

Enrica Amaturo, Biagio Aragona,
Suania Acampa e Francesco Amato

Tecnologie emergenti e ricerca sociale

Prospettive teoriche e pratiche metodologiche



**Percorsi
di ricerca**

FrancoAngeli

OPEN  ACCESS

Percorsi di ricerca

COLLANA DIRETTA DA **RENATO GRIMALDI**

Comitato scientifico-editoriale: Felice Addeo – Dip. Scienze Politiche e della Comunicazione (Univ. Salerno); Roberto Albera – Dip. Scienze Chirurgiche (Univ. Torino); Biagio Aragona – Dip. Scienze Sociali (Univ. Federico II, Napoli); Sandro Brignone – Dip. Filosofia e Scienze dell'Educazione (Univ. Torino); Angelo Cangelosi – School of Computer Science (Manchester, UK); Marco Cantamessa – Dip. Ingegneria Gestionale e della Produzione (Polit. Torino); Elena Cattellino – Dip. Scienze umane e sociali (Univ. Valle d'Aosta); Bernard Cousin – Dépt. Histoire (Aix-Marseille, France); Giuseppe Luca De Luca Picione – Dip. Economia, Management, Istituzioni (Univ. Federico II, Napoli); Marco Devecchi – Dip. Scienze Agrarie, Forestali e Alimentari (Univ. Torino); Maria Adelaide Gallina – Dip. Filosofia e Scienze dell'Educazione (Univ. Torino); Cristina Ispas – Dept. Educational Sciences (Resita, România); Graziano Lingua – Dip. Filosofia e Scienze dell'Educazione (Univ. Torino); Salvatore Livatino – Dept. Engineering and Technology (Hatfield, UK); Vincenzo Lombardo – Dip. Informatica (Univ. Torino); Sergio Margarita – Dip. Management (Univ. Torino); Francesco Mazzeo Rinaldi – Dip. Scienze Politiche e Sociali (Univ. Catania); Giorgio Metta – Istituto Italiano di Tecnologia (IIT); Witold Misiuda-Rewera – Uniwersytet Marii Curie-Skłodowskiej (Lublin, Polska); Silvano Montaldo – Dip. Studi Storici (Univ. Torino); Anabel Moriña Díez – Dept. Didáctica y Organización Educativa (Sevilla, España); Giovanni Onore – Fundación Otonga (Quito, Ecuador); Tania Parisi – Dip. Filosofia e Scienze dell'Educazione (Univ. Torino); Stefano Pesce – Comau Academy, Education and Business Development (Grugliasco); Davide Porporato – Dip. Studi Umanistici (Univ. Piemonte Orientale, Vercelli); Bruno Siciliano – Dip. Ingegneria Elettrica e delle Tecnologie dell'Informazione (Univ. Federico II, Napoli); Pietro Terna – Dip. Scienze Economico-Sociali e Matematico-Statistiche (Univ. Torino); Simona Tirocchi – Dip. Filosofia e Scienze dell'Educazione (Univ. Torino); Roberto Trincherò – Dip. Filosofia e Scienze dell'Educazione (Univ. Torino).

Le scienze umane e le scienze naturali sono destinate a cooperare nonostante la frattura cognitiva esistente. Questa collana, che nasce con il coinvolgimento di studiosi dei due campi, vede nella ricerca e nell'uso delle nuove tecnologie il luogo sia fisico sia concettuale per la creazione di un insieme di modelli di relazioni di riferimento per la costruzione di teorie e per l'orientamento di scelte rilevanti in campo politico, economico, industriale, tecnologico, sanitario, educativo, ambientale, storico, sociale.

Tutti i testi sono preventivamente sottoposti a referaggio anonimo.



Il presente volume è pubblicato in open access, ossia il file dell'intero lavoro è liberamente scaricabile dalla piattaforma **FrancoAngeli Open Access** (<http://bit.ly/francoangeli-oa>).

FrancoAngeli Open Access è la piattaforma per pubblicare articoli e monografie, rispettando gli standard etici e qualitativi e la messa a disposizione dei contenuti ad accesso aperto. Oltre a garantire il deposito nei maggiori archivi e repository internazionali OA, la sua integrazione con tutto il ricco catalogo di riviste e collane FrancoAngeli massimizza la visibilità, favorisce facilità di ricerca per l'utente e possibilità di impatto per l'autore.

Per saperne di più: [Pubblica con noi](#)

I lettori che desiderano informarsi sui libri e le riviste da noi pubblicati possono consultare il nostro sito Internet: www.francoangeli.it e iscriversi nella home page al servizio "[Informatemi](#)" per ricevere via e-mail le segnalazioni delle novità.

Enrica Amaturò, Biagio Aragona,
Suania Acampa e Francesco Amato

Tecnologie emergenti e ricerca sociale

Prospettive teoriche e pratiche metodologiche



**Percorsi
di ricerca**

FrancoAngeli

OPEN ACCESS

Questo volume è stato realizzato con il contributo dei fondi del Progetto di Ricerca Nazionale (PRIN) 2022 «Datafication and Citizenship: the case of Public Administrations in Italy», finanziato dal Ministero dell'Università e della Ricerca (MUR) nell'ambito del bando PRIN 2022, coordinato dall'Università degli Studi di Napoli Federico II, in collaborazione con l'Università degli Studi di Milano-Bicocca e l'Università degli Studi Roma Tre. Codice progetto: 000026-PRIN_2022_SH3-2022H2CFLZ. CUP: E53D23011830006.

Copyright © 2026 by FrancoAngeli s.r.l., Milano, Italy.

Pubblicato con licenza *Creative Commons*
Attribuzione-Non Commerciale-Non opere derivate 4.0 Internazionale
(CC-BY-NC-ND 4.0).

Sono riservati i diritti per Text and Data Mining (TDM), AI training e tutte le tecnologie simili.

L'opera, comprese tutte le sue parti, è tutelata dalla legge sul diritto d'autore.
L'Utente nel momento in cui effettua il download dell'opera accetta tutte le condizioni
della licenza d'uso dell'opera previste e comunicate sul sito
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.it>

Gli eventuali link attivi e QR code inseriti nel volume sono forniti dall'autore.
L'editore non si assume alcuna responsabilità sui link attivi e QR code ivi contenuti
che rimandano a siti non appartenenti a FrancoAngeli.

Indice

Introduzione , di <i>Enrica Amaturò, Biagio Aragona</i>	pag.	7
--	------	---

Parte I – Fare ricerca con e sulle tecnologie emergenti

1. Le tecnologie emergenti come oggetti sociodigitali , di <i>Suania Acampa, Francesco Amato, Biagio Aragona</i>	»	15
2. Le tecnologie emergenti nei disegni di ricerca: oggetto o strumento? , di <i>Suania Acampa, Francesco Amato, Biagio Aragona</i>	»	25

Parte II – Le tecnologie emergenti: genealogie, definizioni e sfide etico-sociali

3. I Large Language Models , di <i>Dario Chianese</i>	»	37
4. Sistemi di decisione automatizzata , di <i>Francesco Amato</i>	»	50
5. Gemelli digitali , di <i>Mattia De Angelis</i>	»	61
6. Le tecnologie quantistiche , di <i>Suania Acampa</i>	»	68

Parte III – Metodi e pratiche di ricerca per le tecnologie emergenti

7. Regolamentare: politiche e governance , di <i>Suania Acampa, Francesco Amato, Vincenzo Esposito</i>	»	81
8. Immaginare: narrazioni, promesse e visioni del futuro , di <i>Suania Acampa, Dario Chianese</i>	»	92

9. Valutare: modelli e pratiche di assessment , di <i>Francesco Amato, Dario Chianese</i>	pag.	104
10. Interrogare: interazione e prompting , di <i>Dario Chianese</i>	»	113
11. Co-creare l'innovazione: attori, reti e infrastrutture , di <i>Mattia De Angelis</i>	»	124
12. Anticipare l'innovazione: consultare gli esperti, costruire scenari , di <i>Vincenzo Esposito, Cristiano Felaco</i>	»	133
Conclusioni. Ricerca sociale e tecnologie emergenti: verso il pluralismo metodologico , di <i>Enrica Amaturo, Biagio Aragona</i>	»	145
Bibliografia	»	155
Gli autori	»	173

Introduzione

di *Enrica Amato*, *Biagio Aragona*

Il volume *Tecnologie emergenti e ricerca sociale: prospettive teoriche e pratiche metodologiche* raccoglie i primi cinque anni di studi e ricerche condotti all'interno dell'Osservatorio F.A.S.T. (Futuro, Algoritmi, Società e Tecnologie), istituito presso il Dipartimento di Scienze Sociali dell'Università Federico II di Napoli. L'Osservatorio ha l'obiettivo di anticipare le implicazioni sociali delle tecnologie, al fine di promuovere forme eque, sostenibili e democratiche di società digitale.

Il tema che lega tutti i capitoli del volume sono le Tecnologie Emergenti (TE), in cui ricadono tecnologie come l'intelligenza artificiale, i gemelli digitali, le tecnologie quantistiche, e i sistemi decisionali automatizzati. Ma, prima di entrare nel vivo del volume, è bene fare chiarezza su cosa intendiamo per TE, e perché utilizziamo questo termine. La locuzione tecnologie emergenti è stata coniata per indicare ambiti tecnologici in fase di sviluppo, caratterizzati da elevati livelli di incertezza rispetto alle applicazioni future e alle implicazioni di natura sociale collegate (Rotolo et al., 2015). L'aggettivo emergente rappresenterebbe cinque caratteristiche principali di queste tecnologie: i) novità radicale, (ii) crescita relativamente rapida, (iii) coerenza, (iv) impatto significativo e (v) incertezza e ambiguità.

È soprattutto l'ultima di queste caratteristiche, quella su cui ci concentreremo, e che ci spinge ad utilizzare l'attributo "emergente". Tralascieremo invece tutte le altre, che spesso vengono utilizzate per giustificare narrazioni epocaliste sull'innovazione tecnologica, che non condividiamo. L'epocalismo consiste nello sviluppo di discorsi che polarizzano i dibattiti sulla tecnologia. Il primo ad averne parlato è stato Morozov, che ha applicato l'epocalismo alle affermazioni e alle narrazioni relative a Internet: «considerato come una rottura netta con tutto ciò che è venuto prima, un apice della civiltà prima irraggiungibile» (Morozov, 2013, p. 60). Tutte le narrazioni epocaliste servono a tracciare una netta distinzione tra un prima

e un dopo, tra epoche, e a concentrarsi sui punti di svolta che segnalano un cambiamento radicale. Il carburante dell'epocalismo è il tecnosoluzionismo: l'idea che la nuova tecnologia possa risolvere problemi vecchi e non ancora definiti nelle nostre società, istituzioni politiche, imprese e scienze. Tutto ciò che è incompatibile con questa tecnologia nuovissima dovrebbe essere spazzato via. L'epocalismo induce radicalizzazione e passività perché il progresso non può essere fermato e non c'è tempo da perdere per capire se il cambiamento rappresenta davvero una rottura con le pratiche esistenti o solo un cambiamento di ordine e di scala. Inoltre, l'epocalismo attenua la critica alla luce della straordinarietà dei tempi e del presunto potenziale rivoluzionario della tecnologia.

Le TE sono tecnologie che si prestano a narrazioni epocaliste. Il calcolo ad alte prestazioni e le tecnologie quantistiche sono presentate come tecniche che stanno ampliando in modo significativo le capacità di elaborazione e analisi dei dati, aprendo nuove prospettive per la ricerca scientifica e per applicazioni in settori come finanza, medicina personalizzata, sicurezza informatica e governance dei dati. I gemelli digitali hanno l'ambizione di costruire rappresentazioni digitali dinamiche di sistemi fisici complessi, consentendo di monitorarne il funzionamento, simulare scenari futuri e supportare processi decisionali basati su modelli predittivi in settori quali la pianificazione urbana, la gestione delle infrastrutture e la produzione industriale. I modelli linguistici di grandi dimensioni (LLM) si crede che siano sul punto di rivoluzionare le modalità di produzione, accesso e circolazione della conoscenza, intervenendo nei processi di scrittura, analisi e interpretazione dei contenuti, oltre che su ampi settori del mercato del lavoro e della cultura.

Che si condividano o meno, queste narrazioni epocaliste includono aspettative di futuro che le TE possono realizzare. Le promesse di trasformazione radicale ad esse associate contribuiscono a mobilitare risorse economiche, orientare programmi di ricerca, definire priorità politiche e costruire consenso attorno a specifiche traiettorie tecnologiche. Le narrazioni sull'intelligenza artificiale (e più recentemente sui LLM) come motore di una nuova economia della conoscenza, sulle tecnologie quantistiche come infrastrutture strategiche per la sicurezza e la sovranità tecnologica o sui gemelli digitali come strumenti capaci di simulare e governare sistemi complessi, sono performative, vanno cioè a produrre attivamente le traiettorie di sviluppo di queste tecnologie. Lontane da essere unicamente il risultato di avanzamenti scientifici e ingegneristici, esse sono parte di configurazioni sociotecniche più ampie nelle quali infrastrutture, istituzioni, pratiche sociali e priorità economiche e politiche si intrecciano in modo profondo.

Ci sono quindi almeno due buoni motivi per cui gli scienziati sociali dovrebbero rivolgere il loro sguardo alle TE. Se, da un lato, la loro penetrazione nella vita sociale, culturale, produttiva e organizzativa delle nostre società digitali le rende interessanti oggetti di ricerca per tutti gli ambiti di studio della sociologia, d'altro canto le TE attirano l'interesse di quei ricercatori sociali che vogliono sperimentare forme più o meno tecnologicamente intensive di ricerca sociale, e che stanno progressivamente integrando IA, gemelli digitali, e la potenza di calcolo delle tecnologie quantistiche nelle loro pratiche di raccolta, organizzazione e analisi dei dati, anche qualitativi, sollevando fondamentali questioni metodologiche. Questi due aspetti sono entrambi centrali nel volume e si concretizzano in una serie di sfide teoriche e metodologiche che ci proponiamo di affrontare, adottando una postura ben precisa, che noi chiamiamo ottimismo critico (Amaturo, Aragona, 2021).

Contrastando l'epocalismo, l'ottimismo critico è un concetto che è servito a molti autori per spiegare i propri orientamenti riguardo al futuro, soprattutto al futuro tecnologico. Significa non essere né meccanicistici nella nostra immaginazione del mondo, né ingenui riguardo alle aspettative che il futuro può riservarci. Il concetto è stato impiegato in diverse discipline, dalla pedagogia (Freire, 1996) alla storia dell'arte e all'ingegneria informatica (Brueckner, 2018). L'ottimismo critico tende a bilanciare le aspettative tecnologiche, evitando gli estremi quando si affrontano le sfide che le TE pongono alle scienze sociali e ai suoi metodi. Se infatti le scienze sociali si rivolgono alle TE come oggetti di studio, perché non c'è sfera del sociale che non sia interessata dalla loro implementazione, d'altro canto il metodo della ricerca sociale sta assumendo forme nuove che vedono una sempre maggiore integrazione delle TE nella ricerca sociale.

Il lavoro su *The Social Life of Methods*, avviato da Mike Savage (2013) (insieme a John Law ed Evelyn Ruppert), ha dimostrato che i metodi di ricerca sociale non si collocano «*al di sopra*» dei mondi che indagano, ma sono, piuttosto, indissolubilmente «*parte integrante*» di quei mondi. I metodi traducono i flussi dinamici ed eterogenei di azione nel mondo in alcune forme di dati, e non in altre: essi «*possono essere identificati come la materia stessa della vita sociale*» (Savage, 2013, p. 5). Allo stesso tempo, i metodi di ricerca stessi sono costituiti, mobilitati e riconfigurati attraverso le loro relazioni con i mutevoli rapporti sociali, economici e politici. Adottare questa visione di Savage, vuol dire considerare le TE non soltanto come oggetto di analisi, ma anche come risorse analitiche che possono essere integrate nei dispositivi di ricerca. Le stesse infrastrutture digitali, i modelli computazionali e gli ambienti simulativi che caratterizzano le TE offrono infatti nuove possibilità per la raccolta, l'elaborazione e l'interpretazione dei dati sociali.

Questa duplice dimensione – tecnologie come oggetti e strumenti di ricerca – contribuisce a ridefinire il rapporto tra metodi e fenomeni studiati. Sempre più spesso, infatti, l'analisi delle TE si svolge all'interno degli stessi ambienti nei quali esse operano, attraverso dati, piattaforme e strumenti che fanno parte dei processi oggetto di indagine. Ne deriva una trasformazione delle pratiche di ricerca che richiede competenze interdisciplinari e una più stretta integrazione tra approcci qualitativi, quantitativi e computazionali. L'impiego delle TE come strumenti di ricerca non è privo di implicazioni epistemologiche e politiche in quanto richiede di interrogarsi sulle forme di conoscenza prodotte, sulle dimensioni privilegiate nell'analisi e sul modo in cui le tecnologie partecipano alla costruzione degli stessi oggetti di ricerca. Per queste ragioni, lo studio delle TE richiede un approccio metodologicamente pluralista e riflessivo, capace di combinare strategie di ricerca diverse e di adattarsi a oggetti in continuo aggiornamento. Piuttosto che applicare metodi predefiniti, la ricerca sociale è chiamata a rielaborare e integrare strumenti analitici differenti, valorizzando le possibilità offerte dalle tecnologie senza rinunciare a uno sguardo critico sulle loro implicazioni.

Il presente volume si inserisce in questo quadro, proponendo una riflessione interdisciplinare sulle TE e sulle modalità attraverso le quali esse possono essere indagate. Invece che limitarsi esclusivamente alle caratteristiche tecniche dei sistemi, i contributi raccolti analizzano le condizioni sociotecniche che ne rendono possibile lo sviluppo e le forme di conoscenza necessarie per comprenderne le implicazioni sociali. Il volume intende offrire uno spazio di confronto tra prospettive teoriche e approcci metodologici diversi, con l'obiettivo di esplorare come le TE vengano progettate, implementate e interpretate da una pluralità di attori, e come contribuiscano a ridefinire le relazioni tra tecnologia, istituzioni e società. Non si intende offrire una trattazione esaustiva, ma proporre una chiave di lettura critica di alcuni ambiti tecnologici particolarmente rilevanti.

La struttura del volume riflette questa impostazione. La prima parte (capp. 1-2) affronta le questioni teoriche, epistemologiche e metodologiche dello studio delle TE. Sul piano teorico, viene proposta una lettura delle tecnologie emergenti come artefatti relazionali, che mette in discussione l'apparente neutralità tecnica a favore di una concezione che ne evidenzia la natura co-prodotta e contestata, facendo emergere i processi di negoziazione, appropriazione e adattamento che caratterizzano il loro sviluppo e la loro implementazione. Sul piano epistemologico e metodologico, viene delineata una prospettiva pluralista che integra approcci e pratiche provenienti non solo dalle scienze sociali, ma anche da altri ambiti disciplinari. In questo quadro, le TE vengono considerate sia come oggetti sia come

strumenti della ricerca sociale, capaci di trasformare le pratiche di raccolta, elaborazione e analisi dei dati.

La seconda parte (capp. 3-6) fornisce un quadro introduttivo delle tecnologie analizzate nel volume, offrendo una mappatura critica e comparativa dei loro ambiti applicativi e delle principali sfide sociotecniche che sollevano. Ogni capitolo è dedicato a una specifica tecnologia – LLM, sistemi di decisione automatizzata, gemelli digitali, tecnologie quantistiche – e ne ricostruisce le traiettorie di sviluppo, le implicazioni e le principali questioni etiche connesse alla loro adozione in contesti specifici, con particolare attenzione alle dinamiche di governance e ai dispositivi di valutazione che ne accompagnano la diffusione.

La terza parte (capp. 7-12) presenta una serie di approcci metodologici e pratiche di ricerca che consentono di investigare empiricamente queste tecnologie. Nel loro insieme, i contributi raccolti mostrano come lo studio delle TE richieda approcci analitici capaci di cogliere la complessità delle configurazioni sociotecniche contemporanee e di interrogare le implicazioni dei processi tecnologici nei diversi contesti in cui prendono forma.

Attraverso questa articolazione, il volume si propone come una guida critica e interdisciplinare per comprendere e studiare le TE nella loro complessità e incertezza, e allo stesso tempo come un luogo di confronto e riflessione sul ruolo della tecnologia nella ricerca sociale, e sulla messa alla prova di pratiche, metodi e categorie interpretative dell'analisi sociologica per la comprensione della società digitale.

