

Senso e causa

Epistemologia dei sistemi complessi tra genomica computazionale e ontologia relazionale

Carlo Mancosu

Gennaio 2026

ABSTRACT

Il presente lavoro indaga la distinzione tra causa e senso come categorie epistemiche irriducibili, assumendo la genomica computazionale contemporanea come caso di studio privilegiato. Partendo dall'analisi delle pratiche inferenziali nei sistemi di *variant calling* basati su intelligenza artificiale, si mostra come la scienza moderna operi sempre più attraverso regimi di senso — configurazioni locali, ritrattabili, dipendenti da scopo — mentre continua a raccontarsi nel linguaggio della causalità forte.

La tesi centrale afferma che il rapporto diadico puro tra causa ed effetto non esiste nella struttura del reale, ma costituisce una proiezione semplificata su campi relazionali complessi. Il senso, diversamente dalla causa, non cerca l'origine ma valuta la collocazione; non opera per concatenazione ma per configurazione; non pretende irreversibilità ma ammette ritrattazione senza contraddizione logica.

Il passaggio dalla rappresentazione lineare del genoma alla struttura pangenomica a grafo viene letto come transizione da un'ontologia sostanziale a un'ontologia relazionale. In questa prospettiva, la variante genetica non è più un oggetto discreto che devia da un archetipo, ma un evento topologico — una biforcazione, un percorso alternativo in uno spazio di possibilità.

La critica si estende alla divulgazione scientifica e al marketing farmaceutico, che trasformano stime probabilistiche in enunciati deterministici, perpetuando un'eredità teologico-metafisica — l'idea platonica dell'armonia e del progetto — incompatibile con la natura dei sistemi complessi.

PROLOGO

La genomica come laboratorio epistemologico

«Il sapere scientifico non può essere fondato se non su se stesso, non può darsi alcuna ragione della propria esistenza e della propria necessità se non nella supposizione di un'etica.»

— Jacques Monod

La possibilità di sequenziare il genoma di una persona consente oggi di diagnosticare e curare malattie genetiche, sviluppare nuovi farmaci, comprendere la variazione biologica della specie. Strumenti come DeepVariant¹ — reti neurali profonde addestrate a distinguere varianti reali da artefatti di sequenziamento — rendono la lettura del DNA più accurata, economica e rapida che mai. Eppure, proprio nel cuore di questa efficienza tecnica, si annida una questione filosofica ineludibile: cosa significa sapere, in questo contesto?

Quando diciamo che un gene è «responsabile» di una malattia, non stiamo descrivendo una proprietà osservabile del DNA. Stiamo formulando un'inferenza statistica basata su correlazioni all'interno di popolazioni, filtrate attraverso modelli di ereditarietà che assumiamo validi, punteggi di deleterietà costruiti su conoscenze biologiche pregresse, soglie di rarità definite rispetto a database storicamente sbilanciati. La responsabilità non è scritta nel genoma: è il risultato di un sofisticato processo di costruzione epistemica.

Questo lavoro assume la genomica computazionale come caso di studio non per contribuire al dibattito tecnico sulle metodologie di sequenziamento, ma per illuminare una questione più generale: la differenza tra causa e senso come categorie epistemiche. La tesi che qui si sostiene è che la scienza contemporanea — in particolare quella che opera su sistemi complessi, multifattoriali, storicamente situati — ha già superato il paradigma causale classico nella pratica, pur continuando a parlarne il linguaggio nella teoria e nella comunicazione pubblica.

Il divario tra pratica e racconto genera conseguenze concrete. La divulgazione generalista e il marketing dell'industria farmaceutica e del *wellness* trasformano stime probabilistiche in verità assolute, approssimazioni efficaci in certezze ontologiche. Il test del DNA viene venduto come rivelazione di un destino inscritto nel codice genetico, quando in realtà produce decisioni praticabili sotto incertezza, mappe utili che non coincidono mai con il territorio.

¹DeepVariant: software sviluppato da Google Research che utilizza reti neurali convoluzionali per identificare varianti genetiche. Trasforma i dati di sequenziamento in immagini e applica tecniche di riconoscimento visivo per distinguere varianti reali da artefatti tecnici.

La struttura del lavoro procede in due movimenti. La prima parte analizza il passaggio dall'osservazione all'inferenza nella genomica moderna, mostrando come la circolarità delle conferme e il pregiudizio metodologico siano condizioni ineliminabili della conoscenza. La seconda parte sviluppa la distinzione tra causa e senso, proponendo il passaggio a un'ontologia relazionale in cui il senso non è un legame diadico tra eventi, ma una proprietà geometrica, topologica e reticolare dello spazio delle configurazioni.

PARTE PRIMA

Dalla correlazione alla causa

CAPITOLO 1

Osservazione e inferenza nella genomica moderna

La genomica moderna opera attraverso due livelli distinti che vengono sistematicamente confusi: l'osservazione strumentale e l'attribuzione di senso. Il primo livello — letture, allineamenti, chiamata di varianti², metriche di qualità — produce dati. Il secondo livello — patogenicità, responsabilità di un gene, narrazione clinica — produce interpretazioni. La confusione tra questi piani è la fonte di gran parte degli abusi epistemici che attraversano il campo.

