

Judea Pearl & Dana MacKenzie, *The book of why. The new science of cause and effect*, Basic Books, 2018, pp. 432, \$ 28,80

I dati, in quanto tali, non sono intelligenti. Non ci parlano, cioè non arrivano ai nostri apparati cognitivi con attaccate etichette che informino di quello che significano. Basterebbe avere capito l'abc dell'epistemologia scientifica per saperlo. Oggi, però, di dati si parla come se contenessero chissà quali verità. "Viviamo in un'era – dice l'informatico e filosofo Judea Pearl, vincitore del premio Turing nel 2011 per avere rivoluzionato l'approccio probabilistico all'intelligenza artificiale – che presuppone che i Big Data siano la soluzione a tutti i problemi. I corsi di 'scienza dei dati' proliferano nelle nostre università, e gli impieghi come 'scienziati dei dati' sono redditizi nelle nostre società, dove l'economia è 'basata su dati'". Di fatto, spiega in una intervista, gli strumenti statistici tradizionali, che guardano solo alle correlazioni, e gli algoritmi di Intelligenza Artificiale (IA) che usiamo per interrogare le banche dati sono come gli uomini "nella famosa caverna di Platone, [...] esplorano le ombre sulla parete della grotta e imparano a prevedere con precisione i loro movimenti. Ma non capiscono che le ombre osservate sono proiezioni di oggetti tridimensionali che si muovono in uno spazio tridimensionale". Ne deriva, per il teorico dell'IA basata su network bayesiani, che sono sciocchezze la "singolarità" (la super-intelligenza che prenderebbe il comando delle tecnologie), l'arrivo di legioni di robot che ci schiavizzeranno o un Armageddon causato dall'IA. L'IA, oggi, è solo in grado, con molta ma molta più efficienza dell'uomo, di rilevare la correlazione tra i dati. Il fatto che vinca a scacchi o a GO, che sappia progettare farmaci a livello molecolare, o guidare auto o fingere di essere un servizio clienti umano, dimostra solo che la gamma di domini dove questa capacità di uso superficiale dei dati si può applicare in modi adattativi, è più vasta di quanto inizialmente si pensava. "I giorni in cui – continua Pearl – l'IA saprà approssimarsi all'intelligenza umana è vicino, ma le sue capacità vanno giudicate su tre livelli di abilità cognitive: vedere (associazione), fare (intervento) e immaginare (controfattuali). L'IA oggi lavora solo al livello più basso, cioè, vedere".

Vedere, fare e immaginare: sono i tre pioli di una scala metaforica che costituisce l'asse narrativo di un bellissimo libro che Judea Pearl ha scritto insieme a Dana Mackenzie, per spiegare come mai solo oggi scienziati e statistici si capiscono quando discutono di dati e di come metterli tra loro in relazioni al fine di trovare spiegazioni causali dei fatti. Per oltre mezzo secolo, ricordano gli autori, ma forse anche di più, nel mondo accademico "il lessico causale era praticamente vietato". E questo ha avuto anche delle conseguenze tragiche. L'argomentazione attraverso cui si snoda la narrazione ruota intorno al disegno di una "scala della causalità", firmato dalla grafica e illustratrice Maayan Harel, molto efficace per dar senso agli argomenti, come altre illustrazioni, anche molto divertenti, presenti nel libro

Il gradino più basso, dove abitano gli statistici, si occupa semplicemente dell'osservazione, e consiste nella ricerca di regolarità nel mondo. In che modo le variabili sono collegate? In che modo cambia quello che so di Y, se osservo X? In altre parole, cosa mi dice un sintomo su una data malattia o un sondaggio sui trend elettorali? Pearl colloca, appunto, anche gli attuali algoritmi di machine learning e deep learning su cui si basa l'IA a questo primo stadio. L'esplosione della potenza di calcolo e i set di dati molto dettagliati accessibili hanno prodotto risultati sorprendenti e importanti, ma i meccanismi funzionano ancora "nello stesso modo in cui un lo statistico cerca di adattare una linea a un insieme di punti" e "i dati grezzi continuano a guidare i processi di adattamento". La rivoluzione causale è ciò che ha permesso ai ricercatori di salire sui gradini più alti della scala.

Nonostante gli sforzi eroici del genetista Sewall Wright (1889-1988) di applicare a diversi domini conoscitivi complessi la tecnica statistica della path analysis, da lui inventata nel 1918, per descrivere quantitativamente le dirette dipendenze tra insiemi di variabili – tecnica oggi di uso comune – la supremazia intellettuale esercitata dai biometristi inglesi allievi di Karl Pearson ha determinato una sorta di autocensura per l'uso di strumenti matematici per gestire le domande causali. Parlare di "causalità" era diventato come parlare di flogisto dopo la teoria cinetica del calore, per cui le statistiche solevano concentrare solo su come raccogliere e collegare tra loro i dati, non su come interpretarli e spiegarli. Ci sono statistici, per Judea Pearl, che ancora "trovano difficile capire che esistono conoscenze al di fuori delle statistiche e che i dati da soli non possono compensare la mancanza di conoscenza scientifica".