Parte I
Fare ricerca con e sulle tecnologie emergenti

1. Le tecnologie emergenti come oggetti sociodigitali

di *Suania Acampa, Francesco Amato, Biagio Aragona*

In un mondo in cui le tecnologie digitali sono inestricabilmente connesse alla struttura e alle dinamiche della società, le tecnologie emergenti non vanno considerate come semplici strumenti, ma come fenomeni socio-digitali. Per sociodigitale (Halford, Southerton, 2024), intendiamo l'ultima evoluzione del quadro teorico da cui partire per esplorare l'intreccio multiforme, variegato ed esteso tra le tecnologie digitali e le società contemporanee. Un quadro teorico che si avvale di concetti tratti dagli studi sulla scienza e sulla tecnologia, dalla teoria dell'assemblaggio, dal post-umanesimo e dalla teoria della pratica sociale, correnti di pensiero che hanno portato a fondamentali riflessioni sugli aspetti materiali e sociali delle tecnologie, coniando termini come sociotecnico, sociomateriale e, infine, sociodigitale. In questo capitolo si cercherà brevemente di richiamare queste teorie fondamentali e di identificare quei costrutti teorici che risultano imprescindibili per affrontare lo studio delle tecnologie emergenti, come, ad esempio, quello di performatività (Law, Singleton, 2000) e le implicazioni epistemologiche che ne derivano.

Sociotecnico, sociomateriale e sociodigitale: evoluzione di un fenomeno

L'evoluzione dei concetti che connettono gli aspetti tecnologici e sociali delle tecnologie va rintracciata in una serie di teorie cognate, la cui eredità maggiore è quella dei Science and Technology Studies (STS).

Gli STS hanno offerto il quadro teorico e gli strumenti metodologici per studiare come le tecnologie siano modellate dalle dinamiche sociali e, al tempo stesso, siano capaci di riconfigurarle, consentendo un'analisi più approfondita e critica delle loro implicazioni sociali.

Uno degli strumenti concettuali chiave sviluppati negli STS è la teoria attore-rete (*Actor-Network Theory*, ANT) (Callon, 1984; 2001; Latour,

2005), che concepisce la realtà sociale come composta da reti eterogenee di attori, umani e non umani, legati da relazioni di reciproca configurazione. Secondo questa prospettiva, gli artefatti tecnologici sono chiamati “attanti” e partecipano alla co-produzione degli assetti sociali, organizzativi e simbolici, poiché non sono concepiti come meri supporti neutri all'azione umana. Questa considerazione evidenzia come il ridurre l'analisi a singoli elementi tecnologici porti ad una comprensione parziale e riduttiva dei processi in atto.

Le tecnologie emergenti non possono perciò essere studiate isolatamente, ma dovrebbero essere considerate all'interno di reti che sono sia sociali che tecnologiche, appunto sociotecniche, più ampie, che comprendono infrastrutture materiali e digitali, pratiche organizzative, routine professionali, utenti finali, personale tecnico, autorità regolatorie, organismi di controllo, comunità epistemiche, imprese e soggetti del terzo settore, nonché i quadri normativi e gli immaginari collettivi che ne orientano progettazione, implementazione e uso.

Pinch e Bijker (1984), nell'ambito dell'approccio della costruzione sociale della tecnologia (*Social Construction of Technology, SCOT*), elaborano il concetto di configurazione sociotecnica, sottolineando come le interazioni sociali, le pratiche culturali e i contesti istituzionali influenzino lo sviluppo e la stabilizzazione delle tecnologie. A differenza della teoria attore-rete, che attribuisce simmetria analitica ad attori umani e non umani nel plasmare gli esiti tecnologici, la SCOT pone l'accento, in primo luogo, sui processi attraverso cui gli artefatti vengono socialmente costruiti mediante l'azione umana, le negoziazioni e i conflitti tra gruppi sociali rilevanti. Secondo questa prospettiva, le tecnologie emergenti non sarebbero strumenti neutri privi di visione, ma oggetti configurati da scelte, valori e interessi sociali, politici ed economici di cui sono precisate, dagli attori coinvolti nella loro produzione e uso, le possibili applicazioni e i criteri di esclusione. Possibilità e limiti non derivano quindi soltanto da vincoli tecnici, ma da vincoli e opportunità socialmente prodotti.

Oltre alle teorie STS, un'altra teoria chiave è quella dell'assemblaggio. Deleuze e Guattari, nel 1980, parlano di “*agencement*” per indicare configurazioni che sfidano il pensiero rappresentazionale e il controllo disciplinare della conoscenza, riconoscendo la produzione di oggetti sociotecnici come esito di campi di forze e processi eterogenei (Deleuze, Guattari, 1980; Aragona, Felaco, 2018). Gli assemblaggi possono essere intesi come intrecci provvisori e dinamiche di realtà interconnesse (Law, 2004). Visti come assemblaggi, i sistemi tecnologici agiscono attraverso specifiche disposizioni di strumenti, attrezzature, esseri umani, artefatti, algoritmi e testi (Callon, 2007; 2016). L'assemblaggio è progettato e gestito per generare,

circolare e reinterpretare continuamente informazioni, rendendo le tecnologie emergenti dinamiche e reattive, poiché vengono costantemente riconfigurate tramite nuove integrazioni di dati, modifiche algoritmiche, attività e obiettivi (Kitchin, Lauriault, 2014). Ogni aspetto delle tecnologie emergenti, che si tratti di un algoritmo, di un sensore o di un'interfaccia utente, interagisce e si integra con altri apparati e sistemi di dati, contribuendo a formare un ecosistema complesso e interconnesso. Questo intreccio rende problematico isolare empiricamente una singola tecnologia senza considerare le reti più ampie che essa modella e da cui è modellata, attraverso interazioni e connessioni continue tra apparati tecnici e sociali (Burrell, 2016; Kitchin, 2019). Non si tratta della semplice somma di elementi sociali e tecnici, ma di una composizione in cui ogni parte plasma e rimodella le altre. Le costruzioni algoritmiche che popolano la società digitale sono il prodotto di questa co-evoluzione, in cui codici, protocolli e infrastrutture si intrecciano con norme sociali, valori culturali e dinamiche economiche. In questo quadro, le decisioni progettuali e le modalità di implementazione non sorgono nel vuoto, ma risultano da processi articolati che coinvolgono una pluralità di stakeholder.

Adottando le teorie della pratica sociale (Gherardi, 2012; 2016), la produzione e il consumo di tecnologie possono essere compresi come una pratica situata, in cui la materialità della tecnologia non funge da semplice supporto inerte, ma partecipa attivamente alla negoziazione degli assetti sociali attraverso intra-azioni tra elementi eterogenei (Gherardi et al., 2019). In tal senso, le tecnologie emergenti si configurano come assemblaggi sociomateriali dove l'agire tecnologico e l'agire umano non sono sfere separate, ma si codeterminano costantemente attraverso l'uso quotidiano. Seguendo l'approccio *practice-based*, la tecnologia non viene semplicemente "adottata", ma "eseguita" (*performed*) all'interno di pratiche che ne ridefiniscono continuamente il significato e l'impatto organizzativo: ciò richiede di osservare tali pratiche sia "dall'esterno" – analizzandone le configurazioni e gli effetti sociali – sia "dall'interno" – considerando il vissuto esperienziale e incarnato degli attori coinvolti (Gherardi, 2012). Anche in questa prospettiva, le tecnologie emergenti non si limitano a essere strumenti neutrali, ma diventano elementi che plasmano e sono plasmati dalle relazioni organizzative, generando trasformazioni che emergono dall'intreccio continuo tra materialità, azione e significato.

Tale prospettiva sociomateriale contrasta la rappresentazione diffusa di queste tecnologie come entità principalmente immateriali o puramente "virtuali" (Suchman, 2007; Gillespie, 2014; Dourish, 2017). Le infrastrutture che le sostengono, server, data center, reti di comunicazione, dispositivi terminali, si fondano su risorse materiali ed energetiche ingenti e

localizzate, con conseguenze ambientali rilevanti in termini di consumo elettrico, estrazione di materie prime, produzione e smaltimento di hardware e rifiuti elettronici (Crawford, Joler, 2018; Crawford, 2021). Questa dimensione materiale non si esaurisce negli aspetti infrastrutturali, ma si intreccia con configurazioni sociali ed economiche specifiche. Le tecnologie emergenti operano sempre all'interno di contesti situati, definiti da norme culturali, pratiche organizzative, regimi regolatori e relazioni di potere. Le implicazioni sociali che esse hanno incidono sull'accesso a servizi e risorse, sulle opportunità economiche e, in molti casi, su diritti civili e sociali, contribuendo talvolta a riprodurre o amplificare disuguaglianze esistenti (Aragona, 2022). La geografia materiale dei dati, dove vengono raccolti, elaborati, archiviati e trasferiti, apre inoltre questioni di sovranità, sicurezza, giurisdizione e responsabilità etica, che non possono essere comprese senza tenere conto della logistica e dell'infrastruttura geopolitica che rendono possibile il funzionamento quotidiano di queste tecnologie. In questa prospettiva, la materialità delle tecnologie emergenti non riguarda soltanto il "supporto" fisico su cui si appoggiano algoritmi e dati, ma anche il modo in cui codici, protocolli, piattaforme e infrastrutture si intrecciano con norme sociali, valori culturali e dinamiche economiche, producendo effetti concreti nella vita quotidiana.

In linea con gli STS, Gherardi (2016) interpreta la sociomaterialità come una declinazione del postumanesimo, che amplia l'orizzonte analitico oltre la centralità esclusiva del soggetto umano, includendo cose, artefatti, infrastrutture e altre entità non umane come attori a pieno titolo. In questo senso, le tecnologie emergenti appaiono non solo come oggetti da valutare o regolamentare, ma come nodi in cui si condensano e ridefiniscono pratiche, saperi, gerarchie e forme di valore. Prendere sul serio la loro sociomaterialità significa, quindi, interrogare congiuntamente le condizioni materiali del loro funzionamento e le conseguenze sociali, culturali e politiche che derivano dalla loro presenza capillare nella società contemporanea.

In sociomateriale, il termine "materiale" non si limita al tangibile, ma include anche la materialità del digitale (Sundin, Haider, 2023). Algoritmi, dati e software sono così considerati entità materiali in senso pieno, poiché partecipano alla configurazione delle pratiche sociali e, al tempo stesso, ne sono configurati (Dourish, 2017). Analogamente a quanto avviene per il sociotecnico, la sociomaterialità non presuppone due domini separati, uno "sociale" e uno "materiale", da mettere in relazione a posteriori, ma un processo di co- costituzione e intreccio continuo. Da questa angolazione, trattare la materialità come separabile dal "sociale" perde significato. In questa prospettiva, il mondo è assunto come fondamentalmente sociomateriale, poiché costantemente ricreato e rimodellato attraverso disposizioni e

pratiche che legano insieme elementi umani e non umani (Sundin, Haider, 2023).

Queste prospettive ontologiche postumaniste sono rafforzate dal concetto di intra-azione sviluppato da Barad (2007), che rifiuta l'idea di entità già date che poi entrano in relazione e sostiene invece che i fenomeni siano il risultato di relazioni costitutive. Quando Barad (2007, pp. 308-309) afferma che «*phenomena are the ontological inseparability of intra-acting agencies*», intende che i fenomeni non sono cose già formate che interagiscono, ad esempio, il sociale e il digitale, ma configurazioni relazionali in cui gli “intra-agenti” non preesistono alla relazione, ma emergono e si definiscono proprio attraverso l'intra-azione stessa. Nonostante ciò, le connessioni che compongono tali configurazioni sono instabili, si trasformano nel tempo e rendono lo studio dei fenomeni un processo necessariamente iterativo. Pur insistendo su questa dinamicità, Barad riconosce che i fenomeni possono stabilizzarsi temporaneamente in forme sufficientemente durevoli da poter essere analizzate empiricamente. È in questo contesto che introduce il concetto di “taglio agenziale”, una delimitazione situata e istituita dal ricercatore per fini analitici che definisce ciò che viene incluso ed escluso dall'osservazione. Un taglio agenziale è dunque una configurazione momentaneamente stabilizzata, resa intelligibile attraverso separazioni temporali, spaziali e relazionali che permettono di descrivere e rilevare ciò che, in realtà, rimane inscritto in un processo più ampio.

È in questo senso che le tecnologie emergenti vanno intese come fenomeni sociodigitali. Questa combinazione di sociomaterialità e intra-azione promuove un approccio costruttivista e realista allo stesso tempo dei sistemi sociotecnici, comprese le tecnologie emergenti. Il taglio agenziale non è una semplice operazione di selezione metodologica, ma, bensì, esplicita che l'oggetto di ricerca non preesiste neutralmente allo sguardo del ricercatore e che, piuttosto, viene costituito attraverso pratiche di demarcazione che partecipano alla realtà che si intende analizzare. Le intra-azioni non si limitano a registrare una realtà già data, ma la plasmano, producendo una rete di significati che emergono dalle pratiche di indagine, progettazione, implementazione e uso. Allo stesso tempo, quando un taglio agenziale è stato prodotto, esso produce stabilità e oggettività di quel fenomeno nelle sue implicazioni sociali.

Il passaggio dal concetto di “sociotecnico” a quello di *sociodigitale* è quindi un cambiamento significativo e distintivo, non semplicemente un altro esempio della natura sociotecnica del mondo. Ci invita a riconsiderare il costruttivismo sociale e il realismo in un'unica cornice onto-epistemologica che Barad chiama realismo agenziale (Barad, 2007).

Agency e performatività delle tecnologie emergenti

Analizzare le tecnologie emergenti attraverso i quadri teorici sopraccitati consente di cogliere come esse risultino da processi di negoziazione e decisione che riflettono rapporti di potere, gerarchie istituzionali e sistemi di valore delle società in cui vengono progettate e utilizzate. Le tecnologie emergenti operano come “attanti” all’interno delle reti, nel senso che influenzano e sono a loro volta influenzate da altri attanti, quali altre infrastrutture e sistemi tecnologici terzi, a cui si aggiungono, come abbiamo esplicitato, anche gli utenti, le organizzazioni e le policy aziendali e istituzionali.

Adottare una prospettiva integrata che combini la capacità descrittiva della teoria attore-rete con l’attenzione al dinamismo proposta dalla teoria dell’assemblaggio consente di seguire nel tempo le pratiche attraverso cui gli attori rendono operativi dispositivi e infrastrutture, e di osservare come tali pratiche si riconfigurino in risposta a nuovi input, obiettivi, vincoli normativi e innovazioni tecniche.

All’interno di questo quadro, la metafora del “districare la matassa” – il groviglio sociotecnico – rimanda a un lavoro analitico di ricostruzione della rete in cui si cerca di rendere visibili i passaggi attraverso cui flussi di dati, regole decisionali, pratiche professionali e vincoli istituzionali vengono progressivamente tradotti in azioni operative e decisioni concrete. L’operazione non mira a separare artificialmente il “sociale” dal “digitale”, ma a seguire i collegamenti che legano fra loro componenti algoritmiche, infrastrutture, procedure organizzative, attori istituzionali, utenti e soggetti regolatori, così da chiarire come determinate configurazioni di rete rendano possibili alcuni esiti, ne impediscano altri e distribuiscano in modo specifico rischi e responsabilità.

In questo quadro, le tecnologie non sono meri strumenti passivi, piuttosto, prendono parte alla definizione di ciò che conta come problema, categoria, rischio, risorsa o valore. I loro significati non sono semplicemente attribuiti dall’esterno, ma emergono dal modo in cui queste entità partecipano a pratiche concrete, contribuendo a stabilire ciò che diventa visibile, misurabile, categorizzabile e quindi governabile. La nozione di performatività aiuta a spiegare l’emergere di proprietà quali l’agency all’interno delle pratiche, e il modo in cui tali proprietà si manifestano operativamente (Sundin, Haider, 2023). La concezione performativa della realtà non si limita a riconoscere che le entità “agiscono”, ma insiste sul fatto che ogni azione, interazione e relazione contribuisce a produrre realtà sociali e materiali in forme contingenti e situate (Latour, 2005). Il concetto è strettamente legato alle genealogie foucaultiane e alle teorie di Butler. Foucault

(1980) evidenzia come conoscenza, corpi e potere siano prodotti attraverso dispositivi e pratiche, anziché dati una volta per tutte, mentre, invece, Butler (1999) concepisce il genere come risultato di azioni performative reiterate, offrendo un modello dinamico e processuale dell'identità di genere (Introna, 2013). In entrambi i casi, fenomeni che potrebbero apparire “naturali” o “stabili” sono invece il risultato di pratiche ripetute che producono e consolidano specifiche forme di realtà.

Applicata alle tecnologie emergenti, la performatività significa che esse non si limitano a “riflettere” il mondo, ma partecipano alla sua configurazione. Integrando algoritmi e modelli predittivi con dati sociali, storici e culturali, le tecnologie emergenti contribuiscono a definire parametri di valore, schemi di classificazione, criteri di inclusione ed esclusione. Attraverso operazioni di categorizzazione e valutazione automatizzata, influenzano comportamenti, aspettative e forme di riconoscimento sociale, attribuendo significati e status a persone, eventi e oggetti. La classificazione automatica degli individui a fini di social scoring o la configurazione di sistemi di sorveglianza avanzati non rappresentano semplicemente realtà preesistenti, ma concorrono a costituire e rinforzare gerarchie, categorie e confini sociali che finiscono per essere legittimati dalla loro stessa applicazione.

Pensiamo ad applicazioni controverse come i sistemi di riconoscimento facciale per identificare e classificare gli individui in tempo reale. Tali sistemi non sono solo strumenti per l'elaborazione di dati biometrici, ma dispositivi che introducono e consolidano nuove pratiche di controllo, incidendo sulle percezioni e sulle politiche di privacy, sicurezza e identità.

In questo senso, le tecnologie emergenti operano come pratiche performative: contribuiscono alla (ri)costruzione di norme e aspettative, ridefiniscono confini di legittimità e vulnerabilità, modificano priorità e strutture decisionali nei contesti in cui vengono adottate. Non interagiscono passivamente con un ambiente dato, ma prendono parte alla sua continua ristrutturazione, producendo effetti materiali, simbolici e normativi che ridisegnano rapporti di potere e forme di cittadinanza.

I risvolti epistemologici

Assumere le tecnologie emergenti come fenomeni sociodigitali performativi, risultato di configurazioni sociotecniche e sociomateriali implica riflettere radicalmente sulla posizione del ricercatore. Non si può immaginare che chi studi questi fenomeni sia una figura epistemica esterna, neutrale, che “osserva” sistemi già dati. La conoscenza che si produce ricercando le tecnologie emergenti partecipa alla loro stessa configurazione, contri-

buendo a stabilizzare alcune versioni del reale e a renderne invisibili altre. L'idea di conoscenza situata proposta da Haraway (1988), secondo cui ogni prospettiva è parziale, incarnata e posizionata, riconosce che non esiste uno sguardo “da nessun luogo”, ma soltanto punti di vista collocati che richiedono l'assunzione di responsabilità per ciò che rendono visibile e dicibile. L'idioma della co-produzione rafforza questa impostazione mostrando come modi di conoscere il mondo e modi di organizzarne l'ordine sociale siano intrecciati tra gli assetti istituzionali, le categorie giuridiche e le forme di partecipazione pubblica che non sono semplicemente il contesto in cui le tecnologie emergenti vengono valutate, ma come, piuttosto, vengono ridefiniti insieme alle stesse tecnologie (Jasanoff, 2004).

L'analisi degli assemblaggi e delle reti, in altre parole, è già una pratica di ri-assemblaggio che sposta i confini tra interno ed esterno al sistema, tra attori centrali e marginali, tra ciò che rientra nel perimetro della responsabilità e ciò che ne viene espulso. Le nozioni di intra-azione e taglio agenziale suggeriscono che ogni disegno di ricerca implichi scelte ontologiche e non solo metodologiche. Decidere quali flussi di dati seguire, quali siti osservare, quali documenti analizzare equivale a definire, almeno temporaneamente, che cosa sia “la” tecnologia emergente oggetto di studio e quali versioni di essa diventino oggetto di scrutinio pubblico (Barad, 2007). Il contributo di Mol (1999) sulla politica ontologica (*ontological politics*) permette di esplicitare questo punto; in quanto descrivere un oggetto sociotecnico significa contribuire a far esistere alcune delle sue molteplici possibili versioni e a metterne in ombra altre. La ricerca, in questo senso, è sempre anche pratica di selezione e di composizione del reale, con effetti che non sono neutri rispetto alle opzioni politiche e normative disponibili.

Le implicazioni epistemologiche si intrecciano così con forme specifiche di giustizia o ingiustizia nella distribuzione delle voci e delle esperienze che entrano a far parte della conoscenza scientifica delle tecnologie emergenti. In questa prospettiva, adottiamo la nozione di “ingiustizia epistemica” per analizzare le condizioni in base alle quali alcuni attori vengono riconosciuti come soggetti di conoscenza credibili, mentre altri risultano sistematicamente esclusi o interpretati in modo distorto (Fricker, 2007). Nel campo delle tecnologie emergenti, ciò può riguardare, ad esempio, gli utenti più vulnerabili, le comunità oggetto di sperimentazioni invasive o gli operatori di prima linea che vivono quotidianamente le conseguenze dei processi automatizzati, ma le cui interpretazioni raramente informano le narrazioni ufficiali sull'innovazione digitale e sull'impatto.