Consideriamo cosa accade quando un sequenziatore «legge» il DNA. I frammenti di acido nucleico attraversano una macchina che traduce segnali biochimici o elettrici in sequenze di lettere. Software come DeepVariant identificano dove la sequenza di un individuo differisce da un genoma di riferimento. A questo livello, osserviamo una proprietà fisica: alla posizione X del cromosoma Y, l'individuo presenta una base diversa da quella attesa. Questa è una misurazione, non un'interpretazione.

Il passaggio all'interpretazione richiede invece molteplici strati di inferenza. Dire che quella variazione è «responsabile» di una malattia comporta assumere che i punteggi predittivi (come CADD³) catturino correttamente l'impatto biologico; che le varianti rare nella popolazione generale siano più probabilmente patogene; che i modelli mendeliani⁴ descrivano adeguatamente i pattern di ereditarietà; che la correlazione tra variante e fenotipo⁵ clinico sia causale e non spuria. Ogni assunzione introduce incertezza. Il risultato finale non è una verità, ma una stima ad alta fiducia condizionata a un sistema di premesse.

Il processo di identificazione di una variante patogena procede tipicamente attraverso fasi successive di filtraggio. Si parte da un insieme di differenze rispetto al riferimento — centinaia di migliaia in ogni individuo — e si applica una serie di criteri: frequenza di popolazione, impatto predetto sulla proteina, compatibilità con il fenotipo clinico, coerenza con i modelli di ereditarietà. Il risultato è una «lista di candidati», solitamente compresa tra dieci e cinquanta varianti, che richiede ulteriore validazione.

² *Variant calling*: processo computazionale di identificazione delle differenze tra il genoma sequenziato di un individuo e un genoma di riferimento. Il termine «chiamata» indica che il software «chiama» (dichiara) una variante quando i segnali superano una soglia di confidenza statistica.

³ CADD (*Combined Annotation Dependent Depletion*): algoritmo che integra diverse annotazioni genomiche per calcolare un punteggio numerico che predice quanto una variante possa essere dannosa per la funzione del gene. Punteggi più alti indicano maggiore probabilità di effetti deleteri.

⁴ Modelli mendeliani: schemi di ereditarietà derivati dalle leggi di Mendel, che descrivono come i caratteri genetici si trasmettono dai genitori ai figli secondo pattern prevedibili (dominante, recessivo, legato al cromosoma X, ecc.).

⁵ Fenotipo: l'insieme delle caratteristiche osservabili di un organismo (aspetto fisico, sintomi clinici, parametri biochimici), risultanti dall'interazione tra il genotipo (patrimonio genetico) e l'ambiente.

Questa validazione, tuttavia, non può mai essere completamente indipendente dalle assunzioni che hanno guidato la selezione iniziale. I gold standard⁶ — i dataset di riferimento usati per addestrare e valutare i modelli — sono essi stessi costruiti storici, prodotti dalle stesse tecnologie e dagli stessi paradigmi che pretendono di validare. Il consorzio Genome in a Bottle fornisce sette campioni «standard» sequenziati con decine di tecnologie e curati manualmente per anni. Ma questi campioni non sono la verità: sono la migliore approssimazione disponibile, soggetta a correzioni quando nuove tecnologie rivelano errori precedentemente invisibili.

L'intelligenza artificiale, in questo contesto, non risolve il problema dell'inferenza — lo amplifica. DeepVariant non viene programmato con regole biologiche; viene addestrato mostrandogli milioni di esempi di immagini di sequenziamento accoppiate a etichette fornite dai gold standard. Attraverso questa esposizione massiccia, il modello «impara» a distinguere varianti reali da errori di sequenziamento, approssimando una funzione statistica complessa che nessun essere umano ha codificato esplicitamente. È induzione pura: il modello generalizza dai dati, inducendo capacità predittive applicabili a campioni mai visti prima.

L'aspetto più rivelatore è che questi modelli sviluppano capacità che vanno oltre il loro addestramento esplicito. DeepVariant, addestrato solo per rilevare piccole varianti ereditarie, ha imparato autonomamente a identificare segnali correlati a varianti mosaico⁷ (non ereditarie) e elementi LINE⁸. Questa «conoscenza nascosta» emerge perché il modello è un approssimatore universale che rileva pattern nei dati che noi stessi non sappiamo ancora descrivere come regole biologiche.

Qui si manifesta la natura profondamente induttiva del sapere genomico contemporaneo. Non partiamo da leggi universali per dedurre il comportamento di un gene; osserviamo miliardi di dati per estrarre pattern di probabilità. I pattern vengono cercati in modo che concordino con ciò che, *rebus sic stantibus*, crediamo di sapere sulla biologia e sull'epidemiologia. L'associazione tra genetica e malattia è sempre un'inferenza basata su assunzioni metodologiche che precedono l'osservazione.

La domanda diventa allora: quando diciamo che un gene è «responsabile» di una malattia, cosa stiamo affermando? Non una proprietà osservabile della materia, ma una validità locale sotto vincoli espliciti. Dato questo modello, questi dati, questi assunti e questo contesto d'uso, trattare quella variante come patogena migliora l'esito atteso. È un fatto decisionale, non ontologico.

⁶Gold standard (o *truth set*): dataset di riferimento considerato «vero» per addestrare e validare i modelli computazionali. In genomica, i principali gold standard sono forniti dal consorzio Genome in a Bottle (GIAB) del NIST (National Institute of Standards and Technology), che ha caratterizzato in modo estremamente dettagliato il genoma di sette individui.

