare le persone a prendere decisioni migliori sul futuro.
- Anche le società di consulenza che si occupano di scienze comportamentali, come il Behavioural Insights Team, possono avere finanziamenti e interesse a svolgere questo tipo di ricerca. Queste organizzazioni si concentrano generalmente sul miglioramento di molte piccole decisioni, piuttosto che sul miglioramento della qualità di poche decisioni molto importanti, ma possono svolgere un lavoro su quest'ultimo aspetto.

2. Fare più ricerca di base per identificare nuove tecniche

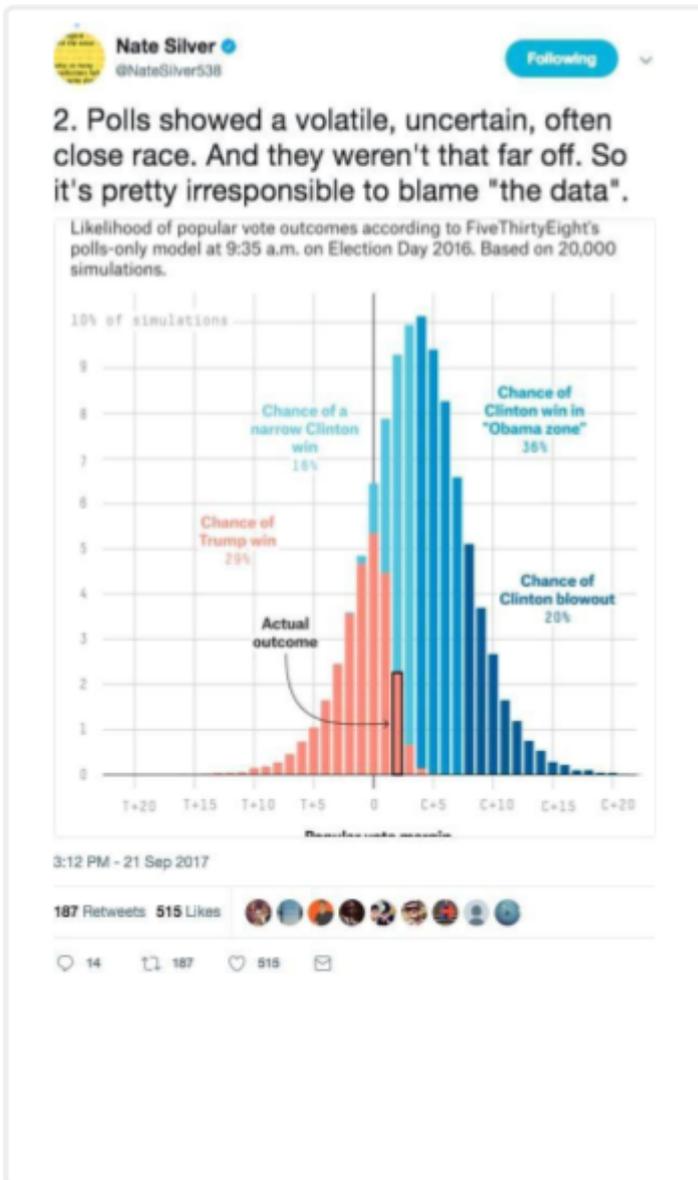
Si potrebbe anche cercare di fare ricerca più sostanziale: sviluppare nuovi approcci per migliorare l'epistemica e il processo decisionale, e poi testarli. Questo è più urgente se non si pensa che gli interventi esistenti, come quelli elencati sopra, siano molto validi.

Un esempio di domanda aperta in quest'area è: come possiamo giudicare un "buon ragionamento" quando non abbiamo risposte oggettive a una domanda? Ovvero, quando non possiamo limitarci a giudicare le risposte o i contributi in base al fatto che portino a previsioni accurate o a risposte che sappiamo essere vere?

Due esempi di programmi di ricerca in corso legati a questa domanda sono il programma Crowdsourcing Evidence, Argumentaion, Thinking and Evaluation (CREATE) della IARPA e il progetto Making Conversations Smarter, Faster (MCSF) di Philip Tetlock.

Anche gli accademici e le istituzioni sopra elencate potrebbero essere luoghi promettenti per lavorare o richiedere finanziamenti se siete interessati a sviluppare nuove tecniche decisionali.

3. Promuovere l'adozione delle migliori tecniche collaudate nelle istituzioni ad alto impatto.



In alternativa, potreste concentrarvi maggiormente sull'implementazione di quelle tecniche che attualmente ritenete più in grado di migliorare il processo decisionale collettivo (come la ricerca sulla previsione di Tetlock, i mercati di previsione o i SAT).

Se si ritiene che un problema specifico sia particolarmente importante, si potrebbe preferire concentrarsi sull'implementazione di tecniche (piuttosto che sullo sviluppo di nuove), in quanto è più facile indirizzarle verso aree specifiche.

Come già accennato, gran parte della "promozione dell'adozione" potrebbe

richiedere una migliore comprensione dei vincoli pratici e degli incentivi dei diversi gruppi che lavorano su problemi importanti, al fine di capire quali cambiamenti sono probabilmente realizzabili. Per questo motivo, lavorare in una delle organizzazioni o dei gruppi elencati di seguito - con l'obiettivo di comprendere meglio le barriere che devono affrontare e creare connessioni - potrebbe essere utile, anche se non ci si aspetta di essere in grado di cambiare immediatamente le pratiche decisionali.