Un approccio sociodigitale alla ricerca richiede di interrogare queste asimmetrie epistemiche, chiedendosi quali esperienze vengano trasformate

in “dati”, quali restino sullo sfondo e con quali effetti sulla definizione dei problemi e delle soluzioni.

Da questa prospettiva, le “tecnologie dell’umiltà” (*Technologies of Humility*) proposte da Jasanoff (2003) non sono soltanto un programma per i decisori pubblici, ma anche un orientamento critico per la ricerca. Mettere a fuoco vulnerabilità, distribuzione degli oneri e dei benefici, modalità di apprendimento istituzionale in condizioni di incertezza significa riconoscere che la produzione di conoscenza sulle tecnologie emergenti deve incorporare, fin dall’inizio, interrogativi su chi rischia che cosa, chi può permettersi di sbagliare, chi dispone delle risorse concettuali e organizzative per rendere visibili i propri danni. L’umiltà evocata da Jasanoff non coincide con una rinuncia alla critica, ma con un diverso modo di esercitarla, meno orientato a rafforzare pretese di controllo totale, più attento a costruire spazi in cui l’ignoranza, l’incertezza e la contestazione possano essere trattate come problemi pubblici anziché scarti da neutralizzare. Questo comporta una ridefinizione della responsabilità epistemica del ricercatore. Se la conoscenza è situata e co-prodotta, la questione non è tanto come eliminare la parzialità, quanto come renderla esplicita e politicamente discutibile. Haraway (1988) insiste sulla necessità di “rispondere di ciò che impariamo a vedere”, sottolineando come ogni configurazione di oggetti e attori osservati implichi anche un posizionamento rispetto alle forme di vita che vengono protette, trascurate o sacrificate. La scelta dei tagli agenziali, in questo senso, può essere letta come una forma di impegno, dove seguire certi percorsi di dati, certi momenti decisionali, certe figure professionali significa costruire alleanze conoscitive con alcune parti della rete e non con altre.

Ciò apre una domanda centrale sul ruolo della ricerca rispetto ai conflitti che attraversano gli assemblaggi sociotecnici. Il riferimento alle epistemologie civiche e alle forme di partecipazione proposto dagli STS permette di trattare lo studio delle tecnologie emergenti come luogo di possibili rinegoziazioni delle relazioni tra esperti, istituzioni e cittadini. La ricerca può limitarsi a confermare gerarchie esistenti, ad esempio rafforzando la centralità di specifici saperi tecnico-giuridici, oppure può contribuire a costruire configurazioni più inclusive di competenza e di testimonianza, in cui altri soggetti vengano riconosciuti come co-produttori legittimi di diagnosi e di criteri di giudizio (Jasanoff, 2003; 2004). In questo senso, l’analisi delle tecnologie emergenti diventa anche un laboratorio per sperimentare forme diverse di accountability, in cui le responsabilità siano distribuite lungo l’intera rete di progettazione, implementazione e uso, piuttosto che concentrate su singoli attori o momenti regolatori.

La ricerca non è neutrale rispetto alle alternative immaginabili (Mol, 1999). Descrivere una tecnologia emergente come “inevitabile” o “tecnicamente necessaria” tende a chiudere lo spazio del possibile, mentre seguire le pratiche, le esitazioni e le deviazioni che ne hanno accompagnato lo sviluppo consente di mostrare che altre traiettorie sarebbero state, e restano, praticabili. In questa prospettiva, le nostre indagini empiriche non si limitano a ricostruire gli assemblaggi delle tecnologie emergenti, ma rendono pensabili configurazioni diverse di attori, regimi normativi, metriche di valutazione e forme di controllo. L’apertura di tali spazi controfattuali non è un esercizio puramente teorico, ma alimenta la capacità collettiva di mettere in questione l’asserita “inevitabilità”, l’epocalismo, di specifici sistemi digitali e di rivendicare modi diversi di organizzare il rapporto tra progettazione, ricerca, decisioni, diritti e democrazia.

2. Le tecnologie emergenti nei disegni di ricerca: oggetto o strumento?

di *Suania Acampa, Francesco Amato, Biagio Aragona*

Fare ricerca su/con le TE: le sfide

Come si diceva nell'introduzione, l'interesse degli scienziati sociali per le tecnologie emergenti è doppio. Se da un lato, la loro penetrazione nella vita sociale, culturale, produttiva e organizzativa delle nostre società digitali le rende interessanti oggetti di ricerca per tutti gli ambiti di studio della sociologia, d'altro canto le TE attirano l'interesse di quei ricercatori sociali che vogliono sperimentare forme più o meno intensive di ricerca sociale. Tuttavia, l'integrazione delle TE nei disegni di ricerca sociale solleva importanti questioni metodologiche. Se sul piano teorico sono oramai numerose le riflessioni su cui ci si può basare, e il panorama di teorie che consente di inquadrare le TE è ricco e consolidato, quando si provano a declinare queste teorie nella ricerca empirica le difficoltà sono tante. La prima grande questione riguarda: le TE vanno considerate oggetti di studio o strumenti utili per la ricerca sociale? Come si vedrà nel corso degli altri capitoli del volume, la risposta più semplice a questa domanda è: entrambi. Ma sarebbe una risposta molto superficiale che non renderebbe la complessità di cosa significhi inserire le TE nei nostri disegni di ricerca.

Per comprendere il ruolo delle TE nella ricerca sociale bisogna interrogarsi non solo su ciò che studiamo, ma anche su *come* e attraverso che *cosa* diventa possibile studiarlo.

Si tratta, evidentemente, di una distinzione sempre più difficile da tracciare, poiché le stesse tecnologie che indaghiamo sono sempre più spesso impiegate per produrre conoscenza. Questa duplicità mette alla prova le nostre categorie tradizionali e impone di ripensare le relazioni tra osservatore, oggetto osservato, metodo e tecnica.

Riconoscere le TE come oggetti e strumenti implica, infatti, ammettere che la conoscenza è sempre situata, ossia costruita in ambienti materiali, algoritmici e simbolici che ne condizionano le forme di produzione.

La sfida, per la ricerca sociale *su* e *con* le TE, è tanto tecnica quanto epistemologica: mantenere la propria capacità critica all'interno di un ecosistema cognitivo in cui la tecnologia non è solo un mezzo, ma una forma di pensiero incorporato nelle pratiche di conoscenza (Leonelli, 2019).

Il capitolo si articola in una discussione che parte dalla genealogia teorica che ha portato a considerare la tecnologia come oggetto di indagine – dal funzionalismo e dalla teoria dei sistemi alle prospettive sociotecniche degli Science and Technology Studies (STS) – per poi analizzare come la rivoluzione digitale e la datificazione abbiano avviato la dissoluzione della netta distinzione tra oggetto e strumento, trasformando le TE in vere e proprie infrastrutture epistemiche (Edwards et al., 2011; Edwards, 2013). Concludiamo, infine, discutendo attraverso casi empirici dei modelli linguistici di grandi dimensioni (LLM), dei gemelli digitali e delle tecnologie quantistiche.

Ripensare il confine: le TE come oggetto e strumento della ricerca

Il modo in cui le TE vengono progettate, implementate e utilizzate dipende da processi di negoziazione, adattamento e contestazione continui (Jasanoff, 2004; Suchman, 2007), che coinvolgono un'ampia gamma di attori: dai decisori politici e dagli ingegneri agli utenti finali e ai gruppi di *advocacy*.

La loro implementazione solleva inevitabilmente sfide etiche ed epistemologiche fondamentali legate alla rappresentazione dei dati e al processo decisionale basato sulla modellazione e sulla simulazione. Le tecnologie emergenti si collocano al crocevia di due trasformazioni convergenti: da un lato, la crescente digitalizzazione dei processi sociali; dall'altro, la tecnicizzazione delle pratiche scientifiche, che introducono nuovi strumenti, ridefiniscono il rapporto tra teoria, metodo e tecnica e influenzano la costruzione stessa dell'oggetto di ricerca e la sua rappresentazione in dati (Amaturo, Aragona, 2019).

In questo contesto, la domanda se tali tecnologie possano essere considerate sia oggetti sia strumenti della ricerca sociale riguarda la natura stessa della conoscenza sociologica e i confini, sempre più porosi, tra ciò *che* indaghiamo e ciò *con cui* indaghiamo. Essa rinvia a una questione di fondo: in che misura la tecnologia è oggetto della realtà che studiamo e in che misura è un mezzo attraverso cui la conosciamo?

Le scienze sociali si sono tradizionalmente confrontate con la tecnologia come *oggetto* d'indagine, studiandone le implicazioni sociali, economiche e culturali. Le prospettive sociologiche classiche – dal funzionalismo di Parsons (1951), alla teoria dei sistemi di Luhmann (1995), fino alla sociologia della modernità di Beck (1992) e Giddens (1990) – hanno trattato la tecnologia come un elemento del sistema sociale, capace di modificare le forme dell'agire e dell'organizzazione collettiva. In questi approcci, la tecnologia viene considerata principalmente come elemento funzionale o strumentale all'interno della società (ad esempio, un sottosistema con funzioni specifiche) o come fattore di cambiamento strutturale (cioè qualcosa che contribuisce a trasformare l'organizzazione della società, le istituzioni e il modo di agire delle persone).

Con l'emergere degli STS, a partire dagli anni Settanta e Ottanta, si afferma invece una prospettiva più relazionale: la tecnologia smette di essere un mero artefatto e diventa un attore sociotecnico (Latour, 2005) che partecipa attivamente alla costruzione del mondo sociale. Questa visione, inaugurata dalla teoria dell'attore-rete (ANT) e dai contributi di autori come Callon (1984), Law (1992) e Latour (1992; 2005), ha trasformato il modo in cui la ricerca sociale concepisce la relazione fra tecnologia e società, mostrando come le tecnologie configurino le condizioni stesse della conoscenza e dell'azione collettiva.

Dunque, la tecnologia continua a essere un *oggetto* privilegiato della riflessione sociologica, ma la sua analisi ha assunto nel tempo prospettive diverse. Nei primi studi STS, l'attenzione si è concentrata sulla costruzione sociale della tecnologia: autori come Bijker (1995), Pinch e Hughes (1987) hanno mostrato come gli artefatti tecnici non emergano da un processo lineare di progresso scientifico, bensì da negoziazioni sociali, controversie e adattamenti culturali. Nel suo celebre studio sulla bicicletta, Bijker (1995) ricostruisce la traiettoria di sviluppo che ha portato alla sua forma moderna, evidenziando come diversi gruppi sociali (sportivi, donne, ingegneri, produttori) abbiano attribuito significati differenti all'oggetto tecnico, contribuendo a stabilizzarne il design. Allo stesso modo, Hughes (1993) ha descritto lo sviluppo dei sistemi elettrici come grandi sistemi sociotecnici, composti da reti di attori, dispositivi e istituzioni che co-evolvono. Questi studi hanno contribuito a porre la tecnologia al centro dell'analisi sociologica, mostrando quanto essa sia un processo di costruzione collettiva oltre a essere il risultato dell'avanzamento scientifico e industriale.

Negli ultimi decenni, tuttavia, l'indagine sulle tecnologie ha assunto forme nuove, spostando l'attenzione dalla costruzione sociale degli artefatti alle pratiche e alle materialità attraverso cui la tecnologia agisce nel sociale (Suchman, 2007; Akrich, 1992; Knorr Cetina, 2001). Gli studi sulla ma-

terialità della conoscenza scientifica e sulle infrastrutture tecniche hanno evidenziato che gli artefatti agiscono come attanti che configurano l'azione e la percezione (Star, Bowker, 2010; Leonelli, 2019). La tecnologia, in questa prospettiva, è considerata parte integrante dei processi sociali, cognitivi e simbolici: produce forme di coordinamento, nuovi regimi di visibilità e specifiche modalità di soggettivazione (Lash, 2007; Beer, 2018). Gli studi sugli immaginari sociotecnici (Jasanoff, Kim, 2015; Konrad, Böhle, 2019) hanno mostrato come studiare le tecnologie sulla base delle loro rappresentazioni – che possono oscillare tra visioni apocalittiche e integrate – orienti le politiche economiche, i programmi di ricerca e le agende politiche. Questo approccio mostra quanto le narrazioni, alla base della costruzione degli immaginari sociotecnici, siano performative (Borup et al., 2006; Jasanoff, Kim, 2015), ossia anticipano e legittimano le direzioni di sviluppo scientifico e tecnologico, funzionando come oggetti discorsivi e normativi che strutturano le possibilità del pensiero, dell'azione collettiva e della governance.

Questo slittamento dal considerare le tecnologie come fenomeni da interpretare al riconoscerle come componenti attive dei processi conoscitivi segna un passaggio cruciale nella riflessione contemporanea. Se nei paradigmi precedenti la tecnologia costituiva un oggetto da spiegare, oggi appare sempre più come condizione di possibilità della spiegazione stessa. In particolare, con la rivoluzione digitale, la crescente pervasività dei dati e l'avvento delle piattaforme algoritmiche, la tecnologia cessa di essere soltanto un oggetto di osservazione per diventare anche una condizione di possibilità dell'osservazione stessa: dunque non si limita più a rappresentare la realtà sociale, ma la modella, agendo come mediatore che rende visibili alcuni aspetti dei fenomeni, oscurandone altri.

È soprattutto la datificazione a trasformare le tecnologie in infrastrutture epistemiche, cioè in dispositivi materiali e organizzativi che rendono possibile la produzione stessa della conoscenza (Kitchin, 2014; Leonelli, 2019). Per infrastruttura epistemica si intende l'insieme di tecnologie, standard, pratiche e attori che consentono di raccogliere, classificare, conservare e interpretare i dati, stabilendo così ciò che può essere conosciuto e in che modo. In altre parole, si tratta di ambienti sociotecnici che strutturano l'accesso alla realtà empirica e definiscono le condizioni di possibilità della ricerca.

In questa prospettiva, il confine tra tecnologia come *oggetto* e tecnologia come *strumento* tende a dissolversi: la ricerca sociale non può più essere separata dalle tecnologie che la rendono praticabile, né dalle infrastrutture materiali e simboliche che sostengono la produzione dei dati.

Kitchin (2014) descrive proprio la rivoluzione dei dati come il passaggio verso assemblaggi sociotecnici in cui tecnologie, pratiche e regimi di potere si intrecciano nella produzione della conoscenza. Le TE, in quanto parte di questi assemblaggi, danno forma a nuovi modi di indagare e descrivere il mondo; ciò che Beer (2018) definisce *data gaze*: uno sguardo che filtra e organizza la realtà attraverso la logica dei dati, producendo visibilità selettive e nuove forme di potere epistemico.

In questo scenario, Marres (2017) parla di *empirismo distribuito* per descrivere la coincidenza contemporanea tra produzione e circolazione dei dati nelle piattaforme digitali: gli stessi atti di uso, navigazione, condivisione e interazione generano tracce che diventano immediatamente risorse analitiche. Un esempio di questa trasformazione è rappresentato dai *data sprint* (Munk et al., 2019), laboratori intensivi in cui ricercatori, sviluppatori e designer collaborano per esplorare insiemi di dati complessi. Qui, la conoscenza emerge dall'interazione situata tra codici, visualizzazioni e pratiche interpretative.

Se gli ambienti digitali sono stati tra i primi esempi di contaminazione tra oggetti e strumenti di ricerca, le tecnologie emergenti radicalizzano tale condizione. La loro logica non è soltanto quella della raccolta o dell'elaborazione automatizzata dei dati, ma anche quella della simulazione, della predizione e della generazione di configurazioni sociali e cognitive prima ancora che queste si manifestino empiricamente. Un gemello digitale urbano, ad esempio, accoppiato a flussi di dati in tempo reale, consente di testare diversi scenari alternativi dello stesso contesto urbano, ridisegnando le linee di trasporto, introducendo *pricing* dinamici o redistribuendo servizi essenziali, e di esplorare gli effetti sociali e spaziali di diverse politiche pubbliche (Batty, 2018).

Non si tratta di “fabbricare” fenomeni artificiali, ma di riconoscere che ogni dispositivo d'indagine configura la porzione di mondo che intende osservare, rendendo visibili certe relazioni e mettendone in ombra altre. In questa prospettiva, le tecnologie emergenti ampliano quella “ricerca reinventata” evocata da Marres (2017, p. 19), avvicinandola a forme di sperimentazione collettiva in cui la conoscenza è co-prodotta da reti sociotecniche che includono ricercatori, infrastrutture, standard, algoritmi e comunità d'uso. Il ricercatore, a questo punto, non può più concepirsi come esterno al proprio oggetto: la tecnologia, in quanto ambiente di interazione, raccolta ed elaborazione dei dati, entra a far parte del dispositivo metodologico (Ruppert et al., 2013), ridefinendo i confini tra oggetto e metodo (Amaturo, Aragona, 2019).

La condizione del ricercatore in questo contesto è inevitabilmente riflessiva: l'integrazione delle tecnologie emergenti nei disegni di ricerca

richiede di interrogarsi non solo su *ciò* che osserviamo, ma anche su *come* e *attraverso che cosa* l'osservazione diventa possibile.

Al tempo stesso, adottare le TE come strumenti di ricerca non significa accettarne passivamente le logiche, ma sviluppare una postura metodologica capace di operare *con* e *contro* (Amaturo, Aragona, 2019): “con”, perché le tecnologie ampliano le possibilità di osservazione e di analisi; “contro”, perché è necessario interrogare criticamente i presupposti epistemici e i modelli di realtà che esse incorporano. In questa prospettiva, la riflessività diventa una condizione critica del fare ricerca con le TE: consente di esplicitare le mediazioni tecniche e cognitive che incidono sulla produzione dei risultati e di assumere responsabilità epistemica nei confronti delle proprie pratiche e strumenti.

Infine, l'utilizzo delle TE come *strumenti* d'indagine deve essere accompagnato da una riflessione sulle implicazioni etiche, politiche e cognitive: la questione dell'opacità algoritmica (Burrell, 2016) solleva interrogativi sulla trasparenza e sulla riproducibilità. La conoscenza prodotta attraverso le TE è sempre situata e condizionata da vincoli materiali, infrastrutturali e politici: il loro utilizzo non può essere interpretato come un progresso lineare della conoscenza (poiché non catturano i processi sociali nella loro interezza e complessità) ma ne restituiscono frammenti mediati dagli artefatti che li generano (Amaturo, Aragona, 2019). Ciò implica riconoscere i limiti della conoscenza prodotta mediante le TE e le condizioni di possibilità che le accompagnano: la dipendenza da infrastrutture proprietarie, l'asimmetria nell'accesso ai dati e alle risorse computazionali, la configurazione tecnico-istituzionale dei processi di raccolta e analisi. Il ricercatore si trova così a operare in un contesto eteronomo, in cui l'autonomia scientifica è negoziata all'interno di ecosistemi digitali governati da logiche economiche e tecnologiche. Accettare questa condizione significa rendere la postura critica del ricercatore parte integrante del processo metodologico.

Queste considerazioni teoriche trovano riscontro concreto nei modi in cui le TE vengono già oggi integrate nei progetti di ricerca, modificando non solo ciò che è osservabile, ma anche le modalità stesse dell'osservazione. In altri termini, le implicazioni teoriche fin qui discusse trovano un corrispettivo diretto nelle pratiche di ricerca contemporanee.

Circoscrivendo questa trattazione a esempi empirici, i modelli linguistici di grandi dimensioni (LLM) offrono un primo osservatorio privilegiato della duplice natura delle tecnologie emergenti. Come *strumenti* di ricerca, i modelli linguistici di grandi dimensioni permettono di automatizzare processi tradizionalmente onerosi: l'estrazione di entità e relazioni da grandi corpora testuali, la classificazione multilabel, la sintesi comparata di documenti o la codifica preliminare per l'analisi del contenuto. Queste

funzioni consentono di ampliare la scala empirica e di sperimentare forme di *computational grounded theory* (Nelson, 2020; Törnberg, Uitermark, 2021), in cui i pattern linguistici vengono interpretati alla luce di categorie teoriche. Tuttavia, gli LLM si configurano anche come *oggetti* di ricerca, in quanto riflettono e incorporano visioni culturali, ideologiche e politiche (Bender et al., 2021; Crawford, 2021), richiedendo un'attenzione specifica ai dataset, ai processi di addestramento e ai modelli di inferenza che ne orientano il comportamento e ne determinano i limiti. Inoltre, la loro opacità, dovuta alla complessità dei training e alla scarsità di dati pubblici, solleva anche questioni di *accountability* (Birhane, Prabhu, 2021), che possono essere indagate esaminando le scelte tecniche, organizzative e politiche inscritte nei modelli (dai dataset utilizzati ai processi di addestramento, dalle strategie di filtraggio agli effetti sociali osservabili) al fine di chiarire chi prende le decisioni, con quali criteri e con quali conseguenze per differenti gruppi sociali.