⁷Mosaicismo: condizione in cui un individuo presenta cellule con genotipi diversi, originati da mutazioni avvenute dopo il concepimento. A differenza delle varianti ereditarie (presenti in tutte le cellule), le varianti mosaico sono presenti solo in una porzione delle cellule.

⁸LINE (*Long Interspersed Nuclear Elements*): sequenze di DNA ripetute lunghe circa 6.000 basi che costituiscono circa il 17% del genoma umano. Sono «elementi mobili» capaci di spostarsi e replicarsi in diverse posizioni del genoma.

CAPITOLO 2

La circolarità delle conferme

Il problema della circolarità delle conferme è un limite intrinseco e riconosciuto nella genomica computazionale. Quando sia l'addestramento che la validazione dei modelli dipendono da un numero ristretto di «verità di base», il rischio è che il sistema impari non la biologia universale, ma le caratteristiche specifiche di quei pochi campioni. Se usiamo gli stessi standard per misurare il successo, stiamo chiedendo al modello di confermare un paradigma che gli abbiamo già fornito.

Questo fenomeno è stato descritto come «effetto lampione» (*streetlamp effect*): la tendenza a cercare — e trovare — solo ciò che rientra nel campo illuminato dalla mappa esistente. Per decenni, la genomica si è basata su un genoma di riferimento lineare (GRCh38⁹) che fungeva da coordinate assolute per localizzare ogni variante. Ma questo riferimento non è il territorio: è una rappresentazione parziale, costruita principalmente da campioni di origine europea, che ignora fino al dieci per cento del genoma umano presente in altre popolazioni.

La circolarità opera a più livelli. I modelli di *variant calling* vengono addestrati sui gold standard del NIST; i gold standard vengono costruiti usando tecnologie che incorporano le stesse assunzioni dei modelli; la validazione avviene confrontando i risultati con gli stessi gold standard usati per l'addestramento. È un sistema chiuso che conferma se stesso, anche quando i suoi presupposti sono incompleti o errati.

Le strategie adottate per mitigare la circolarità sono istruttive, perché rivelano la consapevolezza metodologica della comunità scientifica. L'evidenza ortogonale consiste nel confrontare tecnologie con errori sistematici diversi: se Illumina, PacBio e Oxford Nanopore¹⁰ — tre piattaforme con chimiche e bias differenti — concordano su una variante, si assume che non sia un artefatto del paradigma di sequenziamento. È un criterio pragmatico, non una dimostrazione di verità.

La «silver truth» (verità d'argento) rappresenta un'ammissione ancora più esplicita. Quando mancano standard assoluti, si usa un metodo per generare etichette per un altro metodo, accettando una qualità inferiore pur di espandere i dati oltre il paradigma originale. DeepVariant addestrato sul sequenziamento genomico viene usato per etichettare dati di sequenziamento dell'RNA, producendo

⁹GRCh38 (*Genome Reference Consortium Human Build 38*): la versione più recente e completa del genoma umano di riferimento, pubblicata nel 2013 e periodicamente aggiornata. Consiste in una sequenza lineare di circa 3 miliardi di basi che funge da sistema di coordinate per localizzare geni e varianti.

¹⁰Illumina, PacBio, Oxford Nanopore: le tre principali tecnologie di sequenziamento del DNA, ciascuna con caratteristiche distinte. Illumina produce letture brevi (150-300 basi) ad alta precisione; PacBio e Oxford Nanopore producono letture lunghe (migliaia di basi) che facilitano l'assemblaggio ma con tassi di errore inizialmente più alti. Le diverse tecnologie hanno errori sistematici diversi, rendendo il loro accordo una forma di validazione incrociata.

oltre ottanta milioni di esempi che la verità manuale non potrebbe mai generare. È *bootstrapping*: si tira su il sistema tirandosi per i capelli.

Il passaggio ai pangenomi¹¹ rappresenta il tentativo più ambizioso di rompere la circolarità. Invece di un riferimento lineare, si costruisce un grafo in cui i nodi contengono segmenti di DNA e i percorsi descrivono le variazioni individuali. Il pangenoma attuale combina quarantasette genomi diversi, aggiungendo centodiciannove milioni di basi e rivelando alleli¹² in loci¹³ complessi che la mappa lineare rendeva invisibili. Ma anche il pangenoma è una costruzione: le scelte su come assemblare il grafo, quali individui includere, come rappresentare le divergenze, riflettono visioni diverse della diversità umana.

Il punto cruciale è che affrontare e cercare di prevenire le criticità note non significa esserne immuni, né immunizzarsi verso quelle non ancora riscontrate o immaginate. I modelli attuali incorporano assunzioni che potrebbero rivelarsi errate quando nuove tecnologie o nuove popolazioni entreranno nel campo di osservazione. Le metriche di qualità basate sui k-mer¹⁴ sottostimano drasticamente i tassi di errore nelle regioni ripetitive, perdendo oltre il novanta per cento degli errori residui. I centromeri¹⁵ e le varianti strutturali complesse rimangono esclusi dai benchmark perché la scienza ammette di non avere ancora gli standard per rappresentarli.

Questo non invalida il sapere genomico; lo situa. La conoscenza prodotta è valida entro un regime di senso definito da: lo stato delle conoscenze disponibili, gli strumenti tecnologici impiegati, le finalità operative dichiarate. Cambiane una sola, e il senso può dissolversi senza che nulla «sia stato falso». Una decisione passata non diventa sbagliata perché superata; diventa non più sensata nel nuovo regime.