Queste iniziative potrebbero avere un impatto particolare se concentrate su organizzazioni che controllano molte risorse o su organizzazioni che lavorano su problemi importanti. Ecco alcuni esempi di luoghi specifici in cui potrebbe essere utile lavorare se si vuole fare questo:

- Qualsiasi governo di rilievo, in particolare nei seguenti settori: (1) intelligence/sicurezza nazionale e politica estera; (2) parti del governo che si occupano di politica tecnologica (il Government Office for Science nel Regno Unito, ad esempio, o l'Office of Science and Technology Policy negli Stati Uniti); (3) politica di sviluppo (DFID nel Regno Unito, USAID negli Stati Uniti); o (4) difesa (il Dipartimento della Difesa negli Stati Uniti, o il Ministero della Difesa nel Regno Unito).
- Organizzazioni che destinano grandi quantità di denaro alla soluzione di importanti problemi mondiali, come fondazioni o agenzie di sovvenzioni come la Gates Foundation o Open Philanthropy. Vedi anche il nostro profilo su come lavorare come sovventore di una fondazione.
- Alcune organizzazioni internazionali come l'ONU
- Aziende private coinvolte nella creazione di tecnologie trasformative, o le agenzie che le regolano

Potreste anche provare a testare e implementare tecniche decisionali migliorate in una serie di organizzazioni come consulenti. Alcune organizzazioni specifiche in cui potreste essere in grado di farlo, o almeno di accumulare un'esperienza rilevante, includono:

- Good Judgment, che organizza corsi di formazione per individui e organizzazioni per applicare i risultati della scienza delle previsioni e migliorare le previsioni.
- Lavorare in una società di consulenza specializzata in "scienze comportamentali", come il Behavioural Insights Team o Ideas42. Alcuni accademici di successo in questo campo hanno anche creato piccole società di consulenza, come Hubbard Decision Research (che ha lavorato molto sulla formazione in materia di calibrazione).
- HyperMind, un'organizzazione focalizzata su una più ampia adozione dei mercati di previsione.
- Si tratta di una società di consulenza più generale, con l'obiettivo di specializzarsi nell'aiutare le organizzazioni a migliorare il processo decisionale.

Un altro approccio potrebbe essere quello di cercare di ottenere un posto di lavoro presso un'organizzazione che ritenete stia svolgendo un lavoro davvero importante, e alla fine puntare a migliorare il suo processo decisionale. Nel nostro podcast con Danny Hernandez abbiamo ascoltato la storia di una persona che ha migliorato il processo decisionale di un'azienda dall'interno.

Infine, potreste anche cercare di promuovere l'adozione di pratiche migliori all'interno del governo e di organizzazioni molto importanti come i laboratori di IA o di biologia, o di migliorare il processo decisionale più in generale, se pensate di poter ottenere una buona piattaforma per farlo, lavorando come giornalisti, oratori o magari accademici in questo settore.

Julia Galef è un esempio di chi ha seguito questo tipo di percorso. Julia ha lavorato come giornalista freelance prima di fondare il Center for Applied Rationality. È co-conduttrice del podcast Rationally Speaking e ha un canale YouTube con centinaia di migliaia di follower. Ha anche scritto un libro che ci è piaciuto molto, "Scout Mindset", per aiutare i lettori a mantenere la loro curiosità e a evitare di rimanere radicati nelle loro convinzioni. Per saperne di più sul percorso professionale di Julia, consultate la nostra intervista.

4. Indirizzare maggiori finanziamenti alla ricerca in questo settore

Un altro approccio sarebbe quello di fare un passo indietro nella filiera e cercare di indirizzare maggiori finanziamenti verso il lavoro in tutte le aree sopra citate: sviluppare, testare e implementare migliori strategie decisionali.

Il luogo principale di cui siamo a conoscenza che sembra particolarmente interessato a indirizzare maggiori finanziamenti verso il miglioramento della ricerca sui processi decisionali è lo IARPA negli Stati Uniti. Diventare un responsabile di programma presso lo IARPA - se si è bravi e si hanno idee sulle aree di ricerca che potrebbero beneficiare di maggiori finanziamenti - è quindi un'opportunità molto promettente.

C'è anche la possibilità che Open Philanthropy investa più tempo o fondi nell'esplorazione di questo settore (in passato ha finanziato parte del lavoro di Tetlock sulle previsioni).

Altrimenti, potreste cercare di lavorare presso qualsiasi altra grande fondazione interessata a finanziare la ricerca scientifica, dove potrebbe esserci spazio per indirizzare i fondi verso questo settore.

Un ringraziamento speciale a Jess Whittlestone, che ha scritto la versione originale di questo problema e da cui è stato tratto gran parte del testo precedente.

Note e Riferimenti

"Nella previsione politica, Tetlock (2005) ha chiesto ai professionisti di stimare le probabilità di eventi fino a 5 anni nel futuro - dal punto di vista del 1988. Ci sarebbe stata una fine non violenta dell'apartheid in Sudafrica? Gorbačev sarebbe stato spodestato da un colpo di stato? Gli Stati Uniti sarebbero entrati in guerra nel Golfo Persico? Gli esperti hanno spesso avuto difficoltà a battere i semplici modelli attuariali o persino i dati di base casuali (si veda anche Green e Armstrong, 2007)". Mellers et. al. (2015).

La psicologia dell'analisi di intelligence: I fattori che determinano l'accuratezza delle previsioni nella politica mondiale. *Il Manuale di Psicologia Sperimentale: Applicato*. 21, 1, pp. 1- 14↵

Vedi questo post sul blog per un tentativo di valutare le differenze di performance nelle previsioni su orizzonti temporali più o meno lunghi. Purtroppo non c'è molto su cui basarsi, perché abbiamo pochissimi dati pubblici sulle previsioni a più di un anno di distanza.

Da notare che uno dei nostri consiglieri, Tasha McCauley, è anche membro del consiglio di amministrazione di OpenAI.