Un secondo ambito emblematico è rappresentato dall'High Performance Computing (HPC), che funge contemporaneamente da *strumento* di analisi e da *oggetto* di indagine sociotecnica. Come strumento, l'HPC consente simulazioni e analisi multilivello su fenomeni complessi – dalla mobilità urbana alle catene di approvvigionamento, fino alla diffusione di contenuti sui social media – consentendo di testare ipotesi *what-if* su scala di popolazione (Conte et al., 2012; Kitchin, 2014). Si pensi, ad esempio, a uno studio sulla resilienza informativa in rete (Cinelli et al., 2021; Farrand, Carrapiço, 2015): un cluster HPC può sostenere migliaia di simulazioni su nodi, messaggi e algoritmi di raccomandazione, valutando come regimi di moderazione diversi incidano sulla circolazione delle informazioni.

Ma l'HPC può essere considerato anche oggetto d'indagine, nella misura in cui è parte di un'infrastruttura che determina chi può elaborare, a quale costo e con quali finalità. Analizzando le asimmetrie che genera, ad esempio, tra attori che possono permettersi l'accesso e attori che ne restano esclusi, è possibile comprendere quanto le scelte relative alla progettazione, distribuzione e accesso a tali tecnologie modellano la possibilità stessa di produrre dati, sviluppare modelli complessi e condurre ricerca avanzata.

I gemelli digitali rappresentano forse il caso più evidente di intreccio tra strumento e oggetto. Come strumenti, i gemelli digitali consentono di costruire simulazioni dinamiche e visualizzabili di processi urbani, organizzativi o ambientali, nonché di testare scenari di intervento in modo controllato: nelle politiche urbane *data-driven*, ad esempio, tali modelli offrono spazi di negoziazione fra attori diversi (decisori, cittadini, imprese) attraverso visualizzazioni condivise e metriche comprensibili (Batty, 2018). Ma proprio per la loro forza pervasiva, i gemelli digitali diventano anche

oggetto di indagine, in quanto veicolano visioni normative sul funzionamento della città e sulle forme di partecipazione legittima (Halpern, Gunel, 2021); alcuni studi mostrano che le scelte su quali variabili includere, come rappresentarle e come comunicare l'incertezza riflettono particolari modelli di governance e valori pubblici (Saltelli, Di Fiore, 2023).

Iniziative come Destination Earth¹ – che integra nel proprio framework strumenti di *uncertainty quantification* e *quality assessment* per la modellazione climatica (ECMWF, 2024) – hanno reso strutturale la quantificazione e comunicazione dell'incertezza (per esempio con il tool AQUA e framework di quality assessment), mentre il National Digital Twin² britannico, attraverso i Gemini Principles e la guida Digital Twin Navigator introduce principi di *trust*, tracciabilità e *accountability* (CDBB, 2018; 2021) che abilitano forme di audit sui gemelli e sui dati. Insieme, questi esempi mostrano come la trasparenza del modello sia un risultato tanto politico (governance, responsabilità) quanto tecnico.

Infine, le tecnologie quantistiche (TQ) aprono un orizzonte ancora in definizione, ma già significativo per la ricerca sociale. Esse promettono sensori ad altissima sensibilità per la ricerca ambientale, protocolli di comunicazione sicura per la gestione dei dati sensibili e, in prospettiva, modelli *quantum-inspired* per la simulazione di reti sociali o decisionali complesse (Vermaas, 2025).

Le TQ diventano oggetti di straordinario interesse sociologico perché mettono in crisi le categorie di causalità lineare e di determinismo che hanno a lungo dominato il pensiero scientifico. I concetti di *entanglement* e coevoluzione, ad esempio – che ne rappresentano i pilastri ontologici ed epistemologici – invitano la ricerca sociale a ripensare le proprie condizioni di osservazione e spiegazione: l'*entanglement* esprime l'interdipendenza costitutiva tra osservatore, metodo e mondo sociale, mostrando come le pratiche attraverso cui il sociale viene studiato contribuiscano anche a plasmarlo; la coevoluzione, invece, evidenzia che le trasformazioni delle infrastrutture tecnologiche procedono di pari passo con quelle dei modi in cui il sociale può essere teorizzato, indagato e modellato. In questa prospettiva, i processi che impiegano le TQ come risorsa computazionale di nuova generazione configurano arene sperimentali di co-creazione, in cui tecnologia e metodo si adattano reciprocamente e la modellazione stessa diviene parte integrante dell'indagine.

Le infrastrutture quantistiche, infatti, trasformano l'osservazione in intervento, facendo sì che oggetto e processo di conoscenza coevolvano

¹ Link: <https://digital-strategy.ec.europa.eu/en/policies/destination-earth>.

² Link: www.cdbb.cam.ac.uk/DFTG/GeminiPrinciples.

all'interno di un'"epistemologia *entangled*" dei fenomeni sociali. Questa prospettiva rompe definitivamente l'illusione della neutralità del metodo e trasforma l'incertezza in una risorsa, riconoscendo che l'oggetto d'indagine non è dato una volta per tutte, ma si attualizza nelle pratiche di osservazione e modellazione. Se il digitale e i modelli predittivi tendono a reintrodurre una forma di determinismo computazionale (Beer, 2018; Couldry, Mejias, 2020), le tecnologie quantistiche, al contrario, rimettono in gioco la probabilità, la non linearità e la coesistenza di stati multipli, offrendo una metafora potente per ripensare la complessità sociale, le modalità della sua conoscenza e della sua rappresentazione.

Conclusioni

L'insieme di queste trasformazioni, tuttavia, non risponde ancora alla domanda iniziale: le TE sono oggetti o strumenti della ricerca sociale? La risposta più feconda è che esse sono entrambe le cose, e lo sono in modi che mettono continuamente alla prova le nostre categorie interpretative.

Ogni tentativo di separare il momento dell'osservazione da quello dell'operatività tecnologica risulta oggi inadeguato, perché le stesse condizioni che rendono possibile la ricerca (i dati, gli algoritmi, le infrastrutture) sono parte integrante del fenomeno sociale che intendiamo indagare. Le tecnologie emergenti non si lasciano confinare né nel dominio dell'oggetto, né in quello dello strumento: sono "mediatrici epistemiche" che modellano simultaneamente ciò che conosciamo e il modo in cui lo conosciamo. Esse costituiscono, inoltre, il punto di intersezione tra processi sociali, pratiche scientifiche e regimi di potere, rendendo la ricerca sociale un'attività intrinsecamente relazionale e situata.

In questa prospettiva, la distinzione tradizionale tra teoria, metodo e tecnica deve essere ripensata come una costellazione dinamica in cui ogni elemento influenza e ridefinisce gli altri.

Le TE, con la loro capacità di generare, modellare e simulare fenomeni sociali, impongono un ripensamento delle forme di evidenza e delle logiche di spiegazione: la conoscenza sociologica diventa così un processo di co-produzione tra attori umani e non umani, tra categorie interpretative e infrastrutture materiali, tra pratiche di analisi e dispositivi di calcolo.

La sfida per la ricerca sociale non consiste soltanto nel padroneggiare tecniche computazionali, ma anche nel mantenere una consapevolezza critica della loro natura situata. Come ha ricordato Haraway (1988), la conoscenza è sempre parziale e incarnata: la tecnologia non è mai neutra, ma

intrecciata alle posizioni da cui si osserva. Riconoscere questa parzialità non significa rinunciare all'ambizione di oggettività, bensì ridefinirla in termini di responsabilità epistemica: essere responsabili dei propri strumenti, delle infrastrutture che li sostengono e delle visioni del mondo che essi incorporano.

Le tecnologie emergenti, in questo senso, rappresentano sia una sfida metodologica sia una possibilità di evoluzione epistemologica: un invito a ripensare il ruolo della ricerca sociale come pratica di interpretazione e di intervento, capace di interrogare criticamente i propri strumenti mentre li impiega per comprendere fenomeni sociali complessi che, al tempo stesso, contribuisce a definire. Coltivando un *ethos* metodologico fondato sulla riflessività, sull'apertura e sulla responsabilità epistemica, la ricerca sociale può continuare a produrre conoscenza affidabile e innovativa, capace di misurarsi criticamente con l'accelerazione tecnologica del presente.

Parte II
Le tecnologie emergenti:
genealogie, definizioni e sfide etico-sociali

3. I Large Language Models

di *Dario Chianese*

Introduzione

Negli ultimi anni i modelli linguistici di grandi dimensioni (LLM) sono diventati strumenti di uso quotidiano per milioni di persone. La rapida diffusione di chatbot come ChatGPT ha reso evidente una trasformazione che riguarda non solo la tecnologia, ma anche le pratiche attraverso cui si comunica, si produce conoscenza e si organizzano attività professionali. L'accesso a funzioni come la generazione di testo, la sintesi di documenti, la programmazione assistita o la traduzione automatica tramite interfacce conversazionali ha ampliato in modo significativo l'autonomia operativa degli utenti e la velocità con cui si possono svolgere compiti complessi.

Questi sviluppi hanno collocato i modelli linguistici al centro di un dibattito interdisciplinare che coinvolge informatici, linguisti, filosofi, giuristi, economisti e scienziati sociali. La loro capacità di interagire in linguaggio naturale li inserisce infatti direttamente nei processi comunicativi che strutturano la vita sociale. Gli LLM operano come mediatori di informazioni, strumenti creativi, assistenti professionali e, in alcuni casi, componenti di sistemi decisionali automatizzati. La loro integrazione in settori come l'istruzione, il mercato del lavoro, la salute, il supporto psicologico e la produzione mediale pone una duplice sfida. In primo luogo, emerge la necessità di una prospettiva genealogica che ricostruisca i passaggi storici e concettuali che hanno portato dalle prime sperimentazioni dell'intelligenza artificiale ai sistemi neurali contemporanei, passando attraverso la trasformazione del concetto di algoritmo e l'affermazione del machine learning come paradigma dominante (Airoldi, 2022; Woolgar, 1985). In secondo luogo, è necessario valutare le implicazioni etiche e sociali dello sviluppo di sistemi che apprendono da grandi insiemi di dati, raccolti in modo spesso opaco, incorporando rappresentazioni culturali non neutrali

che possono generare output influenti nella definizione di categorie, identità e decisioni (Noble, 2018; Barocas et al., 2013).

Questo capitolo affronta tali questioni innanzitutto ricostruendo la genealogia dell'intelligenza artificiale, con particolare attenzione al progressivo passaggio dalla programmazione simbolica ad approcci di apprendimento automatico. Si introducono alcune definizioni essenziali utili a comprendere come funzionano i modelli linguistici e quali sono i loro punti di forza e le loro limitazioni, ma soprattutto le principali sfide etico sociali che pongono, considerando la loro relazione con la produzione della conoscenza, la costruzione della verità, l'esercizio del potere, la discriminazione algoritmica, la governance dei dati e le trasformazioni nel mondo del lavoro e dell'educazione. L'obiettivo è fornire una lettura critica e accessibile per orientarsi in un campo in rapida evoluzione, giacché comprendere i modelli linguistici significa comprendere una parte significativa dei processi attraverso cui oggi si produce, circola e si legittima la conoscenza. Questo comporta interrogarsi su temi centrali per le società contemporanee: quali forme di autorità epistemica vengono ridisegnate dall'uso di sistemi generativi; quali responsabilità emergono per sviluppatori, istituzioni e utenti; quali strumenti di governance sono necessari per affrontare rischi e opportunità che attraversano l'intero ecosistema informativo.

L'avvento degli LLM come tecnologie emergenti

La storia degli LLM si inserisce in un percorso più ampio che attraversa l'evoluzione dell'intelligenza artificiale, la trasformazione del concetto di algoritmo e la crescente disponibilità di dati digitali, e questa genealogia permette di comprendere come visioni teoriche, interessi istituzionali, innovazioni tecniche e pratiche sociali abbiano contribuito a definire il tipo di intelligenza artificiale oggi associata agli LLM.

Negli anni Cinquanta e Sessanta l'IA prende forma come progetto scientifico fondato sulla convinzione che il pensiero umano potesse essere formalizzato attraverso regole logiche. A partire da questa visione, che privilegiava l'astrazione simbolica, si giunge, qualche decennio dopo, alla costruzione dei cosiddetti sistemi esperti, programmi capaci di risolvere problemi in domini specifici grazie a regole esplicitamente codificate da esperti umani. In questa fase di ambizioni teoriche elevate, l'IA veniva spesso presentata come una scienza in rapido avvicinamento alla comprensione della cognizione umana (Woolgar, 1985). A partire dagli anni Ottanta e Novanta il paradigma simbolico entra tuttavia in crisi. La complessità del mondo reale, la difficoltà di codificare regole esaustive e l'imprevedi-

bilità dei contesti applicativi rendono evidente la necessità di un approccio meno rigido. Parallelamente si assiste all'espansione delle infrastrutture digitali e alla crescente capacità dei computer di elaborare grandi quantità di dati, condizioni favorevoli allo sviluppo dei nuovi metodi di apprendimento automatico.

A partire dagli anni Duemila si consolida un paradigma basato sull'apprendimento da dati. L'idea non consiste nel programmare regole che descrivono un fenomeno, ma nel permettere ai sistemi di inferire tali regole osservando grandi quantità di esempi. È in questo contesto che si sviluppano tecniche come le reti neurali profonde, capaci di individuare schemi complessi all'interno dei dati e di adattarsi a compiti diversi tramite processi di addestramento iterativi. Il successo di questo paradigma dipende da fattori sociotecnici eterogenei: l'aumento esponenziale dei dati online, la crescita delle capacità di calcolo, la diffusione delle infrastrutture cloud e la disponibilità di grandi investimenti pubblici e privati. La ricerca sull'IA, sostenuta anche da attori militari e industriali, si orienta così verso modelli che richiedono quantità sempre maggiori di dati e potenza computazionale (Crawford, 2021).

La storia degli LLM incrocia quella della trasformazione del concetto di algoritmo, che ha assunto significati diversi in epoche diverse. Questa trasformazione, però, non riguarda soltanto il significato tecnico del termine, ma il suo ruolo sociale: gli algoritmi entrano infatti in un'ampia varietà di processi organizzativi, amministrativi e culturali, e condizionano in modo crescente la circolazione delle informazioni e le pratiche della vita quotidiana. L'emergere degli LLM va compreso dunque come esito di una lunga evoluzione dell'IA che, tramite innovazioni nel calcolo e nell'elaborazione dei dati, ha dato vita a una nuova forma di produzione di conoscenza, incorporata in infrastrutture sempre più pervasive e alla portata di pochi attori economici con grandi disponibilità di informazioni.

E se oggi i modelli vengono così impiegati, devono tuttavia il loro successo ad alcune innovazioni tecniche nel campo dell'elaborazione del linguaggio naturale, come il lavoro sulle rappresentazioni distribuite delle parole, che permette ai modelli di cogliere relazioni semantiche e sintattiche trattando le parole come vettori numerici in uno spazio matematico (Mikolov et al., 2013). La vera svolta è legata all'architettura Transformer proposta da Vaswani et al. (2017), che consente di elaborare sequenze di testo in modo più efficiente e di modellare relazioni complesse tra le parole all'interno di un contesto testuale. Questi contributi hanno reso possibile la costruzione di modelli linguistici via via più estesi, come il Bidirectional Encoder Representations from Transformers (Devlin et al., 2019) e le prime versioni della famiglia dei Generative Pretrained Transformers

(Radford et al., 2021). La loro caratteristica principale consiste nella capacità di prevedere le parole successive a partire da quelle precedenti. Da questa operazione di base derivano capacità ulteriori, tra cui la generazione di testi coerenti, il completamento di frasi, la traduzione automatica e la risposta a domande. In molti casi queste abilità emergono senza che siano state programmate in modo diretto, ma come risultato delle statistiche apprese dai dati.

Negli anni più recenti si afferma la categoria dei modelli fondazionali, ovvero sistemi addestrati su dati molto vasti e capaci di adattarsi a una grande varietà di compiti. Il loro ruolo risulta centrale, poiché fungono da base per ulteriori specializzazioni tramite tecniche di *fine tuning*. L'incremento della dimensione dei dataset e del numero di parametri dei modelli alimenta un ciclo di sviluppo che punta a espandere ulteriormente capacità generative e versatilità, allo scopo di creare modelli di rappresentazione del mondo via via più elaborati, attraverso testi, immagini, video, suoni e rappresentazioni simulate della realtà. Con il rilascio di ChatGPT nel 2022 queste tecnologie raggiungono una diffusione senza precedenti. La possibilità di interagire con un modello linguistico attraverso una semplice interfaccia conversazionale produce cambiamenti repentini: le rappresentazioni dell'IA mutano rapidamente e si diffondono narrazioni che associano ai LLM capacità cognitive avanzate o il potenziale raggiungimento di una intelligenza artificiale generale, un tema che anima tuttora un ampio dibattito scientifico (Bubeck et al., 2023). Ad oggi vi sono innumerevoli implementazioni di LLM in interfacce utente sia dedicate, che di supporto ad altri servizi o a specifiche attività lavorative, e tutte queste istanze sono figlie della ribalta dell'IA come leva tecnologica dell'innovazione, dell'efficienza e della semplificazione delle attività umane. Una tecnologia fortemente umanizzata, spesso incompresa, talvolta anche pericolosa.

Che cosa è un LLM e cosa fa

Comprendere gli LLM richiede un vocabolario essenziale e alcuni concetti fondamentali. Per poter cogliere i principi di funzionamento che ne guidano il comportamento serve chiarire quali operazioni compiono, quali forme di apprendimento utilizzano e come interagiscono con i dati che costituiscono la loro base di conoscenza.

Innanzitutto, un LLM è un modello linguistico, ossia un modello statistico capace di elaborare testo in linguaggio naturale. L'obiettivo primario di questi modelli consiste nella previsione di una parola (o meglio, di una sua porzione, detta *token*) successiva a partire dal contesto precedente,

ossia dalle parole che la precedono in un testo. Gli LLM odierni ampliano questa logica di base fino a generare testi articolati e coerenti, rispondere a domande, sintetizzare documenti o svolgere attività conversazionali. Queste capacità “emergono”, senza essere direttamente impartite, poiché il modello impara a riconoscere regolarità, associazioni, stili e strutture all’interno dei dati su cui è addestrato, e li utilizza per produrre nuove sequenze testuali coerenti con i compiti richiesti.

Il funzionamento degli LLM deriva principalmente da un processo di addestramento condotto su grandi basi di dati testuali. Il modello viene esposto a miliardi di parole provenienti da libri, articoli, siti web e altre fonti digitalizzate, e a partire da questi esempi sviluppa una rappresentazione matematica del linguaggio. Le parole/token non vengono trattate come unità isolate, ma come vettori in uno spazio matematico che cattura somiglianze e differenze, un approccio introdotto dai lavori sulla rappresentazione distribuita del significato (Mikolov et al., 2013) e che, in estrema sintesi, cerca di comprendere il senso di un termine a partire dai raggruppamenti di termini a cui si accompagna più di frequente. La procedura di addestramento modifica progressivamente i parametri del modello per ridurre quanto più possibile l’errore di previsione e genera una struttura complessa che incorpora relazioni lessicali, sintattiche e semantiche. Durante il processo, perciò, non viene fornito un insieme di regole esplicite che descrivono come funziona il linguaggio: l’efficacia del modello deriva piuttosto dalla capacità di interiorizzare schemi ricorrenti nelle produzioni linguistiche umane, comprese le loro ambiguità e irregolarità.

Gli LLM moderni vengono generalmente realizzati attraverso più fasi di addestramento. La prima fase è il pre-addestramento, in cui il modello apprende una conoscenza generale del linguaggio. La qualità e la diversità dei dati disponibili in questa fase determinano una parte significativa delle capacità finali del modello. La seconda fase è il fine tuning. In questo passaggio il modello viene “raffinato” e specializzato per compiti più specifici, come l’assistenza a clienti di un’azienda, la redazione di testi scientifici o la generazione di codice informatico, e soprattutto lo rendono più adeguato all’interfaccia con gli esseri umani basata su turni di interlocazione, come avviene nelle chat. Le tecniche di fine tuning consentono di produrre un modello fondazionale e di adattarlo a contesti applicativi molto diversi, riducendo la necessità di addestramenti separati per ogni compito.

Una delle innovazioni recenti più influenti nella costruzione di modelli come ChatGPT è l’addestramento tramite rinforzo con feedback umano (RLHF). Questo approccio mira a rendere il comportamento del modello più coerente con aspettative sociali e norme culturali condivise (Christiano et al., 2017). Per fare ciò vengono raccolte valutazioni da parte di indivi-

dui reclutati per giudicare la qualità delle risposte prodotte dal modello. Il sistema apprende così a preferire risposte considerate più chiare, più utili o più eticamente accettabili. Il risultato è un comportamento conversazionale più controllato, che privilegia forme discorsive adeguate al contesto di utilizzo e riduce la probabilità di output problematici. Questo processo ha conseguenze significative, perché introduce nel modello rappresentazioni normative che derivano direttamente dalle scelte e dalle interpretazioni dei valutatori umani.