La formulazione più onesta sarebbe dunque: questa affermazione è valida rispetto a una certa rappresentazione, un certo insieme di assunzioni, un certo dominio di applicazione, un certo costo dell'errore. Ma questa formulazione non vende, non tranquillizza, non è insegnabile come dogma.

¹¹Pangenoma: rappresentazione del genoma di una specie che include non solo una sequenza di riferimento, ma l'insieme delle variazioni presenti in una popolazione. A differenza del genoma lineare tradizionale, il pangenoma è strutturato come un grafo che cattura la diversità genetica di molteplici individui.

¹²Allele: ciascuna delle forme alternative che un gene può assumere in una determinata posizione (locus) del cromosoma. Ad esempio, il gene che determina il colore degli occhi può esistere in diverse varianti (alleli) che producono occhi marroni, blu o verdi.

¹³Locus (plurale: loci): posizione specifica di un gene o di una sequenza di DNA su un cromosoma. Ogni gene ha un proprio locus, identificato da coordinate cromosomiche precise.

¹⁴K-mer: sequenza di k nucleotidi consecutivi. L'analisi dei k-mer è una tecnica computazionale che scompone le sequenze di DNA in frammenti sovrapposti di lunghezza fissa per stimare la qualità dell'assemblaggio genomico confrontando la frequenza attesa con quella osservata.

¹⁵Centromero: regione centrale del cromosoma, essenziale per la corretta segregazione durante la divisione cellulare. I centromeri sono costituiti da sequenze altamente ripetitive che risultano particolarmente difficili da sequenziare e assemblare con le tecnologie tradizionali.

CAPITOLO 3

Il pregiudizio come condizione della conoscenza

«I pregiudizi del singolo, molto più che i suoi giudizi, costituiscono la realtà storica del suo essere.»

– Hans-Georg Gadamer

Non solo la mappa non è il territorio; la mappa è incompleta e risente di un sistema di categorizzazione stabilito a priori, prima di avere una visione complessiva del genoma stesso. Questa non è una debolezza da correggere, ma una condizione strutturale di ogni impresa conoscitiva. Come insegna l'ermeneutica gadameriana, non possiamo prescindere dal pregiudizio: possiamo tenerlo presente, ma il nostro sguardo e le nostre mappe non saranno mai neutri.

Quando si disegna una mappa, cosa rappresentare e come rappresentarlo dipende dagli obiettivi che riteniamo di avere e dall'immagine a priori che abbiamo del paesaggio da mappare. La genomica ha operato per decenni con un genoma di riferimento lineare perché l'obiettivo era avere un sistema di coordinate univoco – una semplificazione necessaria data la complessità del materiale. Ma questa scelta ha reso invisibili le regioni che non si conformavano alla linearità: inserzioni, delezioni, riarrangiamenti strutturali, variazioni che esistevano nel territorio ma non avevano posto sulla mappa.

Le soglie categoriali che usiamo sono convenzioni tecniche, non distinzioni biologiche. La differenza tra una «piccola variante» (meno di cinquanta basi) e una «variante strutturale» (cinquanta basi o più) riflette i limiti dei vecchi algoritmi, non una discontinuità nella natura. Le regioni «difficili» – centromeri, eterocromatina, loci immunitari complessi – sono state sistematicamente escluse dai benchmark perché non rientravano nei modelli disponibili. Se non sapevamo come mapparli, li definivamo non analizzabile.

Il problema è, in fondo, teologico. L'idea platonico-aristotelica poi rilanciata da Plotino, Sant'Agostino e la scolastica – l'immagine di un grande architetto e dell'armonia della creazione – continua a informare le nostre aspettative cognitive anche quando la teologia esplicita è stata abbandonata. Che si chiami Dio o Natura non modifica il concetto: ci aspettiamo un ordine, un progetto, una corrispondenza tra la struttura del reale e la struttura del pensiero.

Il genoma lineare ha funzionato come archetipo platonico dell'umano: un modello ideale da cui ogni individuo devia in modo misurabile. La variante è stata pensata come errore o differenza rispetto a una norma, non come percorso alternativo ugualmente legittimo. Questo schema concettuale precede l'osservazione e la condiziona. Non vediamo varianti; vediamo deviazioni da un'attesa.

Il passaggio al pangenoma a grafo rappresenta, in questo senso, una svolta epistemologica prima ancora che tecnologica. Non c'è più un riferimento privilegiato; ogni individuo contribuisce con sequenze che aggiungono nodi e percorsi al grafo. L'identità genetica non è più deviazione da un centro, ma posizione in una rete. La variante diventa caratteristica topologica — un punto in cui il cammino si biforca per poi ricongiungersi — e non più oggetto discreto che infrange un ordine.

Derrida direbbe che le nostre mappe sono sempre già inserite in un testo — un sistema di differenze e rimandi che precede ogni singola osservazione. Non usciamo mai dal testo delle nostre categorie. Possiamo spostare i confini, aggiungere distinzioni, raffinare le rappresentazioni, ma ogni nuova mappa eredita il linguaggio e i presupposti di quelle precedenti.