Al secondo piolo si passa dall'osservare al fare, all'intervenire. Ci si chiede cosa è successo o cosa sarebbe successo in base alle possibili modificazioni attive di una situazione. Cosa accadrebbe a Y se faccio X? Come posso fare in modo che accada Y? In altre parole: se prendo un analgesico se ne andrà il mal di testa? cosa accadrà rispetto a diverse variabili se vieto le sigarette? Secondo Pearl "molti scienziati sono rimasti traumatizzati apprendendo che nessuno dei metodi che hanno imparato con la statistica è sufficiente per articolare, per non dire rispondere, a una semplice domanda come 'Cosa succede se raddoppiamo il prezzo?'. Il libro spiega come e quando un modello da solo può rispondere alla domanda in assenza di esperimenti dal vivo.

Al gradino più alto della scala entra in gioco l'argomentazione "controfattuale", cioè l'immaginazione, la retrospezione e la comprensione. X è stato causato da Y? Se X non fosse accaduto come sarebbero oggi le cose? Cosa sarebbe successo se avessi agito diversamente? In altre parole: è stata l'aspirina a farmi passare il mal di testa?; se i terroristi non avessero abbattuto le torri gemelle ci sarebbero state le guerre degli Stati Uniti contro paesi islamici? Siamo al livello della scienza, ma anche dell'etica. Si tratta di usare la capacità di guardare indietro e immaginare cosa avrebbe potuto governare i nostri giudizi in termini di successo e fallimento, giusto e sbagliato, etc. Un tempo questo modo di argomentare era considerato distintivo dell'intelligenza umana, e negli ultimi decenni sono entrati in gioco strumenti di modellizzazione sempre più complessi. Tali strumenti sono stati applicati a vari problemi sociali e scientifici, tra cui l'efficacia delle procedure mediche, l'impatto dei cambiamenti climatici e l'utilità delle politiche sociali.

Gran parte del libro discute di perché e come gli statistici e gli scienziati rifiutavano la causalità come argomento scientifico. Si studiavano le correlazioni, ci si ripeteva il mantra "la correlazione non implica causalità" e si riteneva che

fosse illusorio cercare di andare oltre le correlazioni. Gli scienziati sperimentali avrebbero voluto discutere e pianificare le implicazioni causali della loro ricerca, ma gli statistici rifiutarono la maggior parte dei tentativi di prendere in considerazione le cause. Con l'invenzione degli studi controllati e randomizzati (RCT) gli statistici hanno creduto di poter dimostrare che la correlazione implica la causalità. Quindi gli RCT sono diventati sempre più importanti. L'uso degli RCT sdoganava il concetto di causalità, ma fino a un certo punto. Quando per esempio gli scienziati hanno notato che il fumo potrebbe causare il cancro del polmone, hanno dovuto aspettare che gli RCT osservassero che il fumo è associato al cancro. Questo alla fine ha spinto gli esperti a immaginare criteri utili per stabilire la causalità. Ma anche nelle circostanze ideali, quei criteri non erano abbastanza convincenti da produrre un consenso fra gli statistici. Alcuni statistici influentissimi come Ronald Fisher e Jacob Yerushalmi hanno usato il loro prestigio e la loro intelligenza che eludere l'argomento della causalità e mettere in discussione le correlazioni. Per cui le prese di posizione degli organi sanitari sul fumo e sul cancro sono state ritardate per anni dal disagio degli scienziati nel parlare di causalità.

Pearl emerge, attraverso le pagine del libro ma anche da interviste, come un uomo molto simpatico, pieno di curiosità e di gioia per il suo lavoro, e anche molto amato da studenti e colleghi viene da pensare. Il suo entusiasmo per la ricerca filosofica deve averlo aiutato non poco a fronteggiare la tragedia del figlio giornalista sequestrato e assassinato dai talebani in Pakistan. La sua reazione è stata molto istruttiva. Intanto ha creato una fondazione a nome di Daniel fondata sui principi di tolleranza e il rispetto per tutte le culture e sui suoi valori che in ultima istanza gemmano da una cultura scientifica o che coltiva l'oggettività e rifugge il conformismo. "Mio figlio è stato ucciso dall'odio per cui sono deciso a combattere l'odio" ha detto, non ha lesinato dure critiche a chi, soprattutto liberal e progressisti, accusano Israele di essere una causa del terrorismo e così razionalizzano vergognosamente il terrorismo fondamentalista.