Grazie a queste integrazioni, però, gli LLM possono svolgere un ampio ventaglio di attività. Tra queste rientrano la generazione di testi lunghi e coerenti, la traduzione automatica, la spiegazione di concetti complessi, la sintesi di documenti e la risposta a domande aperte. In alcuni casi l'ampiezza e la precisione delle loro prestazioni suggeriscono un comportamento che appare simile al ragionamento astratto o alla comprensione semantica, seppure sia questione di dibattito la validità della rilevazione di questi comportamenti come indicatori di una effettiva comprensione o consapevolezza dei compiti trattati, per via di altrettanti comportamenti che sarebbero altrimenti assenti. Tra i più rilevanti si collocano le cosiddette allucinazioni, ovvero risposte che includono informazioni plausibili, ma inventate. Questo fenomeno è documentato anche in contesti critici, come la citazione di sentenze giuridiche mai esistite o la creazione di riferimenti bibliografici falsi o inesatti. La tendenza del modello a generare contenuti verosimili ma non verificati deriva dal fatto che il suo compito primario consiste nella previsione linguistica, non nella consultazione diretta di fonti esterne. Ad oggi esistono forme di mitigazione di questo fenomeno, che però può comunque celarsi anche negli ambienti più controllati di gestione della conoscenza.

La fortuna degli LLM è legata anche al modo in cui questi si presentano agli utenti. Le interfacce conversazionali eliminano gran parte della complessità tecnica, rendendo possibile l'uso di modelli sofisticati attraverso richieste in linguaggio naturale. Questa scelta progettuale permette una forma di interazione più intuitiva, ma produce anche alcune tensioni. In particolare, l'aspetto dialogico può favorire proiezioni antropomorfe che incidono sul modo in cui gli utenti interpretano le risposte, attribuiscono autorità ai contenuti o stabiliscono forme di fiducia nei confronti del sistema (Kieras, 1990; Crawford, 2021). Le interfacce determinano, dunque, non solo come il modello viene utilizzato, ma anche come viene percepito. La conversazione continua, la capacità di mantenere coerenza contestuale e l'apparente empatia contribuiscono a configurare una relazione comunicativa che può influenzare decisioni, comportamenti e aspettative, aspetti cruciali nella riflessione sulle loro implicazioni etiche e sociali.

Il trattamento dei dati e gli LLM

Gli LLM dipendono in modo strutturale dalle basi di dati utilizzate per il loro addestramento. La qualità, l'origine, la composizione e le modalità di raccolta di questi dati influenzano in modo diretto le prestazioni del modello, ma anche le visioni del mondo che esso incorpora e riproduce. Per questa ragione la sociologia dell'IA considera i dataset come oggetti centrali di analisi, legati a processi di estrazione, classificazione e standardizzazione che sono sempre situati in contesti materiali e culturali specifici. Le basi di dati non rappresentano una semplice collezione neutrale di testi, e sono piuttosto il risultato di scelte, opportunità e vincoli che riflettono strutture sociali, valori e rapporti di potere. Come osserva Crawford (2021), gli artefatti del mondo vengono trasformati in dati attraverso pratiche di estrazione, misurazione e categorizzazione che definiscono ciò che può essere trattato come informazione computabile. Questo processo produce una forma di oggettività apparente che nasconde le condizioni e le asimmetrie da cui i dati derivano. La corsa alla produzione di LLM fa poi sì che molte basi di dati utilizzate oggi siano costruite a partire da materiali raccolti negli anni passati, spesso senza procedure chiare di valutazione della qualità, della rappresentatività o del rispetto della privacy (Joyce et al., 2021). L'urgenza di disporre di grandi quantità di dati ha spesso prevalso su considerazioni deontologiche o sui criteri di rigore altrimenti preminenti nella produzione di tecnologie dal profilo di rischio elevato, nel cui novero possono ora ritrovarsi gli LLM. E tali scelte di convenienza non sono che alcune delle numerose che vengono prese nell'addestrare modelli di dimensioni man mano più grandi.

Airoidi (2022) descrive la relazione tra dati e modelli attraverso l'espressione «cultura nel codice»: le aspettative, le categorie e le pratiche degli esseri umani che raccolgono, selezionano e annotano i dati vengono incorporate nelle strutture algoritmiche. Nei sistemi di apprendimento supervisionato questa influenza è evidente: la classificazione umana costituisce il punto di partenza per l'apprendimento della macchina. Con l'apprendimento non supervisionato, che caratterizza gran parte dell'addestramento di un LLM, tale influenza è meno esplicita ma altrettanto sostanziale, perché i modelli apprendono da testi prodotti da comunità specifiche, storicamente favorite nella produzione e nella circolazione di documenti, senza che vi sia scrutinio umano sulla stragrande maggioranza del materiale impiegato. Di conseguenza, gli LLM tendono a riflettere e riprodurre disuguaglianze nella rappresentazione del genere, delle etnicità, della classe sociale e delle identità minoritarie. Le relazioni statistiche apprese dai dati favoriscono ciò che è più frequente e meglio documentato nelle fonti di-

sponibili, mentre rappresentano in modo distorto ciò che è meno visibile o meno presente in quelle fonti.

Oltretutto, la relazione tra dati e modelli è bidirezionale: i modelli, una volta dispiegati, producono output che influenzano il modo in cui le persone cercano informazioni, costruiscono giudizi e prendono decisioni; il “codice nella cultura”. Le distorsioni presenti nei dati non rimangono confinate all’interno dei sistemi, ma circolano attraverso gli usi sociali delle tecnologie che li incorporano. Nel caso degli LLM, questi divengono produttori di dati sintetici, ad uso della successiva generazione di modelli, ponendosi come mediatori culturali primari nell’addestramento dei sistemi successivi e depositari di specifiche rappresentazioni, che appaiono nei testi generati. Gli LLM contribuiscono così a stabilizzare categorie non necessariamente neutre: Crawford (2021) osserva che classificazioni apparentemente tecniche possono consolidare idee di genere o di razza basate su semplificazioni storicamente radicate, e questo processo assume un effetto normativo perché i sistemi automatici vengono spesso percepiti come più oggettivi di quanto non siano effettivamente, soprattutto quando operano all’interno di servizi quotidiani come motori di ricerca, piattaforme digitali o sistemi di raccomandazione.

L’autorità talvolta accreditata a queste tecnologie si accompagna alla difficoltà nel comprendere come i modelli linguistici elaborino e trasformino i dati, un tema centrale nella sociologia dell’IA. Pasquale (2015) descrive i sistemi algoritmici come “scatole nere” quando le loro logiche operative risultano opache, e Christin (2020) identifica tre fonti principali di tale opacità: la segretezza intenzionale da parte delle aziende, l’analfabetismo tecnico che limita la capacità di molti attori di interpretare il codice e la complessità intrinseca dei sistemi di apprendimento automatico, i cui parametri evolvono attraverso processi non sempre interpretabili. Woolgar (1985) evidenziava già ai tempi dei sistemi esperti la difficoltà di definire in modo univoco il comportamento dei sistemi di IA, e con gli LLM questa difficoltà assume una dimensione nuova, perché la quantità di dati e la scala dei modelli sanciscono l’impossibilità di ricostruire con precisione come le informazioni vengano trasformate.

Le caratteristiche dei dataset, dunque, intervengono direttamente sulla qualità delle decisioni e dei contenuti generati dagli LLM, e se i dati sono sbilanciati, i modelli potranno riprodurre tale sbilanciamento, seppur mitigato da interventi di fine-tuning, creando effetti discriminatori quando vengono richieste, per esempio, descrizioni di professioni, identità sociali o caratteristiche personali, oppure quando l’utente si affida al modello per ottenere valutazioni o suggerimenti. Le discussioni sulla governance dei dati si concentrano quindi su questioni come la selezione delle fonti, la tra-

sparenza delle procedure, il riconoscimento dei diritti degli autori e la tutela delle persone rappresentate nei dataset. In assenza di politiche robuste e puntuali, la raccolta intensiva di dati diviene un processo estrattivo che produce vantaggi economici per pochi attori e costi sociali per molti altri (Crawford, 2021).

Analizzare le basi di dati come oggetti sociologici implica riconoscere che, come appaiono oggi, esse rappresentano una forma organizzata di conoscenza sociale, modellata da interessi istituzionali, da infrastrutture materiali e da pratiche concentrate in luoghi specifici. La produzione dei dataset seleziona, ordina e talvolta scarta aspetti della realtà, e gli LLM ereditano tanto le potenzialità quanto le limitazioni e le distorsioni dei dati su cui si basano. La loro capacità di generare testi credibili rende ancora più importante interrogarsi su come i dati sono stati raccolti, quali soggetti sono rappresentati in modo adeguato e quali invece risultano marginalizzati.

D'altronde, gli LLM non operano in isolamento. Fanno parte di sistemi articolati che coinvolgono infrastrutture, istituzioni, pratiche professionali, attori economici, culture tecniche e aspettative sociali. Osservarli come oggetti sociotecnici consente di analizzarne il funzionamento al di là della loro dimensione algoritmica, mettendo in evidenza le reti di relazioni e gli assemblaggi che li sostengono e li modellano (Aragona, Felaco, 2018). Il campo dei Critical Algorithm Studies (CAS) muove dall'intenzione di considerare gli algoritmi come parti integrate di sistemi più ampi, dove elementi tecnici e sociali sono inestricabili (Joyce et al., 2021). Gli LLM incarnano questa interdipendenza, dato che la loro esistenza dipende da infrastrutture materialmente estese, come data center, processori specializzati, reti di telecomunicazione e piattaforme cloud, ma anche da risorse immateriali come basi di dati, competenze specialistiche, reti di ricerca e investimenti su larga scala. Le aziende che sviluppano questi modelli svolgono un ruolo decisivo: non solo stabiliscono priorità di ricerca e strategie commerciali, ma determinano anche cosa debba essere considerato un uso corretto del modello. In questo senso la governance privata influenza costantemente le condizioni di accesso, le regole di utilizzo e gli orientamenti etici incorporati nei modelli linguistici. Oltretutto, gli LLM si inseriscono in un ecosistema economico altamente concentrato e consolidato, poiché lo sviluppo e il funzionamento di modelli di grandi dimensioni richiedono investimenti sostanziali e accesso a hardware avanzato, risorse alla portata di ben poche realtà private. Il rapido progresso e la diffusione capillare di queste tecnologie, assieme alla poca reattività dell'attività legislativa e, talvolta, alla resistenza nei riguardi di politiche regolatorie rigide, che potrebbero fiaccare lo sviluppo tecnologico, si sostanzia in un triangolo di

relazioni tra imprese, istituzioni e utenti in cui il potere decisionale tende a concentrarsi nella parte privata, mentre gli effetti sociali si distribuiscono in modo più diffuso.

Gli LLM come classificatori e decisori nel mondo sociale

Gli LLM partecipano a forme di potere algoritmico che si manifestano a vari livelli, intensificando l'automazione delle procedure di classificazione tipiche delle interfacce digitali. Molti sistemi automatici sono progettati per classificare, e le categorie utilizzate non sono mai neutrali, perché riflettono visioni del mondo, priorità e convenzioni culturali (Crawford, 2021). Quando un modello linguistico genera descrizioni di gruppi sociali o ruoli professionali, lo fa a partire dai dati con cui è stato addestrato, e contribuisce a stabilizzare immagini stereotipate che possono riprodurre o amplificare disuguaglianze esistenti. Barocas et al. (2013) mostrano come l'analisi automatizzata dei dati possa influenzare l'accesso a risorse, servizi e opportunità, e nel caso dei modelli linguistici queste influenze riguardano anche la circolazione delle informazioni e la definizione di categorie sociali. Nelle piattaforme digitali, ad esempio, gli algoritmi di raccomandazione agiscono come filtri che selezionano contenuti e organizzano l'esperienza degli utenti. Anche queste selezioni possono rafforzare stereotipi e squilibri di potere, e gli LLM, integrandosi in motori di ricerca, assistenti digitali e strumenti di comunicazione, ampliano ulteriormente la portata di tali processi. La classificazione algoritmica assume un ruolo particolarmente delicato nei contesti decisionali ad alto impatto. La letteratura mostra come sistemi di valutazione automatica possano penalizzare gruppi minoritari a causa di bias incorporati nei dati o nelle procedure di addestramento (Barocas et al., 2013; Noble, 2018).

L'ampia adozione degli LLM è legata, come si diceva, alla loro capacità di interagire tramite linguaggio naturale, che rende possibile un utilizzo immediato da parte di un gran numero di utenti, senza bisogno di particolari competenze tecniche. Questo tipo di interfaccia crea una relazione particolare, e per certi versi inedita, tra gli utenti e il modello, per cui specifici artifici retorici e stili di conversazione possono avere grande rilievo nell'utilizzo e nella soddisfazione degli utilizzatori. Queste interazioni dipendono perciò sia dal modo in cui l'interfaccia è progettata che dalle aspettative dell'utente e dal contesto d'uso. Ad esempio, un assistente digitale che risponde con cortesia e tono empatico può indurre un utente a considerare più affidabili le informazioni così fornite. Al contrario, un

modello troppo rigido nelle sue restrizioni può generare frustrazione o favorire strategie per aggirarne i filtri, fenomeno osservabile nei tentativi di eludere le limitazioni introdotte tramite RLHF. Al contempo, in una dinamica di mutuo adattamento, tanto gli utenti quanto i modelli stessi, e i loro produttori, possono evolvere e condizionare nel tempo le forme di queste interazioni.

Infine, va ricordato quanto spesso gli LLM oggi intervengano nelle pratiche attraverso cui le informazioni vengono cercate, interpretate e sintetizzate. Quando gli utenti si rivolgono a un modello linguistico per ottenere spiegazioni, chiarimenti o soluzioni, attribuiscono implicitamente un ruolo epistemico al sistema, grazie alla capacità del modello di riorganizzare informazioni che provengono dai dati di addestramento, ma che non sempre è corroborato da un accesso chiaro o veridico alle fonti. La presenza degli LLM nei processi conoscitivi introduce una forma di intermediazione che modifica i criteri con cui si valuta la veridicità delle affermazioni e la qualità delle argomentazioni. In alcuni casi questa intermediazione aumenta l'efficienza o facilita l'accesso a contenuti complessi. In altri casi rischia di diffondere informazioni non verificate o di sostituire competenze esperte con risultati generati automaticamente. Tali dinamiche occupano un posto centrale nelle discussioni sulle implicazioni etiche e sui modelli di governance dell'IA.

Conclusioni

L'espansione dei modelli linguistici di grandi dimensioni ha inaugurato una fase in cui l'intelligenza artificiale entra in modo diretto nei processi comunicativi, cognitivi e organizzativi della vita associata. La genealogia degli LLM mostra come queste tecnologie derivino da una lunga storia di idee, innovazioni e pratiche che attraversano la matematica, l'informatica, le scienze cognitive, l'organizzazione delle infrastrutture digitali e l'economia delle piattaforme. Le definizioni e i principi di funzionamento riportati chiariscono come i modelli linguistici apprendano schemi ricorrenti dai dati, come si specializzino attraverso il fine tuning e come vengano orientati tramite processi di allineamento che incorporano valori e interpretazioni umane.

L'analisi delle basi di dati mette in evidenza che i modelli linguistici non sono semplici strumenti di calcolo. Sono il risultato di processi di estrazione, selezione e classificazione che trasformano fenomeni sociali in materiale computazionale. L'attenzione della letteratura dei CAS ai rap-

porti tra dati, potere e rappresentazione mostra come gli LLM ereditino e a loro volta riproducano le strutture sociali da cui originano. La questione dei dati non riguarda solo la qualità tecnica dei modelli, ma la loro capacità di intervenire nelle immagini del mondo che circolano negli ambienti informativi contemporanei. Considerare gli LLM come oggetti sociotecnici permette di coglierne la collocazione dentro reti che integrano infrastrutture materiali, attori economici, norme culturali, pratiche professionali e modelli di governance. Le tecnologie generative sono parte di assemblaggi che distribuiscono competenze, responsabilità e potere. Le loro prestazioni non possono essere comprese senza interrogare le condizioni di produzione, gli interessi coinvolti e i contesti applicativi. La letteratura sugli algoritmi come sistemi sociali (Aragona, Felaco, 2018; Joyce et al., 2021) fornisce strumenti utili per interpretare questi fenomeni.

Gli LLM intervengono nei processi di costruzione della verità e dell'autorità epistemica, trasformano l'accesso alla conoscenza, modificano le mansioni lavorative e ridefiniscono obiettivi e pratiche educative. L'efficacia comunicativa dei modelli genera nuove forme di dipendenza informativa e solleva problemi relativi alle competenze critiche degli utenti. Allo stesso tempo crea opportunità in termini di accessibilità, creatività ed efficienza. Le istituzioni formative e professionali sono chiamate a ripensare modelli di apprendimento e di valutazione che tengano conto della presenza costante di strumenti generativi. Le sfide etiche e di governance poste dagli LLM non possono essere affrontate attraverso soluzioni meramente tecniche. La trasparenza e l'accountability richiedono strumenti normativi e organizzativi che tengano conto dell'opacità strutturale dei modelli (Burrell, 2016). La questione dell'equità comporta profonde riflessioni su come prevenire e mitigare bias che derivano da strutture sociali pregresse e da decisioni prese in fase di sviluppo (Noble, 2018). Espansione delle forme di sorveglianza e concentrazione del potere sono aspetti particolarmente rilevanti della diffusione dei LLM, e rievocano critiche più ampie alla distribuzione dei benefici e dei rischi del capitalismo digitale (Zuboff, 2019), nonché sulla sostenibilità materiale ed energetica dei suoi processi.

In prospettiva, il futuro degli LLM dipenderà dalla capacità di sviluppare forme di governance multilivello che coinvolgano ricercatori, aziende, istituzioni pubbliche, comunità professionali e cittadini. Come suggerisce Floridi (2022), l'etica dell'IA richiede un approccio che integri responsabilità progettuale, vigilanza istituzionale e forme intense di partecipazione, nell'auspicio di poter definire standard condivisi, promuovere audit indipendenti, sostenere modelli aperti e trasparenti, proteggere i diritti delle

persone rappresentate nei dati e valorizzare il ruolo delle competenze umane nei processi decisionali. Tuttavia, molte delle caratteristiche discusse rendono tali obiettivi tanto più ardui quanto sono desiderabili, e però essi animano lo studio dei modelli linguistici, che offre un'occasione per interrogare il rapporto tra umani e macchine, tra conoscenza e informazione, tra tecnologia e potere. Le scienze sociali possono contribuire a questa interpretazione fornendo categorie, metodi e prospettive che rendono visibili le condizioni che sostengono e orientano il funzionamento di queste TE.

4. Sistemi di decisione automatizzata

di *Francesco Amato*

Verso una definizione dei sistemi di decisione automatizzata

I sistemi di decisione automatizzata (SDA) costituiscono un insieme di tecnologie emergenti la cui visibilità rimane limitata rispetto a quella assunta, nel dibattito pubblico, da applicazioni ormai consolidate dell'intelligenza artificiale, come i modelli linguistici di grandi dimensioni (LLM) e gli algoritmi di raccomandazione. Questi sistemi automatizzano i processi decisionali elaborando dati secondo regole, modelli o algoritmi predefiniti che ne determinano i criteri di scelta e di valutazione. In tal modo, il processo decisionale viene trasferito, in tutto o in parte, da operatori umani a procedure computazionali standardizzate. I SDA sono utilizzati in vari ambiti della società, tra cui l'assistenza sanitaria (Grote, Berens, 2020), l'approvazione dei prestiti (Agarwal et al., 2020), il reclutamento (Mariani, Vega-Lozada, 2023) e la loro applicazione è in continua espansione. Le decisioni automatizzate sono sempre più spesso integrate nelle tecnologie emergenti (Carter, Egliston, 2021), nei dispositivi dell'Internet delle Cose (Mostafa et al., 2019) e nei processi offline che diventano digitalizzati (McAfee, Brynjolfsson, 2018). Applicazioni concrete dei SDA riguardano i veicoli a guida autonoma che stanno rivoluzionando il settore dei trasporti, promettendo maggiore sicurezza ed efficienza, ma sollevando, al contempo, questioni etiche sull'attribuzione di responsabilità in caso di incidenti (Goodall, 2014). I robot che automatizzano il lavoro di cura e assistenza, che contribuiscono ad alleviare la carenza di operatori sanitari, ma che sollevano interrogativi etici sul ruolo umano nella cura dei più vulnerabili (Sharkey, 2014). Le tecnologie che abilitano nuove forme di apprendimento e organizzazione del lavoro che offrono possibilità senza precedenti di personalizzazione e flessibilità, ma che potrebbero anche incidere sulla precarietà del lavoro e la sorveglianza dei dipendenti (Zuboff, 2019). I dispositivi

per il monitoraggio in tempo reale delle condizioni di salute, come gli smartwatch e altri sensori biometrici, che stanno trasformando la medicina preventiva e il monitoraggio dei pazienti, ma sollevano preoccupazioni sulla privacy e la sicurezza dei dati personali (Meskó et al., 2017). I sensori che governano i sistemi di trasporto locale e che consentono una gestione più efficiente del traffico e dei servizi pubblici, evidenziando la sempre più stretta integrazione tra infrastrutture fisiche e digitali, ma che richiedono sistemi di sicurezza robusti per prevenire gli attacchi informatici (Hahn et al., 2019).