Questo non condanna alla relatività assoluta. I vincoli fisici, computazionali e pragmatici restringono lo spazio delle mappe possibili. Non tutte le rappresentazioni sono ugualmente utili; alcune permettono previsioni migliori, decisioni più efficaci, comprensioni più profonde. Ma l'utilità non va confusa con la verità. Una mappa può essere estremamente utile — e la genomica lo è — senza essere vera nel senso forte di corrispondenza con un ordine indipendente dall'osservatore.

Il pregiudizio, dunque, non è un ostacolo da rimuovere ma una condizione da esplicitare. La scienza matura non è quella che pretende di eliminare le assunzioni, ma quella che le dichiara, le mette alla prova, le rivede quando necessario. Il problema nasce quando il pregiudizio viene naturalizzato — quando una convenzione tecnica viene scambiata per struttura del reale, quando un criterio operativo viene ontologizzato in verità.

PARTE SECONDA

Dal modello causale al regime di senso

CAPITOLO 4

L'illusione della causa diadica

Il modello causale classico pretende tre cose insieme: asimmetria (A produce B, non viceversa), necessità condizionata (dato un certo modello, se A avviene allora B segue), irreversibilità semantica (una volta che B è accaduto, A viene qualificato retroattivamente come «la causa»). Questa struttura funziona in casi limite — sistemi isolabili, variabili ben definite, relazioni lineari — ma collassa sistematicamente di fronte alla complessità.

In biologia, e ancor più in genomica, la causa forte è rara. Funziona quasi solo nei casi mendeliani semplici, dove un singolo gene con penetranza¹⁶ completa determina un fenotipo discreto. Ma appena entrano in gioco contesto, interazioni, temporalità, sviluppo, la catena causale si dissolve in una rete di influenze reciproche che nessun modello diadico può catturare.

Il problema non è che non riusciamo ancora a identificare tutte le cause; è che il rapporto diadico puro $A \rightarrow B$ non esiste nella struttura del reale. Ogni «causa» che identifichiamo è già una proiezione semplificata su un campo relazionale. Non esistono micro-causalità che sommandosi producono macro-effetti; esiste un tessuto di regolarità locali che trattiamo *come se* fossero cause perché ci serve operativamente, ma il «come se» non va dimenticato.

La confusione tra predittività e causalità è la fonte di gran parte degli abusi epistemici. La predittività è la capacità di anticipare un esito con errore controllato; la causalità pretende una relazione necessitante tra eventi. Nel regime contemporaneo — AI applicata alla genomica — quasi tutto è predittivo, pochissimo è causale in senso forte, ma il linguaggio continua a usare categorie causali.

Quando diciamo che una variante «causa» una malattia, stiamo in realtà affermando che: trattare questa configurazione come rilevante migliora la predizione dell'esito clinico sotto questi vincoli. Non è una spiegazione del meccanismo; è una stima di associazione che funziona finché funzionano le condizioni che l'hanno generata. Se domani cambia la popolazione, cambia la tecnologia, cambia il costo dell'errore — la «causa» può dissolversi senza che il mondo si contraddica.

I modelli di intelligenza artificiale come DeepVariant non imparano cause; imparano funzioni di verosimiglianza condizionate. Il modello non dice «questa variante produce quella malattia»; dice «dato questo insieme di segnali, la probabilità che questa configurazione sia una variante reale è X%». È una stima, non una spiegazione. E la stima dipende interamente dal regime di addestramento — i dati visti, le etichette assegnate, la funzione di perdita ottimizzata.

¹⁶Penetranza: in genetica, la proporzione di individui portatori di una specifica variante genetica che manifestano effettivamente il fenotipo associato. Una penetranza del 100% (completa) significa che tutti i portatori della variante presentano il carattere; penetranze inferiori indicano che altri fattori (genetici o ambientali) modulano l'espressione.

Perché allora continuiamo a preferire la causa, anche quando non regge? Perché la causa rassicura, chiude il discorso, consente narrazione lineare. Il marketing farmaceutico ha bisogno di certezze da vendere; la divulgazione scientifica ha bisogno di storie da raccontare; il sistema educativo ha bisogno di contenuti da trasmettere. Il senso — con la sua località, la sua ritrattabilità, la sua dipendenza dal contesto — è molto meno appetibile.

Culturalmente, siamo educati a cercare fondamenti, non criteri di orientamento. Il residuo teologico di cui parlavamo — l'aspettativa di un ordine, di un progetto, di una corrispondenza — si manifesta nella preferenza per spiegazioni che «chiudono» piuttosto che per valutazioni che «orientano». Ma questa preferenza non è epistemicamente innocua: produce aspettative irrealistiche, decisioni mal informate, delusioni quando la complessità riemerge.

La genomica dimostra quotidianamente che le cose stanno diversamente. I flussi di lavoro diagnostici producono liste di candidati, non verdetti definitivi. La validazione clinica richiede anni di follow-up e spesso non arriva mai. Le varianti di significato incerto (VUS¹⁷) — quelle per cui non sappiamo se sono patogene o benigne — costituiscono la maggioranza dei risultati. La pratica scientifica vive già nel regime del senso; è solo il racconto che resta ancorato al mito della causa.

¹⁷VUS (*Variant of Uncertain Significance*): variante genetica per cui non esistono sufficienti evidenze scientifiche per classificarla come patogena (causa di malattia) o benigna (innocua). Le VUS rappresentano una sfida diagnostica significativa perché non permettono di trarre conclusioni cliniche definitive.