Questi esempi mostrano come l'innovazione tecnologica e il crescente impiego di tali sistemi nella società contemporanea stiano generando nuove opportunità, favorendo progressi significativi in numerosi ambiti della vita quotidiana. Allo stesso tempo, tuttavia, essi introducono nuove sfide e rischi sociali, tra cui il cosiddetto rischio algoritmico (Aragona, Amato, 2022), inteso come la potenziale perdita di controllo umano e responsabilità su processi decisionali automatizzati che, a causa della loro natura opaca e della capacità di agire autonomamente, possono produrre esiti discriminatori, manipolazioni sociali o asimmetrie di potere non ispezionabili. Tale scenario solleva nodi etici imprescindibili, alimentando il dibattito pubblico e scientifico sulle implicazioni di questi sistemi in relazione a principi quali trasparenza, equità e responsabilità (Richardson, 2021).

Uno sguardo approfondito alla struttura e al funzionamento dei sistemi di decisione automatizzata (SDA) mette in evidenza la loro natura complessa e stratificata. Si tratta di sistemi il cui operato non è sempre immediatamente percepibile. Mentre i soggetti coinvolti nella loro progettazione, sviluppo e implementazione ne comprendono in genere le logiche operative, altri attori, come i partner o gli utenti istituzionali, ad esempio i clienti delle organizzazioni che li impiegano, possono non essere pienamente consapevoli né della loro presenza né delle modalità con cui operano. Questa distanza cognitiva è ancora più marcata per gli utenti finali e per il pubblico generale, che spesso mancano delle competenze necessarie per riconoscere o interpretare i processi decisionali automatizzati incorporati nelle tecnologie di uso quotidiano. Come osservato in letteratura (Gutierrez Lopez, Halford, 2024), la visibilità dei SDA varia in base al punto di osservazione. Per molti individui che subiscono le conseguenze delle decisioni automatizzate senza interagire direttamente con tali sistemi, essi appaiono di fatto invisibili. Il funzionamento dei processi decisionali automatizzati tende infatti a rimanere silenzioso, operando in background senza richiedere attenzione o intervento umano diretto (Pasquale, 2015; O'Neil, 2016; Noble, 2018).

Diversi studi hanno sottolineato che questi sistemi diventano riconoscibili solo quando producono effetti negativi o risultati inattesi (Pasquale, 2015). Tale emersione improvvisa può essere paragonata al cosiddetto effetto glitch, ovvero quell'interruzione che disturba l'esperienza dell'utente e lo costringe a prendere coscienza della natura digitale del sistema con cui sta interagendo (Manon, Temkin, 2011). Per comprendere appieno tale dinamica di impercettibilità, è necessario ripercorrere l'evoluzione storica che ha condotto alla nascita dei sistemi di decisione automatizzata.

Le origini dei sistemi di decisione automatizzata possono essere ricondotte ai primi processi di automazione informatica sviluppatasi negli anni Cinquanta, in parallelo con l'affermazione dell'informatica come campo scientifico autonomo e con la diffusione, tra gli esperti, delle prime idee relative all'intelligenza artificiale (McCarthy et al., 2006). In questa fase, i sistemi automatizzati venivano progettati per eseguire compiti ripetitivi e monotoni, ridurre il carico cognitivo e liberare gli esseri umani da attività considerate alienanti, alimentando l'entusiasmo tecnologico per le nuove possibilità computazionali e per le promesse di innovazione che l'automazione sembrava offrire.

Una svolta significativa si ebbe nel 1956, quando la RAND Corporation, centro di ricerca strategica statunitense fondato nel 1946 con il sostegno del Dipartimento della Difesa, istituì la System Development Corporation (SDC) per la progettazione di software destinato al sistema Semi-Automatic Ground Environment (SAGE) (Campbell-Kelly, 2003). SAGE rappresentava un avanzato sistema automatizzato di difesa aerea per il rilevamento, il tracciamento e l'intercettazione di velivoli nemici nell'ambito del North American Aerospace Defense Command (NORAD). La SDC fu una delle prime organizzazioni a occuparsi della progettazione di software automatizzati, collaborando con il governo degli Stati Uniti e con l'Advanced Research Projects Agency (ARPA) e aprendo così la strada alle prime forme di automazione decisionale su larga scala.

Studi recenti hanno mostrato che tali sistemi vengano percepiti come meno soggetti a bias o giudizi di valore (Luciani, 2018) e ritenuti più affidabili rispetto ai processi decisionali umani, spesso considerati parziali o incoerenti (Castelo et al., 2019). La loro adozione è stata inoltre favorita da dinamiche economiche e organizzative, come la crescente centralità dei dati e le pressioni competitive dei mercati, che hanno spinto le organizzazioni a ricorrere all'automazione per migliorare efficienza e scalabilità (Bergemann, Bonatti, 2024). In questo quadro, l'implementazione dei SDA si è progressivamente estesa fino a diventare una componente strutturale delle tecnologie digitali contemporanee (Mökander, Axente, 2023), dove essi

possono operare in modo autonomo o contribuire attivamente al processo decisionale insieme agli esseri umani (Davenport, Harris, 2005).

Ad oggi i sistemi di decisione automatizzata non possiedono ancora una definizione univoca né un riconoscimento terminologico consolidato nella letteratura scientifica. Questa incertezza riflette la natura eterogenea e in continua evoluzione di tali sistemi e della ricerca scientifica che li riguarda, in cui il concetto di automazione decisionale si intreccia non solo con l'intelligenza artificiale e il machine learning, ma anche con campi quali l'analisi dei dati, la statistica computazionale, la teoria delle decisioni e le scienze cognitive. L'intersezione si estende inoltre a domini applicativi e normativi, come la governance algoritmica, l'etica delle tecnologie, il diritto della responsabilità automatizzata e gli studi sull'organizzazione e sulla gestione dei processi decisionali. In questo quadro, gli SDA emergono come oggetti sociotecnici complessi, situati all'incrocio tra innovazione tecnica, regolazione istituzionale e pratiche di decisione umana e automatizzata.

L'assenza di una definizione condivisa rende complessa la delimitazione di un corpus di studi specificamente dedicati agli SDA, contribuendo alla frammentazione del dibattito scientifico e all'emergere di interpretazioni diverse, talvolta sovrapposte. Tale ambiguità, tuttavia, è indicativa del carattere ancora emergente di questo dominio di ricerca e della necessità di definizioni più rigorose, capaci di chiarirne i confini epistemologici e applicativi (Mittelstadt et al., 2016).

In termini operativi, i sistemi di decisione automatizzata possono essere intesi come tecnologie che impiegano dati, modelli e algoritmi per incidere, in modo autonomo o congiunto con gli esseri umani, sui processi decisionali. Il loro intervento può assumere forme diverse, dall'esecuzione diretta di decisioni alla formulazione di raccomandazioni o valutazioni, ma in ogni caso contribuisce a orientare l'esito del processo decisionale. Tale capacità non dipende soltanto dal grado formale di autonomia tecnica, bensì dalle modalità con cui i sistemi vengono progettati, impiegati e interpretati nel contesto sociale e organizzativo (Pasquale, 2015; O'Neil, 2016; Noble, 2018; Zuboff, 2019).

Comprendere come un SDA operi significa analizzare congiuntamente le sue dimensioni tecniche e sociali. Gli aspetti tecnici riguardano la progettazione, i dati utilizzati, le logiche algoritmiche e i meccanismi di esecuzione delle decisioni; quelli sociali e organizzativi riguardano invece l'uso che ne fanno gli utenti, il grado di fiducia accordato, la distribuzione delle responsabilità e le implicazioni economiche, etiche, giuridiche, politiche e sociali che ne derivano.

Stabilire se e in quale misura un sistema prenda decisioni non è dunque una questione puramente tecnica, ma richiede di considerare le relazio-

ni tra tecnologia, utenti e contesti d'uso, mettendo in luce la natura intrinsecamente sociotecnica dei processi decisionali automatizzati.

Apparati e caratteristiche dei sistemi di decisione automatizzata

Alla luce di questa complessità, un sistema di decisione automatizzata può essere descritto come un insieme integrato di componenti digitali, organizzative e cognitive finalizzate a rendere automatici, parzialmente o totalmente, i processi di scelta e di valutazione. Tali sistemi si fondano sull'elaborazione di dati attraverso algoritmi o modelli di intelligenza artificiale capaci di generare decisioni o raccomandazioni in contesti specifici, dalla gestione amministrativa ai servizi pubblici, dalla finanza alla sanità. La loro operatività si sviluppa all'interno di infrastrutture più ampie, dove interagiscono con utenti, istituzioni e altri attori sociali, incidendo sui processi di governance e sulla distribuzione del potere decisionale. Per comprendere in modo analitico la configurazione di un SDA è utile scomporlo nei suoi principali apparati costitutivi, che rappresentano le dimensioni attraverso cui si articola la sua materialità e il suo funzionamento:

- infrastruttura tecnologica;
- obiettivi e prestazioni;
- governance.

L'infrastruttura tecnologica comprende l'insieme di componenti hardware, software, strumenti informatici e aspetti infrastrutturali che consentono il funzionamento del sistema (Tanenbaum, 2009). È la dimensione materiale e computazionale dell'SDA, che include elementi fisici come server per l'archiviazione e l'elaborazione dei dati o sensori che raccolgono informazioni in tempo reale, ad esempio nei terminali per l'acquisizione di dati biometrici e socioeconomici volti all'assegnazione automatizzata di sussidi o servizi sociali. Le componenti software comprendono sistemi operativi, algoritmi specializzati e programmi applicativi progettati per specifiche funzioni, come l'analisi predittiva in ambito sanitario o la gestione decisionale nelle amministrazioni pubbliche. A questi si affiancano strumenti informatici di supporto, come sistemi di controllo versione per la gestione del codice, piattaforme API per l'integrazione di dati esterni e servizi cloud per l'elaborazione scalabile. L'infrastruttura comprende inoltre configurazioni di rete, flussi di dati e sistemi di archiviazione distribuita o reti private virtuali (VPN), che garantiscono la sicurezza e l'efficienza nella trasmissione e gestione delle informazioni. Insieme, questi elementi costituiscono l'ossatura tecnologica dell'SDA, assicurandone la funzionalità e l'adattabilità ai diversi contesti operativi.

L'apparato degli obiettivi e delle prestazioni riguarda gli indicatori di efficacia, affidabilità, scalabilità e performance del sistema, nonché gli obiettivi tecnologici e finanziari che ne orientano lo sviluppo (Kaplan, Norton, 1992; Osterwalder, Pigneur, 2010; Osterwalder et al., 2015). Questa dimensione consente di comprendere come viene valutata la bontà di un SDA e quali benefici produce per gli utenti finali. Gli indicatori di successo possono includere, ad esempio, tassi di accuratezza nelle decisioni, come l'affidabilità nel rilevamento di frodi in ambito finanziario. L'affidabilità è misurata attraverso parametri quali la continuità di funzionamento o la tolleranza all'errore, particolarmente cruciali in ambiti come la sanità, dove l'interruzione del servizio può avere conseguenze gravi sulla salute dei pazienti. La scalabilità si valuta invece in relazione alla capacità del sistema di gestire carichi crescenti, ad esempio l'aumento del numero di domande di prestito in un sistema bancario, senza compromettere le prestazioni. La performance riguarda la velocità e l'efficienza del sistema, come nel caso di un SDA per la gestione del traffico in tempo reale. Gli obiettivi tecnologici possono includere l'integrazione di nuovi modelli di intelligenza artificiale o l'adeguamento a standard di sicurezza e protezione dei dati, mentre gli obiettivi economici possono mirare alla riduzione dei costi operativi o all'automazione di compiti ripetitivi. Insieme, queste componenti forniscono una cornice per valutare se l'SDA raggiunge i risultati attesi e genera valore misurabile per i soggetti coinvolti.

L'apparato della governance comprende le dimensioni relative alla proprietà e agli aspetti economici del sistema, alle collaborazioni con gli attori coinvolti, alle politiche aziendali, alla conformità normativa e alla gestione delle risorse umane e non umane (Ahmad et al., 2020; Gal, Aviv, 2020). Include inoltre la sfera delle competenze e delle capacità richieste per la produzione, l'uso e il mantenimento delle tecnologie. La governance descrive quindi la rete di attori e relazioni che accompagna la progettazione, lo sviluppo e l'implementazione di un SDA, evidenziando le interdipendenze e le responsabilità che ne derivano. La proprietà, ad esempio, può variare tra sistemi sviluppati internamente e soluzioni fornite da terzi, con conseguenze rilevanti sul controllo dei dati e sulle politiche di sicurezza. Le collaborazioni tra soggetti pubblici e privati, come quelle tra amministrazioni e imprese tecnologiche, risultano centrali per adattare gli SDA alle specifiche esigenze di contesto, come nel caso dei sistemi di gestione automatizzata dei servizi sociali. La conformità normativa costituisce un ulteriore aspetto essenziale: i sistemi devono rispettare quadri regolatori come il GDPR o le linee guida europee sulla trasparenza algoritmica. La gestione delle risorse comprende la distribuzione di competenze e ruoli,

come l'impiego di esperti di machine learning, etica e user experience per garantire un equilibrio tra efficienza tecnica e responsabilità sociale.

I tre apparati, infrastruttura tecnologica, obiettivi e prestazioni, governance, sono profondamente interconnessi nelle fasi di progettazione, sviluppo e implementazione dei sistemi di decisione automatizzata. Ciascuno di essi è soggetto a variazioni e adattamenti in base al contesto e all'evoluzione tecnologica, riflettendo la natura dinamica dei processi sociotecnici in cui sono inseriti (Latour, 1992; Orlikowski, Iacono, 2001). La non fissità di tali apparati costituisce una delle caratteristiche distintive delle tecnologie emergenti, la cui configurazione varia a seconda delle innovazioni, delle condizioni d'uso e degli attori coinvolti (Latour, 1987; Akrich, 1992; Orlikowski, 1992; Pinch, Bijker, 1984). Pur nella loro variabilità, questi tre apparati, e le rispettive componenti, costituiscono l'ossatura di ogni SDA, risultando indispensabili per assicurarne la coerenza operativa e la funzionalità complessiva. Tale complessità può essere indagata attraverso la definizione delle sue dimensioni costitutive; a questo scopo, la Tabella 1 delinea un quadro analitico dettagliato, classificando i diversi sistemi in base alle sfide etico-sociali che sollevano e alle strategie di mitigazione emerse dalle pratiche di ricerca e di gestione e anticipazione del rischio algoritmico.

Tab. 1 - Quadro analitico delle dimensioni dei sistemi decisionali automatizzati e delle sfide associate

Dimen- sione	Descrizione dimensione	Tipi della dimensione	Sfida princi- pale	Strategie di mitigazione
Livelli di Opa- cità	Trasparenza e ispezionabilità del funziona- mento interno del sistema.	Opaco (I); Parzialmente opaco (II); Ispezionabile (III).	Il livello di opacità del si- stema può ren- dere difficile ispezionare il suo funzio- namento e attribuire le responsabilità dei processi eseguiti.	Explainable AI (1): Integrazio- ne di strumenti che spiegano la logica dietro ogni singola decisione. Model Cards (2): Documentazione standardizzata che spiega limiti e funzionamento del modello agli utenti. Audit Algoritmici (3): Verifi- che periodiche da parte di enti terzi per certificare la trasparenza.
Gradi di Automa- zione	Misura in cui il sistema agi- sce autonoma- mente rispetto all'intervento umano.	Autonomo (I); Semi- autonomo (II); Non autonomo (III).	Rischio di perdere il con- trollo umano su decisioni critiche.	Human-in-the-loop (1): Obbligo di conferma umana per le decisioni ad alto impatto. Kill Switch & Over- ride (2): Possibilità per l'operatore di disattivare l'IA o di forzare una decisione manuale in qualsiasi momento. Automazione Graduale (3): Implementare l'autonomia a stadi, partendo da "suggerimento" fino ad "azione autonoma" solo post-validazione.

Modalità di Apprendimento	Come il sistema viene addestrato ed elabora le informazioni.	Apprendimento manuale (I); Apprendimento supervisionato (II); Apprendimento automatico (III).	Il sistema impara da dati errati e sviluppa bias.	Data Curation & De-biasing (1): Pulizia preventiva dei dataset per rimuovere pregiudizi storici o statistici. Apprendimento Continuo Supervisionato (2): Monitoraggio costante delle performance per correggere le “derive” nell’apprendimento. (3) Testing in ambienti sintetici: Addestrare il sistema su dati simulati per verificare come reagisce a condizioni rare prima del rilascio.
Tipo di Automazione	Definizione delle regole e delle logiche che definiscono l’esecuzione dei processi e dei meccanismi all’interno del sistema.	Sistemi basati su regole predefinite (logiche deterministiche) (I); Sistemi basati su Machine Learning (apprendimento dai dati/adattivi) (II).	Rigidità (se basato su regole) o Imprevedibilità (se basato su ML).	Sistemi Ibridi (1): Usare regole rigide per la sicurezza e Machine Learning per l’efficienza. Validazione formale (2): Dimostrazione matematica che il codice rispetta i requisiti. Sandboxing (3): Eseguire i processi complessi in un ambiente di prova isolato prima del rilascio su larga scala.
Funzioni e Compiti	Lo scopo operativo, le funzioni e i compiti rappresentano l’obiettivo e le mansioni per cui il SDA è stato progettato.	Classificazione di dati (I); Gestione di risorse (II); Supporto decisionale (III); Valutazione predittiva (IV).	Utilizzo del sistema fuori contesto o per scopi non etici.	Assessment (1): Valutazione preventiva dell’impatto sociale ed etico dello specifico compito affidato all’SDA. Limitazione di scopo (2): Blocchi software che impediscono all’IA di operare su dati o contesti per cui non è stata progettata. Valutare l’equità dei risultati (3): Non misurare solo l’accuratezza del compito, ma anche l’equità dei risultati prodotti.
Sicurezza del Sistema	Affidabilità e conformità del sistema, con particolare attenzione alla protezione, integrità e riservatezza dei dati.	Protezione dell’integrità dei dati (I); Prevenzione accessi non autorizzati (II); Resistenza a minacce informatiche/manipolazioni (III).	Vulnerabilità ad attacchi hacker o manipolazione dei dati.	Crittografia e anonimizzazione (1): Proteggere i dati in ingresso e in uscita per garantire la privacy. Red Teaming (2): Impiego di “hacker etici” per tentare di ingannare o penetrare il sistema e scoprirne le falle. Robustezza avversaria (3): Addestrare il sistema a riconoscere tentativi di manipolazione.

La prima dimensione riguarda i livelli di opacità, strettamente connessi alla trasparenza e alla possibilità di ispezionare il funzionamento interno del sistema. Poiché gli SDA operano spesso attraverso strutture algoritmiche complesse, comprendere in che modo le decisioni vengono generate e se i processi siano accessibili al controllo umano è essenziale per garanti-

re fiducia, responsabilità e legittimità nelle loro applicazioni. La seconda concerne i gradi di automazione, ossia la misura in cui il sistema agisce autonomamente rispetto all'intervento umano. Analizzare questo aspetto consente di valutare quanto controllo rimanga effettivamente nelle mani degli utenti, dei gestori, delle istituzioni o delle aziende che supervisionano il processo decisionale automatizzato. La terza dimensione è rappresentata dalle modalità di apprendimento implementate dal sistema. Esse influenzano sul modo in cui l'SDA viene addestrato ed elabora le informazioni, distinguendo tra apprendimento manuale, supervisionato o automatico. Questa distinzione è importante per comprendere la capacità del sistema di adattarsi ai dati e alle condizioni del contesto, ma anche per identificare possibili fonti di bias o di errore strutturale. La quarta dimensione riguarda la tipologia di automazione, ossia la natura dei processi automatizzati e i meccanismi che li sostengono. Essa permette di distinguere tra sistemi basati su regole predefinite, che seguono logiche deterministiche, e sistemi fondati su modelli di machine learning, che apprendono dai dati e modificano il proprio comportamento nel tempo. La quinta dimensione si riferisce alle funzioni e ai compiti svolti dal sistema. Gli SDA possono essere progettati per scopi molto diversi, dalla classificazione di dati alla gestione di risorse, dal supporto decisionale alla valutazione predittiva, e la loro analisi consente di chiarire l'ambito operativo, l'obiettivo e la pertinenza del sistema rispetto al contesto di applicazione. Infine, la sesta dimensione concerne la sicurezza del sistema, elemento imprescindibile per garantirne un funzionamento affidabile e conforme ai principi di tutela dei dati e della privacy. Essa include la protezione dell'integrità dei dati, la prevenzione di accessi non autorizzati e la capacità del sistema di resistere a minacce informatiche o manipolazioni esterne.

Considerate congiuntamente, queste sei dimensioni costituiscono una cornice utile per comprendere la struttura e il comportamento dei sistemi di decisione automatizzata, evidenziando gli elementi che ne determinano l'efficacia, la flessibilità e la credibilità nei diversi contesti d'uso.

Implicazioni sociali dei sistemi di decisione automatizzata

La ricerca scientifica sulle implicazioni sociali dei sistemi di decisione automatizzata ha assunto negli ultimi anni un ruolo centrale nello studio dei rischi connessi all'uso di tecnologie digitali (Pink, 2019). Questo ambito di indagine mira a comprendere gli impatti sociali, economici e culturali di tali sistemi, evidenziandone le criticità e promuovendo approcci orientati alla mitigazione dei rischi (Grange, Pinsonneault, 2021). Istituzioni

pubbliche, organizzazioni e comunità scientifiche hanno progressivamente sviluppato strategie per analizzare e ridurre i problemi derivanti dall'impiego di strumenti algoritmici (Aragona, Amato, 2022).

Un punto di partenza fondamentale è rappresentato dal riconoscimento della non neutralità algoritmica, ossia dalla consapevolezza che gli algoritmi riflettono le scelte, i valori e i pregiudizi di chi li progetta (Airoidi, Rokka, 2022). Questa consapevolezza ha favorito lo sviluppo di ricerche dedicate a comprendere come i sistemi automatizzati influenzino le dinamiche sociali e a individuare gli effetti potenzialmente dannosi prodotti dalle loro applicazioni. La letteratura su questi temi è ampia e frammentata e si concentra su aspetti specifici delle tecnologie digitali, come la trasparenza e la responsabilità algoritmica (Aragona, 2022), le dimensioni etiche e deontologiche (Mökander et al., 2021), i bias e le forme di discriminazione (Noble, 2018), la governance dei sistemi automatizzati (Pasquale, 2015) e le pratiche di resistenza o contro-egemoniche che emergono in risposta alla loro diffusione (Bonini, Treré, 2024). L'eterogeneità di tali approcci riflette la complessità intrinseca dei SDA, che rende difficile una valutazione unitaria dei rischi e dei benefici associati al loro impiego. Tuttavia, questa pluralità di prospettive costituisce anche un punto di forza, poiché l'interdisciplinarietà della ricerca consente di combinare contributi provenienti dalle scienze sociali, dall'informatica, dalla Human-Computer Interaction (HCI) e dagli studi organizzativi, coinvolgendo attivamente sviluppatori, utenti e cittadini interessati dagli effetti delle tecnologie (Salvini et al., 2023).

Un ulteriore filone di ricerca riguarda l'analisi degli effetti latenti e imprevisti dei sistemi di decisione automatizzata, che possono manifestarsi nel lungo periodo o in forme non immediatamente osservabili (Airoidi, 2023). Per comprendere tali effetti, è necessario considerare in profondità i contesti sociali, istituzionali e culturali in cui gli algoritmi vengono prodotti e implementati (Watson, Kharrufa, 2021). Questa esigenza impone di adottare approcci contestuali e situati, capaci di cogliere le interazioni complesse tra logiche tecnologiche e ambienti sociali.

Tra le principali linee di indagine emergono lo studio dei pregiudizi e delle disuguaglianze algoritmiche (Noble, 2018), l'analisi delle pratiche d'uso (Geiger, 2017), le ricerche sui consumi mediati dagli algoritmi (Couldry, Mejias, 2020) e sugli effetti quotidiani di tali mediazioni (Bucher, 2019). Ad esempio, O'Neil (2016) ha dimostrato come l'impiego di modelli predittivi nei processi di assunzione e nella valutazione del merito creditizio possa trasformarsi in "armi di distruzione matematica", capaci di codificare i pregiudizi sistemici e negare opportunità economiche a intere fasce di popolazione. Altri studi hanno analizzato l'implementazione di sistemi di monitoraggio della produttività basati su algoritmi granulari, che

trasformano l'ambiente lavorativo in un Panopticon digitale dove la valutazione automatizzata delle prestazioni incide direttamente sulla stabilità contrattuale e sulla salute psicofisica dei lavoratori (Moore, 2020). Parallelamente, l'ascesa dei sistemi di Social Scoring e di polizia predittiva ha sollevato profonde preoccupazioni circa la capacità degli SDA di ridefinire i confini della cittadinanza, trasformando la gestione dei servizi pubblici in uno strumento di sorveglianza e disciplina sociale che limita l'autonomia individuale in base a parametri di conformità algoritmica (Eubanks, 2018). Queste ricerche consentono di illuminare aspetti specifici, come i bias, gli impatti comportamentali e le implicazioni psicosociali, che risultano cruciali per comprendere il ruolo delle tecnologie decisionali nella società contemporanea. Le conoscenze che ne derivano contribuiscono, anche indirettamente, alla costruzione di strumenti di ricerca più efficaci, traducendo gli esiti teorici in indicatori e criteri di analisi. L'individuazione di distorsioni algoritmiche o di effetti negativi sulle interazioni sociali, ad esempio, può costituire la base per la definizione di metriche di equità o per la valutazione dell'impatto sociale dei sistemi. In questo senso, la ricerca sociale sugli SDA non si limita a una funzione descrittiva, ma agisce come ponte tra analisi teorica e pratica valutativa, fornendo gli elementi concettuali necessari per comprendere, misurare e governare l'impatto dei processi decisionali automatizzati nella vita collettiva.

5. *Gemelli digitali*

di *Mattia De Angelis*

Breve storia dei gemelli digitali urbani

La genealogia del concetto di gemello digitale si sviluppa all'interno delle discipline ingegneristiche, dove l'esigenza crescente di comprendere, monitorare e ottimizzare sistemi complessi ha portato alla costruzione di modelli capaci di rappresentare fedelmente il comportamento dei loro corrispettivi fisici. Le prime formulazioni riconoscono al gemello digitale la capacità di fornire una rappresentazione fisica e funzionale completa di un elemento fisico, permettendo di integrare in un unico ambiente digitale, i dati progettuali, operativi e predittivi, lungo l'intero ciclo di vita del prodotto (Boschert, Rosen, 2016). Ciò che emerge è un'idea di gemello digitale come un'infrastruttura cognitiva pensata per supportare decisioni tecniche mediante un allineamento costante tra mondo reale e rappresentazione virtuale.

La formulazione di Grieves (2014) segna una svolta teorica fondamentale. Nel suo modello, il gemello digitale è definito da una triade concettuale costituita da un'entità fisica, un'entità virtuale e un flusso informativo che li collega. Questa struttura chiarisce che la natura del gemello digitale è intrinsecamente relazionale, esso esiste nella misura in cui il legame informativo rimane attivo e coerente. Il flusso dati non è un elemento accessorio, ma il principio che rende possibile l'aggiornamento, la simulazione e l'analisi in parallelo rispetto all'evoluzione dell'elemento fisico. Non sorprende, quindi, che le interpretazioni successive enfatizzino la dimensione dinamica e predittiva del processo di definizione del gemello digitale.

In questa direzione si colloca la prospettiva avanzata da Glaessgen e Stargel (2012), i quali interpretano il gemello digitale come un sistema di simulazione avanzata capace di integrare modelli multi-fisici e multiscala con informazioni storiche e operazionali. Il gemello digitale viene così concepito come uno strumento in grado di anticipare il comportamento del

sistema fisico, generando scenari alternativi, valutando condizioni operative estreme e identificando possibili anomalie prima che si manifestino. La componente predittiva assume qui un valore centrale, introducendo un elemento che nei decenni successivi diventerà un tratto distintivo delle applicazioni più mature.

Con lo sviluppo teorico, si consolidano le prime implementazioni concrete in vari settori. Nel settore manifatturiero, viene impiegato per ottimizzare processi produttivi, simulare linee di assemblaggio, ridurre i tempi di fermo e migliorare la manutenzione predittiva (Negri et al., 2017). Le applicazioni nelle infrastrutture e nei sistemi tecnici complessi, come turbine, aeromobili e componenti mecatronici, consentono il monitoraggio di condizioni operative reali, valutare prestazioni e anticipare guasti. Allo stesso tempo, l'introduzione del gemello digitale nell'ambito delle filiere energetiche ha mostrato la capacità di integrare risorse e processi distribuiti, supportando l'ottimizzazione di generazione, stoccaggio e consumo in contesti ad alta variabilità (Ghenai et al., 2022).

La prospettiva di Kritzinger et al. (2018), che definisce il gemello digitale come una controparte digitale sincronizzata del sistema fisico, si rivela particolarmente adeguata a descrivere questa fase di implementazione eterogenea. I gemelli digitali mostrano infatti una notevole versatilità nel modellare componenti fisiche, macchine, processi, infrastrutture e sistemi interdipendenti, adattandosi alle specificità di diversi settori tecnologici. Questa flessibilità applicativa contribuisce alla loro progressiva affermazione come infrastruttura digitale trasversale.

Anche gli sviluppi teorici successivi convergono nel riconoscere che il valore del gemello digitale risiede nella capacità di aggregare dati, modelli e processi all'interno di un sistema coerente e integrato, capace di restituire informazioni difficilmente ottenibili mediante altre modalità (Grieves, Vickers, 2017). L'espansione verso domini sempre più articolati prepara il terreno alla sua successiva applicazione nei sistemi urbani.

I gemelli digitali si confrontano con ambienti urbani che mostrano alti gradi di complessità: sistemi aperti, infrastrutture sovrapposte, elevata variabilità spazio-temporale e numerosi stakeholder. Queste caratteristiche rendono l'applicazione diretta dei modelli industriali meno immediata (Weil et al., 2023) e impediscono di applicare meccanicamente il modello ingegneristico ai sistemi urbani. Gli studi che si occupano di definire il gemello digitale urbano insistono dunque sulla necessità di integrare dati provenienti da fonti eterogenee, reti fisiche, sensori ambientali, sistemi di mobilità, informazioni amministrative, e di costruire un ambiente digitale capace di rappresentare, simulare e analizzare fenomeni distribuiti nello spazio e nel tempo (Batty, 2018).

In linea con la letteratura sui gemelli digitali urbani, la proposta di Schrotter e Hürzeler (2020) interpreta il gemello digitale urbano come una piattaforma digitale che integra modelli 3D, dati territoriali e infrastrutturali, e che consente di simulare scenari urbani e valutare gli effetti delle decisioni pubbliche. Gli studi che definiscono il gemello digitale urbano lo interpretano come uno strumento avanzato per il governo urbano, in grado di integrare dati da fonti eterogenee, dai sensori ambientali alle infrastrutture, gestire fenomeni sociali, infrastrutture urbane e fornire supporto decisionale nella pianificazione urbana (Ferré-Bigorra et al., 2022). La letteratura più recente definisce il gemello digitale urbano come ambiente digitale integrato in cui i dati in tempo reale, i modelli predittivi e gli strumenti analitici convergono per supportare la pianificazione e la governance urbana (Argota Sánchez-Vaquerizo, 2025).

La definizione di gemello digitale che riportiamo in questo lavoro tiene conto degli elementi chiave emersi nella letteratura: un gemello digitale è una replica digitale avanzata di entità fisiche, progettata per monitorare, simulare e ottimizzare in tempo reale e/o in semi-tempo reale. Grazie alle tecnologie, crea una connessione bidirezionale con il mondo reale, supportando le decisioni strategiche e operative relative al governo urbano.

Questa definizione esprime tre passaggi concettuali rilevanti. Anzitutto, il gemello digitale è una replica avanzata, cioè una rappresentazione dinamica che integra dati, modelli e capacità simulative in un sistema coerente. In secondo luogo, la distinzione tra tempo reale e semi-real-time riconosce la natura ibrida delle fonti informative urbane: i dati derivano infatti da flussi automatizzati e da aggiornamenti manuali, che introducono una temporalità stratificata nel processo di sincronizzazione. Infine, la bidirezionalità non è un attributo tecnico secondario, ma il principio attraverso cui il modello digitale influisce sulle decisioni reali, orientando scelte operative e strategiche. Il gemello digitale urbano non è dunque un mero riflesso della città, ma un ambiente cognitivo e analitico che contribuisce attivamente alla sua gestione.

La natura sociale delle implementazioni dei modelli di gemelli digitali urbani

Le modalità con cui le città vengono rappresentate, osservate e governate hanno una lunga storia che precede di molto l'emergere delle infrastrutture digitali contemporanee. Le tecniche cartografiche, le classificazioni amministrative e le statistiche territoriali hanno svolto un ruolo cruciale nella costruzione della città come spazio leggibile, trasformato-

do fenomeni complessi in entità misurabili e comparabili (Harley, 1989; Cosgrove, 2003; Pickles, 1995). Rappresentare lo spazio urbano significa sempre operare selezioni, definire categorie e stabilire ciò che conta come fenomeno rilevante, secondo logiche istituzionali e culturali che hanno accompagnato la storia della governamentalità moderna (Scott, 1999; Curry, 1998). Le tecnologie digitali non fanno che estendere queste grammatiche di rappresentazione: rendono più capillare la raccolta dei dati, intensificano le possibilità di correlazione e accelerano la circolazione delle informazioni, ma non modificano la natura intrinsecamente politica delle operazioni di classificazione (Crampton, 2011; Del Casino, Hanna, 2005).

L'espansione delle infrastrutture urbane di monitoraggio e dei sistemi di sensing ha ulteriormente articolato questo regime di visibilità. Le piattaforme di raccolta e integrazione dei dati e gli strumenti di analisi automatizzata hanno reso le città ambienti sempre più datificati, in cui elementi eterogenei, mobilità, ambiente, servizi, interazioni, vengono continuamente trasformati in flussi operativi (Gabrys, 2014; Shelton et al., 2015). In questo scenario, non sono solo i dati a essere centrali, ma le infrastrutture che li generano, li normalizzano e li distribuiscono. La città diventa leggibile attraverso un insieme di dispositivi di misurazione e interoperabilità che ne mediano l'esperienza e ne orientano la governance (Graham, Marvin, 2001; Crang, Graham, 2007). È in questa traiettoria di lunga durata che si collocano i modelli di gemelli digitali urbani.

I Science and Technology Studies (STS) offrono un quadro interpretativo essenziale per comprendere la natura socio-materiale di queste implementazioni. Fin dagli studi fondativi, la tecnologia è stata letta non come un elemento neutrale, ma come un assemblaggio di attori, procedure e materiali che prende forma attraverso negoziazioni e traduzioni (Callon, 1984; Latour, 1987; Law, 1992). La distinzione tra “tecnico” e “sociale” viene rifiutata, poiché si sostiene che ogni tecnologia incorpora visioni del mondo, norme implicite e scelte organizzative che influenzano ciò che la società considera conoscibile o governabile. La svolta infrastrutturale ha ulteriormente consolidato questa riflessione mostrando come le infrastrutture funzionino grazie alla loro invisibilità, come elementi che operano in background stabilizzando routine e pratiche (Star, Ruhleder, 1996; Star, 1999). Bowker e Star (1999) hanno evidenziato come classificazioni e standard siano dispositivi attraverso cui la società ordina il mondo, includendo alcuni fenomeni e rendendone altri epistemicamente marginali.

Applicate ai gemelli digitali urbani, queste intuizioni mostrano come tali modelli non rappresentino semplicemente la città, ma partecipino alla sua costruzione materiale e cognitiva. Essi si inseriscono in reti eterogenee composte da sensori, database, modelli predittivi, procedure amministra-

tive e protocolli tecnici, formando un assemblaggio che produce visibilità selettive e forme operative di conoscenza (Edwards, 2010; Ribes, Bowker, 2009). La città diventa così un ambiente sociotecnico in cui attori umani e non umani co-producono rappresentazioni, simulazioni e possibilità d'azione (Farias, Bender, 2010; Jensen, Morita, 2017). Come mostrano Ribes e Jackson (2013), ciò che viene raccolto, mantenuto e reso interoperabile non è mai neutro, ma è il risultato di pratiche, convenzioni e priorità istituzionali che strutturano il dominio del governabile.

Una componente fondamentale dei gemelli digitali urbani risiede nel loro radicamento in un'epistemologia del dato. Come osserva Leonelli (2019), i dati non sono entità oggettive o immediatamente disponibili, ma materiali situati, prodotti all'interno di pratiche locali, strumenti specifici e procedure di validazione. In questa prospettiva, le infrastrutture digitali non si limitano a organizzare informazioni già date, ma contribuiscono a definire quali fenomeni possano essere inclusi nel modello, con quali attributi, a quale scala e secondo quali standard. Di conseguenza, il gemello digitale non si configura come un contenitore neutrale della realtà urbana, bensì come un dispositivo di classificazione che contribuisce a costruire ciò che appare come realtà urbana (Gitelman, 2013). L'oggettività, in questo senso, non deriva dalla distanza o dalla neutralità dello sguardo tecnico, ma dalla capacità di garantire tracciabilità, coerenza e continuità attraverso l'infrastruttura (Edwards, 2013). Tale condizione richiede inoltre un lavoro continuo di manutenzione, aggiornamento e riallineamento. Come mostrano Denis et al. (2015), le infrastrutture esistono infatti solo nella misura in cui vengono costantemente mantenute, riparate e curate. Il funzionamento di un gemello digitale urbano dipende quindi da forme di lavoro distribuito che ne assicurano tanto la stabilità operativa quanto la credibilità epistemica.

Le prospettive dei Critical Algorithm Studies (CAS) consentono di evidenziare un ulteriore livello di complessità, legato alla crescente centralità della computazione e della predizione (Gillespie, 2014; Kitchin, 2017; Seaver, 2017). Le tecnologie algoritmiche operano come forme di governo che modellano il futuro attraverso inferenze statistiche, trasformando la città in uno spazio di correlazioni e probabilità (Amoore, 2011; 2020; Rouvroy, Berns, 2013). Gli algoritmi non sono semplici strumenti, ma attori epistemiche che definiscono ciò che può essere anticipato o ottimizzato.

Considerate in questa prospettiva, le implementazioni dei gemelli digitali urbani appaiono come esito di una co-produzione continua tra infrastrutture materiali e razionalità algoritmiche, tra epistemologie del dato e pratiche istituzionali. Essi articolano diverse temporalità, dalle routine amministrative ai flussi real-time e diverse forme di conoscenza, dalla classifi-

cazione alla predizione, producendo ambienti in cui la città è costantemente resa osservabile, simulabile e anticipabile. La loro natura sociale emerge nel modo in cui selezionano i fenomeni rilevanti, ne definiscono le relazioni, e trasformano tali relazioni in possibilità operative per il governo urbano. In questo senso, i gemelli digitali urbani non sono soltanto strumenti di rappresentazione, ma dispositivi attraverso cui la città diventa un oggetto di calcolo, una forma di conoscenza e un campo di decisione.

I gemelli digitali urbani, tematiche locali e future domande di ricerca

In questo paragrafo, il contributo delle scienze sociali consiste nel chiarire quali forme di conoscenza, quali relazioni di potere e quali questioni locali emergono dall'adozione di questi sistemi nei contesti urbani.

Quindi si mostrano una serie di tematiche locali rilevanti per le scienze sociali connesse all'implementazione dei gemelli digitali urbani. Tematiche come la mediazione algoritmica che introduce nuove forme di agency distribuita, poiché gli algoritmi non si limitano a “supportare” decisioni preesistenti, ma contribuiscono a strutturare ciò che appare come problema, rischio o opportunità, secondo logiche operative che sono al tempo stesso tecniche ed epistemiche (Mackenzie, 2017; Beer, 2018; Seaver, 2017; Crawford, 2021). Questioni relative a standard, protocolli e formati e come questi contribuiscono a definire chi può contribuire ai processi di modellizzazione e in che modo (Timmermans, Epstein, 2010; Plantin et al., 2018).

Inoltre, i gemelli digitali urbani possono contribuire a rafforzare disuguaglianze socio-spaziali preesistenti, poiché la possibilità di rappresentare territori, gruppi sociali o problemi urbani dipende dalla disponibilità, dalla qualità e dalla granularità dei dati; in questo senso, ciò che non viene misurato o formalizzato rischia di restare ai margini tanto del modello quanto delle priorità di intervento pubblico (Taylor, 2017; Kitchin et al., 2015). A ciò si aggiungono questioni relative alla privacy, alla sorveglianza e alla governance dei dati, poiché l'integrazione continua di flussi informativi eterogenei può ampliare le capacità di monitoraggio urbano e rendere più sfumati i confini tra gestione della città, osservazione dei comportamenti e controllo infrastrutturale (van Zoonen, 2016; Kitchin, 2016).