CAPITOLO 5

Il senso come problema geometrico

Se il rapporto diadico puro non esiste, cosa resta? Resta il senso — ma il senso va compreso correttamente, non come versione degradata della causa, bensì come categoria epistemica autonoma con struttura propria.

Il senso muove nella direzione opposta rispetto alla ricerca causale. La causa cerca l'origine, risale una catena, identifica un punto privilegiato. Il senso parte da una situazione data — un esito, un contesto, una decisione da prendere — e valuta se una certa lettura tiene. Non risale; si dispone nello spazio delle possibilità. Non risponde alla domanda «da dove viene?» ma a «come si colloca questo, rispetto a ciò che sappiamo e vogliamo fare?».

Questa è già una direzione opposta: non genealogica, ma posizionale. Il senso non cerca antenati; cerca coordinate.

La seconda caratteristica fondamentale è che il senso non è diadico. La causalità vive nella coppia causa/effetto, prima/dopo, input/output. Il senso no. Il senso non esiste tra due termini, ma emerge da un insieme di relazioni, un contesto articolato, una configurazione. Non può essere ridotto a una coppia, non può essere espresso come legge, non può essere stabilizzato come nesso necessario.

Per questo il senso è intrinsecamente geometrico. Non è una freccia; è una collocazione. Una decisione è sensata se sta bene in un certo spazio di vincoli, se è coerente con le distanze, le prossimità, le curvature di quello spazio. Non «segue» qualcosa; si situa. Questo spiega perché il senso può cambiare senza che il mondo cambi: basta che cambi la metrica dello spazio — nuove conoscenze, nuovi obiettivi, nuovi costi dell'errore.

Il parallelo con la geometria riemanniana è illuminante. In una varietà riemanniana, la metrica non è data a priori ma dipende dalla posizione; le distanze si misurano lungo geodetiche che curvano in risposta alla struttura locale dello spazio. Analogamente, il senso di un'affermazione non è dato a priori ma dipende dalla configurazione epistemica; la sua coerenza si valuta rispetto a traiettorie che curvano in risposta agli obiettivi, ai vincoli, alle conoscenze disponibili. Non si misura una distanza da un punto esterno chiamato «verità», ma una tenuta interna allo spazio delle configurazioni possibili.

Il senso è anche topologico. La topologia studia le proprietà che si conservano sotto deformazioni continue — connettività, numero di buchi, orientabilità — indipendentemente dalle coordinate specifiche. Analogamente, il senso non dipende da coordinate assolute ma dalla connettività: una lettura resta sensata se le relazioni fondamentali tengono, anche quando cambiano i dettagli locali.

Questo lo rende capace di attraversare rappresentazioni diverse — lineare, grafo, campo — senza pretendere un fondamento unico. La causalità, invece, è fragile alle deformazioni: basta rompere un anello della catena e collassa. Se A causa B e B causa C, eliminare B distrugge la catena. Ma se il senso di una configurazione dipende dalla sua posizione in una rete, la rimozione di un nodo richiede solo una riconfigurazione delle connessioni.

Infine, il senso è reticolare. La causa si cerca, si individua, si punta. Il senso non sta in un punto; emerge da una rete di coerenze parziali. Non puoi dire «il senso è qui»; puoi solo dire «questa configurazione regge meglio di altre». È distribuito, non localizzabile. È esattamente ciò che fanno i modelli probabilistici, i grafi, i sistemi di decisione sotto incertezza.

Il senso, dunque, è un problema che tiene conto della posizione delle componenti e non di un loro presunto legame di causa-effetto diretto. Una componente può non causare direttamente nulla, ma essere decisiva perché sposta l'equilibrio della configurazione. Nei sistemi complessi — biologici, economici, sociali — questo è la norma, non l'eccezione.

La conseguenza è forte: se il senso è posizione e non legame causale, allora non puoi «scoprirlo» come una causa, non puoi «dimostrarlo» una volta per tutte, non puoi «attribuirlo» a un singolo elemento. Puoi solo argomentarlo, mostrarne la stabilità locale, difenderlo rispetto a uno scopo dichiarato. Ed è per questo che il senso è ritrattabile senza contraddizione, cambia quando cambia lo spazio, ma non per questo è arbitrario.

CAPITOLO 6

Verso un'ontologia relazionale

L'approccio topologico suggerisce il passaggio a un'ontologia relazionale. Mentre l'ontologia sostanziale chiede «cosa esiste?», l'ontologia relazionale chiede «quali relazioni costituiscono ciò che esiste?». Nel primo caso, le relazioni sono secondarie — proprietà accidentali di sostanze preesistenti. Nel secondo, le relazioni sono primarie — ciò che esiste è costituito dalle sue connessioni.

La transizione dalla genomica lineare a quella pangenomica illustra concretamente questo passaggio. Nel paradigma lineare, una variante era un oggetto discreto definito per differenza rispetto a un riferimento assoluto. Nel paradigma a grafo, la variante è un percorso alternativo in una rete di possibilità — non un'entità che devia, ma un cammino che attraversa.

L'identità genetica stessa si trasforma. Non è più la somma delle deviazioni da un archetipo; è la specificazione di un percorso attraverso il grafo. Due individui non differiscono per le varianti che «hanno»; differiscono per i cammini che percorrono. La sostanza — il DNA come cosa — cede il passo alla relazione — il DNA come posizione in uno spazio di variazione.

I concetti di «snarls» e «bubbles»¹⁸ nella rappresentazione pangenomica sono istruttivi. Uno snarl è un sottografo annidato corrispondente a una regione di divergenza; una bubble è una biforcazione che poi si ricongiunge. La variante non è più un punto che «è o non è»; è una caratteristica topologica del campo — una struttura relazionale che ammette attraversamenti diversi.

Questo cambia radicalmente cosa significa «diagnosticare». Nel paradigma sostanziale, diagnosticare è identificare un oggetto (la variante patogena) e attribuirgli una proprietà (la responsabilità causale). Nel paradigma relazionale, diagnosticare è individuare una configurazione (un pattern di attraversamenti) e valutarne la coerenza (il senso rispetto al fenotipo clinico). Non si cerca più il colpevole; si mappa il paesaggio.

L'intelligenza artificiale si inserisce naturalmente in questa ontologia. I modelli come DeepVariant non identificano oggetti; stimano probabilità condizionate a configurazioni. Il «canale AF»¹⁹ (frequenza allelica) non è un dato sulla variante, ma un vincolo relazionale — come quella variante si posiziona rispetto alla popolazione. Il modello impara relazioni tra segnali, non proprietà di sostanze.

¹⁸Snarls e bubbles: terminologia tecnica della rappresentazione pangenomica a grafo. Una *bubble* (bolla) è una struttura in cui due percorsi divergono da un nodo e poi si ricongiungono, rappresentando una regione di variazione. Uno *snarl* (groviglio) è una generalizzazione che può contenere strutture annidate più complesse.

¹⁹AF (*Allele Frequency*): la frequenza con cui un particolare allele compare in una popolazione. Le varianti molto rare nella popolazione generale (bassa AF) sono considerate più probabilmente patologiche per malattie rare, mentre varianti comuni (alta AF) sono generalmente benigne.

L'ontologia relazionale ha conseguenze per il concetto stesso di verità. Se ciò che esiste è costituito da relazioni, e le relazioni dipendono dal sistema di riferimento, allora non c'è una verità assoluta «là fuori» da scoprire. C'è un campo di configurazioni possibili, alcune più stabili di altre, alcune più utili di altre, alcune più coerenti di altre — ma nessuna «vera» in senso forte.

Questo non è relativismo. I vincoli fisici, computazionali e pragmatici restringono lo spazio delle configurazioni ammissibili. Non tutte le letture sono equivalenti; alcune reggono meglio sotto stress, generalizzano meglio a nuovi domini, producono decisioni migliori. Ma «migliore» va sempre specificato: migliore per quale scopo, in quale contesto, secondo quali criteri.

La verità, in questo quadro, viene sostituita dalla legittimazione. Ciò che conta non è se un'affermazione è vera, ma se è legittimamente producibile e usabile in un certo contesto. È un cambio di ontologia del sapere scientifico: dalla verità come corrispondenza alla validità come funzione di un sistema socio-tecnico.

I «regimi di senso» che operano nella genomica contemporanea illustrano questa pluralità di legittimazioni. Nel regime della silver truth, un'etichetta è legittima se è stata generata da un modello affidabile per addestrarne un altro. Nel regime topologico, un allele è legittimo se costituisce un percorso coerente nel grafo. Nel regime probabilistico, una chiamata è legittima se supera una soglia di verosimiglianza calibrata empiricamente. Nel regime dell'accuratezza situazionale, una variante è «vera» se identificarla in tempo permette un intervento salvavita.

Nessuno di questi regimi coincide con la verità assoluta; tutti producono sapere operativamente valido. La scienza contemporanea lavora già così, anche quando non lo teorizza esplicitamente. Il compito della filosofia è rendere questa pratica riflessiva — esplicitare i regimi, mostrarne i confini, prevenire le ontologizzazioni abusive.

EPILOGO

Scienza senza fondamento

«Se devo stabilire un criterio, non posso chiamarla verità.»

Nel momento stesso in cui stabilisci un criterio, hai già rinunciato alla verità in senso forte. Quello che stai facendo non è fondare il vero, ma rendere operabile una distinzione. La verità pretende indipendenza dal punto di vista, stabilità rispetto al contesto, validità che non dipende dallo scopo. Un criterio, invece, è sempre situato, strumentale, revocabile.

La genomica contemporanea non cerca la verità. Costruisce mappe legittime per agire, sapendo che non coincidono col territorio, che sono storiche, che sono revocabili, che sono valutabili solo rispetto a un fine. E la domanda corretta non è più «è vero?» ma «in quale regime di senso questa affermazione è legittima, e a quale costo?».

Questo non indebolisce la scienza; la rende finalmente adulta. Una scienza che dichiara i propri presupposti è più onesta di una che li nasconde. Una scienza che ammette la revisibilità è più robusta di una che pretende l'eternità. Una scienza che esplicita i costi dell'errore è più responsabile di una che promette certezze.

Il senso non è un vincolo causale in senso stretto; è verifica di predittività statistica entro funzioni obiettivo date rispetto a utilità dichiarate. Non spiega il mondo; misura quanto bene riusciamo a prevederne alcuni aspetti rispetto a scopi che abbiamo scelto. Il senso è una forma di senso diversa dalla causa: locale, non assoluto, e soprattutto privo di quel rapporto di irreversibilità con l'effetto che la causa pretende.

La causa pretende necessità; il senso valuta riuscita. La causa guarda all'origine; il senso guarda all'uso. La causa è irreversibile; il senso è sempre ritrattabile. La causa è una freccia; il senso è una geometria. La causa è locale e puntuale; il senso è distribuito e reticolare.

Dire «è stata sensata rispetto a uno scopo, *rebus sic stantibus* ciò che pensiamo di sapere» significa che il predicato non appartiene alla cosa, ma all'assetto cognitivo-operativo in cui la cosa è stata trattata. Ogni giudizio di senso incorpora almeno tre dipendenze: lo stato delle conoscenze (ciò che in quel momento riteniamo affidabile), gli strumenti disponibili (tecnologie, modelli, rappresentazioni), le

finalità dichiarate (diagnosi, riduzione del rischio, equità). Cambiane una sola, e il senso può dissolversi senza che nulla «sia stato falso».

Il problema resta la comunicazione. La divulgazione generalista e il marketing continuano a scambiare stime per fatti, approssimazioni per verità, criteri per fondamenti. Il «gene responsabile» diventa slogan; il test del DNA diventa oracolo; la predizione diventa destino. L'abuso non è solo epistemico — è etico. Crea aspettative irrealistiche, alimenta ansie infondate, distorce decisioni che hanno conseguenze reali sulla vita delle persone.

La risposta non può essere il ritiro della scienza dalla comunicazione pubblica. Deve essere una comunicazione diversa — una che spieghi cosa significa davvero «ad alta fiducia», che mostri le barre di errore insieme ai risultati, che dichiari i regimi di senso invece di nasconderli. Una comunicazione che tratti il pubblico come capace di gestire l'incertezza, invece di proteggerlo con certezze fittizie.

La scienza senza fondamento non è scienza debole. È scienza consapevole dei propri limiti, capace di operare efficacemente nell'incertezza, onesta sulla natura delle proprie affermazioni. È, paradossalmente, la scienza più forte — perché non crolla quando il fondamento presunto si rivela illusorio.

La genomica ce lo insegna ogni giorno: possiamo prendere decisioni vitali — letteralmente, decisioni che salvano vite — senza pretendere di possedere la verità. Possiamo orientarci nello spazio delle possibilità senza una mappa definitiva. Possiamo agire con senno anche quando la causa ci sfugge.

Forse è questo il compito della filosofia oggi: non fondare la scienza, ma aiutarla a fare a meno del fondamento.

NOTA BIBLIOGRAFICA

Il presente lavoro si situa all'intersezione tra filosofia della scienza, epistemologia e studi sulla tecnologia. Le fonti primarie che hanno informato l'analisi della genomica computazionale provengono dalla documentazione tecnica di Google Research sui sistemi DeepVariant, DeepTrio e sulle metodologie pangenomiche sviluppate dal Human Pangenome Reference Consortium (HPRC). Questi materiali forniscono una descrizione dettagliata delle pratiche inferenziali, dei limiti metodologici e delle strategie di validazione impiegate nella ricerca genomica contemporanea.

Sul versante filosofico, il debito principale è verso l'ermeneutica di Hans-Georg Gadamer, in particolare la teoria del pregiudizio come condizione della comprensione sviluppata in *Wahrheit und Methode* (1960). La nozione derridiana di «testo» — l'idea che ogni interpretazione sia sempre già inserita in un sistema di differenze che la precede — informa la critica alla pretesa di neutralità delle rappresentazioni scientifiche.

La distinzione tra causa e senso qui proposta attinge alla critica humane della causalità, ma se ne distanzia per il rifiuto di ridurre il senso a mera abitudine o convenzione. Il senso, come qui inteso, ha una struttura propria — geometrica, topologica, reticolare — che non coincide né con la necessità della causa né con l'arbitrarietà della convenzione.

Per la critica al riduzionismo causale in biologia, sono stati rilevanti i lavori di Ernst Mayr sulla distinzione tra cause prossime e ultime, e la più recente letteratura sulla biologia dei sistemi, che sottolinea la natura emergente e relazionale dei fenomeni biologici. La metafora del «campo» come alternativa alla «catena» causale trova antecedenti nella teoria dei campi di Kurt Lewin e, più recentemente, nelle applicazioni della topologia algebrica alla biologia computazionale.

La critica alla divulgazione scientifica e al marketing farmaceutico si inserisce nel più ampio dibattito sulla comunicazione del rischio e dell'incertezza, con riferimento ai lavori di Gerd Gigerenzen sulla rappresentazione delle probabilità e di Nassim Taleb sui limiti della previsione nei sistemi complessi.

Riferimenti essenziali

Derrida, J. (1967). *De la grammatologie*. Minuit.

Gadamer, H.-G. (1960). *Wahrheit und Methode*. Mohr.

Gigerenzen, G. (2002). *Calculated Risks: How to Know When Numbers Deceive You*. Simon & Schuster.

Human Pangenome Reference Consortium (2023). A draft human pangenome reference. *Nature*, 617, 312–324.

Lewin, K. (1936). *Principles of Topological Psychology*. McGraw-Hill.

Mayr, E. (1961). Cause and effect in biology. *Science*, 134, 1501–1506.

Monod, J. (1970). *Le hasard et la nécessité*. Seuil.

Poplin, R. et al. (2018). A universal SNP and small-indel variant caller using deep neural networks. *Nature Biotechnology*, 36, 983–987.

Taleb, N. N. (2007). *The Black Swan: The Impact of the Highly Improbable*. Random House.