La letteratura recente sui gemelli digitali urbani ha iniziato a interrogare anche le possibilità di orientare questi sistemi verso forme di “participatory steering”, esplorando come i modelli possano essere usati per aprire, anziché chiudere, spazi di coinvolgimento e deliberazione (Dawkins, Kitchin, 2025). A questi aspetti si collega il problema della trasparenza e della contestabilità dei modelli, ovvero quanto più i gemelli digitali entrano

nei processi decisionali, tanto più diventa rilevante chiarire quali assunzioni, soglie e criteri siano incorporati nelle simulazioni e chi sia effettivamente in grado di comprenderli o di metterli in discussione (Burrell, 2016; Ananny, Crawford, 2018). Da qui derivano interrogativi cruciali: su chi ha il potere di interrogare il modello, di modificare i parametri, di contestare gli scenari, su quali soggetti e quali saperi vengono inclusi o esclusi da queste procedure, su quali esperimenti di uso partecipativo dei gemelli digitali urbani sono in corso.

È possibile delineare ulteriori traiettorie di ricerca per le scienze sociali, ad esempio: come i gemelli digitali urbani contribuiscono a ridefinire le ontologie della città, quali criteri di validità vengono applicati alla conoscenza prodotta, come si riconfigurano i rapporti tra dato, modello e decisione. Altri interrogativi riguardano le dimensioni organizzative e politiche, chiedendosi in che modo l'adozione di modelli predittivi ristruttura i processi decisionali, redistribuisce responsabilità tra attori umani e non umani e trasforma le forme di accountability istituzionale (Amoore, 2020; Rouvroy, Berns, 2013). Un'ulteriore linea di ricerca riguarda il lavoro ordinario e spesso poco visibile che sostiene questi sistemi, fatto di aggiornamento, verifica, manutenzione e riallineamento continuo dei dati e delle infrastrutture, così come il ruolo dei fornitori privati e delle piattaforme proprietarie, che possono ridefinire i rapporti tra amministrazioni pubbliche e attori tecnologici e produrre nuove forme di dipendenza tecnica (Dennis et al., 2015; Plantin et al., 2018; Voorwinden, 2021).

Le questioni qui richiamate non esauriscono le possibilità di ricerca, ma rappresentano alcune traiettorie preliminari attraverso cui le scienze sociali possono contribuire a comprendere e problematizzare i gemelli digitali urbani. La letteratura su questi sistemi è ancora in larga misura embrionale, ma in rapido sviluppo; proprio per questo, i gemelli digitali urbani offrono un laboratorio privilegiato per interrogare come la città venga continuamente ricomposta come oggetto di calcolo, di anticipazione e di governo, aprendo un campo di indagine che richiederà, nei prossimi anni, un dialogo serrato tra studi urbani, tecnologici e scienze sociali, anche perché parliamo di applicazioni in fase effettiva di sviluppo.

6. Le tecnologie quantistiche

di Suania Acampa

Tra le tecnologie emergenti che stanno ridefinendo profondamente la società, le tecnologie quantistiche (TQ) occupano un ruolo del tutto particolare. Le TQ aprono certamente una nuova fase dello sviluppo tecnologico e, per molti versi, implicano anche un cambiamento di prospettiva epistemologica. È come se un intero modo di intendere la fisica (e ciò che possiamo farci) si stesse spostando su un altro piano: da un mondo in cui la fisica quantistica serviva a descrivere i fenomeni microscopici, a un mondo in cui quelle stesse leggi vengono ingegnerizzate (Dowling, Milburn, 2003) per costruire nuovi dispositivi, infrastrutture e modelli di calcolo. Il Quantum Manifesto¹ (2016) ha descritto questa transizione come l'avvento di una «*seconda rivoluzione quantistica*», capace di generare «*una nuova ondata di tecnologie basate sulle leggi della meccanica quantistica [...] che avranno impatti rivoluzionari sulla scienza, sull'industria e sulla società*» (p. 7).

Ciò che distingue le TQ da altre innovazioni è legato tanto alla potenza di calcolo e di misura quanto al cambiamento radicale di paradigma tecnologico che comportano. Come si legge nella European Declaration on Quantum Technologies (2023), sono spesso descritte come tecnologie *paradigm-shifting*, ossia in grado di modificare «*le capacità tecnologiche fondamentali e il modo stesso di concepire l'informazione, la materia e la conoscenza*».

¹ Un documento pubblicato nel 2016 dalla comunità scientifica europea per delineare una visione condivisa sullo sviluppo delle tecnologie quantistiche e per promuovere un investimento strategico da parte dell'Unione Europea. Il manifesto ha svolto un ruolo decisivo nel convincere la Commissione Europea ad avviare la Quantum Flagship, il grande programma da un miliardo di euro lanciato nel 2018 per sostenere ricerca, innovazione e applicazioni industriali nel campo quantistico. In sintesi, è il testo che ha segnato l'inizio della strategia europea sulle tecnologie quantistiche, definendone priorità, obiettivi e rilevanza per la competitività e la sicurezza dell'UE.

Kaku (2023) interpreta la nascita delle tecnologie quantistiche come la fine dell'età digitale, intesa non come semplice evoluzione tecnologica, bensì come trasformazione del modo di pensare la conoscenza. Kaku descrive il digitale come un paradigma fondato sulla linearità e sulla logica binaria, che riduce la realtà a entità discrete e prevedibili. Questo modello, che ha funzionato per decenni, oggi fatica sempre più a rendere conto della complessità dei sistemi naturali e sociali. Mentre il calcolo classico, basato su sequenze e opposizioni, restituisce un'immagine semplificata del reale, le tecnologie quantistiche ne scardinano le fondamenta, sostituendo la logica binaria con una computazione su stati multipli, fondata sulla sovrapposizione e sull'entanglement.

La rivoluzione quantistica segna così un mutamento profondo nella relazione tra scienza e società: da un modello di razionalità orientato al controllo a un principio relazionale e probabilistico, in cui conoscere significa partecipare ai processi del reale.

Negli ultimi vent'anni, le istituzioni europee hanno riconosciuto il valore strategico del settore: dal Quantum Flagship (European Commission, 2018) alla Dichiarazione Europea sulle Tecnologie Quantistiche (2023), fino al futuro European Quantum Act², l'obiettivo dichiarato è quello di costruire un ecosistema integrato di ricerca, innovazione e formazione, capace di trasformare le scoperte della fisica quantistica in applicazioni concrete per l'industria, la sicurezza, l'ambiente e la società civile. Ma, come ogni innovazione tecnologica, anche questa comporta nuove sfide etiche, sociali e politiche, tra cui la necessità di garantire un accesso equo alle risorse computazionali, di evitare concentrazioni di potere tecnologico e di sviluppare forme di governance trasparenti e inclusive (OECD, 2024; Royal Society, 2020).

In questo quadro, l'Europa e i Paesi che stanno investendo nelle TQ hanno progressivamente ampliato l'attenzione non solo al "cosa" sviluppare, ma anche al "come" orientarne le traiettorie. Sebbene questo spostamento verso le modalità di sviluppo caratterizzi oggi molte politiche dell'innovazione tecnologica, nel caso delle TQ esso assume una particolare rilevanza per l'elevata incertezza delle traiettorie, l'intensità degli investimenti strategici e le potenziali implicazioni per la sicurezza e l'economia globale. Non si tratta più soltanto di promuovere la prossima generazione di computer o sensori, ma di iscriverne lo sviluppo entro cornici discorsive che richiamano responsabilità, sostenibilità e valori democratici. In

² Al momento in cui si scrive è aperta una consultazione pubblica aperta a ricevere contributi per dare forma al documento previsto per il 2026: <https://digital-strategy.ec.europa.eu/en/news/commission-invites-contributions-shape-future-eu-quantum-act>.

questo contesto, il paradigma della Responsible Research and Innovation (Owen et al., 2013) e quello dell'*anticipatory governance* (Guston, 2014) vengono sempre più frequentemente mobilitati nelle politiche di settore, insieme a modelli di Quantum Risk Management e di Shared Stewardship proposti da agenzie internazionali. Più che configurarsi come soluzioni stabilizzate, tali cornici contribuiscono a costruire le tecnologie quantistiche come uno spazio sperimentale di governance in cui si ridefiniscono (non senza tensioni) le relazioni tra conoscenza, potere e responsabilità.

Per questo motivo, parlare di TQ solo come un ambito della fisica non è sufficiente: implica anche un diverso modo di pensare la conoscenza e il ruolo dell'osservazione nel costruire il reale: esse trasformano la materia e l'informazione in dimensioni interdipendenti, in cui l'osservazione, la misura e l'elaborazione coincidono. Questa prospettiva, come osservano Calzati e De Kerckhove (2024), tende a spostare il baricentro del sapere dall'oggetto al processo, dal contenuto all'interazione e da una logica dell'essere a una del divenire.

Ecco che, come scienziati sociali, ci interroghiamo su come questa nuova infrastruttura del sapere trasformi i confini stessi della conoscenza, della responsabilità e del potere. Proveremo, in questo capitolo, a comprendere che cosa siano le TQ, che cosa rendono possibili sul piano tecnico-scientifico e come ridefiniscano il rapporto tra scienza e società.

Genealogia e definizioni della seconda rivoluzione quantistica

Le tecnologie quantistiche possono essere definite come insiemi di conoscenze e dispositivi che utilizzano direttamente i principi della meccanica quantistica per elaborare, trasmettere e misurare informazioni. Esse rappresentano un passaggio dall'osservazione dei fenomeni quantistici alla loro manipolazione controllata: un cambio di prospettiva che trasforma la fisica in ingegneria del mondo subatomico. Come sottolineato nel Quantum Manifesto (2016), mentre la prima rivoluzione quantistica ha posto le basi della società digitale, la seconda ne eredita i principi fisici e li trasforma in risorsa tecnologica.

Ma andiamo con ordine. L'origine delle TQ risale alle prime scoperte del XX secolo: la quantizzazione dell'energia di Planck (1900), la descrizione dell'effetto fotoelettrico di Einstein (1905), il modello atomico di Bohr (1913), la meccanica ondulatoria di Schrödinger e il principio di indeterminazione di Heisenberg (1927).

Da queste intuizioni nacquero la fisica moderna e, con essa, la prima rivoluzione tecnologica del secolo scorso.

Kaku (2023) descrive la nascita della fisica quantistica come un evento che trasformò radicalmente il modo di concepire la realtà e la conoscenza: la materia cessò di essere descritta da leggi deterministiche e continue per assumere la forma di stati probabilistici e discreti. La disputa tra Einstein e Bohr rese evidente la portata epistemologica di questa rottura: per il primo, la fisica doveva restare ancorata a una causalità oggettiva; per il secondo, l'osservatore diventava parte integrante del fenomeno osservato (Kaku, 2023). Da questa tensione nacque una nuova visione del mondo, in cui l'indeterminazione e la relazione sostituiscono la certezza e la separazione tra soggetto e oggetto.

Questa prospettiva ha progressivamente aperto la strada all'idea del calcolo quantistico, a partire dai contributi di Dirac e Feynman (Kaku, 2023). Feynman, in particolare, dimostrò che una particella può percorrere simultaneamente tutti i percorsi possibili, anticipando la logica del qubit e della computazione parallela. È in questo senso che la "*rinascita del quantico*" rappresenta, per Kaku, il punto d'origine della seconda rivoluzione quantistica: il momento in cui la fisica cessa di limitarsi a descrivere il mondo e inizia a emularlo.

Tuttavia, è solo alla fine degli anni Novanta che la ricerca ha iniziato a sviluppare tecniche per controllare singoli stati quantici della materia e della radiazione, aprendo la strada a una nuova era della tecnologia (Dowling, Milburn, 2003). Questa fase segna il passaggio dalla teoria alla manipolazione sperimentale dei fenomeni quantistici e pone le basi della seconda rivoluzione quantistica: se la prima aveva spiegato i principi fondamentali della fisica subatomica, la seconda li trasforma in risorsa ingegneristica, traducendoli in protocolli, algoritmi e dispositivi.

È su queste basi teoriche che si fonda questa trasformazione radicale:

1. quantizzazione, secondo cui energia e materia non variano in modo continuo ma si presentano in quantità discrete (i "quanti");
2. sovrapposizione, per cui un sistema quantistico può trovarsi contemporaneamente in più stati possibili fino al momento della misura;
3. *entanglement*, ovvero la correlazione profonda e simultanea tra sistemi anche distanti, tale che modificare uno influenza istantaneamente l'altro;
4. decoerenza, che indica la perdita delle proprietà quantistiche di coerenza e di sovrapposizione quando il sistema interagisce con l'ambiente esterno.

Questi principi costituiscono la base fisica comune ai quattro ambiti principali delle tecnologie quantistiche: Quantum Computing, Quantum Communication, Quantum Simulation e Quantum Sensing and Metrology (Quantum Flagship Strategic Research Agenda, 2018).

Questi settori condividono la stessa logica: sfruttare la coerenza quantistica per migliorare le capacità di calcolo, di trasmissione e di misura dell'informazione.

Il quantum computing è forse il simbolo più potente di questa rivoluzione. Come scrivono Benenti, Casati e Montangero (2025), il computer quantistico non è semplicemente più veloce di uno classico, ma si basa su una logica diversa, fondata sulla sovrapposizione e sull'interferenza che consente di elaborare le informazioni in modo parallelo e probabilistico, anziché sequenziale come nei computer tradizionali. Il passaggio dal bit al qubit rappresenta una trasformazione concettuale: mentre un bit classico può assumere solo due valori (0 o 1), un qubit può esistere in una combinazione dei due stati, consentendo al processore di eseguire operazioni in parallelo su un numero enorme di possibili stati.

Un modo semplice per capirlo è pensare al tentativo di aprire una serratura: un computer classico prova le chiavi una alla volta; quello quantistico, per semplificare, lavora come se le stesse testando tutte insieme. È proprio questo meccanismo – ossia la possibilità di elaborare in parallelo e di selezionare la soluzione ottimale tramite la sovrapposizione – che sta alla base degli algoritmi quantistici, con i quali è possibile affrontare problemi oggi considerati irrisolvibili dai calcolatori tradizionali.

Gli algoritmi di Shor e Grover, descritti da Nielsen e Chuang (2010), ne sono due esempi. Semplificando molto, il primo sfrutta la capacità del calcolo quantistico di elaborare molte combinazioni in parallelo per scomporre rapidamente numeri interi molto grandi, un'operazione che nei computer classici richiederebbe tempi astronomici e che costituisce la base della crittografia moderna; il secondo utilizza l'interferenza quantistica per accelerare la ricerca di un elemento in un grande insieme di dati, riducendo il numero di tentativi necessari. È così chiaro che, in entrambi i casi, il vantaggio non deriva da un aumento della potenza di calcolo in senso tradizionale, bensì da una logica di elaborazione diversa, basata appunto sulle proprietà fisiche dei sistemi quantistici.

Oggi, come spiega la Strategia Italiana per le Tecnologie Quantistiche (2025), siamo ancora in una fase di sviluppo intermedio: i computer quantistici attuali, detti NISQ (Noisy Intermediate-Scale Quantum), utilizzano processori con alcune decine o centinaia di qubit, ma restano instabili e soggetti a errori, poiché i qubit risentono facilmente delle interferenze dell'ambiente.

La sfida dei prossimi anni sarà quindi quella di realizzare sistemi più affidabili e “tolleranti agli errori” (Fault-Tolerant Quantum Computing) cioè capaci di correggere automaticamente le imprecisioni che si generano durante i calcoli e di garantire risultati coerenti e ripetibili.

Accanto al calcolo, il secondo pilastro delle tecnologie quantistiche è la comunicazione quantistica, che sfrutta due proprietà fondamentali della fisica dei quanti: l'*entanglement*, cioè una correlazione profonda tra particelle che restano connesse anche a grandi distanze, e il principio di indeterminazione, secondo cui osservare un sistema ne modifica lo stato. Queste due leggi, combinate, permettono di costruire canali di comunicazione teoricamente sicuri, poiché ogni tentativo di spiare il messaggio lascia una traccia fisica.

Il sistema più noto oggi è la Quantum Key Distribution (QKD), un metodo per generare e condividere una chiave di cifratura segreta tra due interlocutori. La chiave è formata da coppie di fotoni "*entangled*" (lett. impigliati), cioè collegate tra loro come se fossero un'unica entità. Se un terzo tenta di intercettare il flusso di fotoni, l'atto di osservazione altera il loro stato quantico, segnalando immediatamente l'intrusione.

Per capire questo principio, si può pensare al gatto di Schrödinger nel suo famoso esperimento mentale: finché non apriamo la scatola per osservarlo, si trova in una condizione di sovrapposizione quantistica – cioè può essere sia vivo che morto contemporaneamente – ma, nel momento in cui lo misuriamo (aprendo la scatola), il suo stato si riduce (o collassa) a uno dei due stati possibili (vivo o morto). Allo stesso modo, osservare un fotone equivale a modificarne lo stato perché nel mondo quantistico la semplice azione del guardare produce un effetto fisico. Questa proprietà rende la comunicazione quantistica uno dei campi più promettenti della seconda rivoluzione quantistica, con applicazioni che vanno dalle reti ultrasicure per la difesa e la finanza fino alla futura "internet quantistica", una rete globale basata su collegamenti fotonici in grado di trasmettere informazioni con livelli di sicurezza e velocità mai raggiunti prima (Quantum Manifesto, 2016).

Il terzo pilastro delle tecnologie quantistiche è rappresentato dalla misurazione e dalla metrologia quantistiche (*quantum sensing e metrology*), un campo che applica gli stessi principi fisici (sovrapposizione, *entanglement* e indeterminazione) per sviluppare strumenti di misura estremamente precisi. In termini semplici, questi sensori riescono a rilevare variazioni minime di tempo, di gravità o di campo magnetico, operando al limite della precisione consentita dalle leggi della natura.

Per capire il salto di qualità che comportano, si può pensare a un termometro capace di misurare differenze di temperatura milioni di volte più piccole di quelle percepibili da un sensore tradizionale, oppure a un orologio atomico in grado di perdere un solo secondo in milioni di anni. Questi dispositivi non solo migliorano la nostra capacità di osservare il mondo, ma offrono anche nuove possibilità in ambiti molto diversi: in medicina, ad

esempio, i magnetometri quantistici possono registrare minuscole variazioni del campo magnetico cerebrale, consentendo di mappare l'attività neuronale con una precisione mai raggiunta prima. In geofisica, interferometri basati su atomi ultra-freddi vengono utilizzati per misurare le variazioni del campo gravitazionale terrestre, consentendo di individuare movimenti tettonici impercettibili o risorse sotterranee. Nella navigazione e nella comunicazione digitale, infine, gli orologi quantistici consentono una sincronizzazione delle reti estremamente accurata, migliorando la localizzazione di veicoli e satelliti anche in assenza di GPS (Strategia Italiana per le Tecnologie Quantistiche, 2025).

In tutti questi casi, la differenza rispetto ai precedenti strumenti di misurazione non è solo di precisione, ma di principio: mentre la tecnologia classica misura grandezze già determinate, la metrologia quantistica sfrutta direttamente le proprietà quantistiche della materia per definire nuovi standard di misura. In un certo senso, è come se la natura, attraverso atomi e fotoni, entrasse nell'apparato di misura, diventandone parte attiva.

Le tecnologie quantistiche costituiscono oggi un ecosistema integrato, in cui i diversi ambiti si sostengono a vicenda: il calcolo quantistico alimenta la simulazione dei materiali e degli algoritmi; la comunicazione quantistica costruisce reti sicure per la trasmissione dei dati; la sensoristica quantistica fornisce informazioni ultra-precise che migliorano entrambi.

Non a caso, l'Unione Europea le definisce *platform technologies*, ossia piattaforme abilitanti per l'intero sistema dell'innovazione (European Commission, 2018).

La portata della seconda rivoluzione quantistica non si limita alle applicazioni tecniche, poiché, come sottolinea la European Declaration on Quantum Technologies (2023), le TQ sono spesso descritte come tecnologie "*paradigm-shifting*". L'espressione non è casuale: indicano un cambio di orizzonte nel modo di pensare la realtà fisica e ciò che possiamo controllare o manipolare attraverso la tecnologia. Esse spostano la scienza da una logica di rappresentazione a una logica di co-costruzione del reale, in cui conoscere e intervenire diventano due aspetti dello stesso processo. In questo senso, la seconda rivoluzione quantistica segna un cambiamento ontologico ed epistemologico: la distinzione tra osservatore e oggetto si fa porosa, la misura diventa parte del sistema e la realtà assume una natura relazionale.

Questa trasformazione, tuttavia, «*non risiede soltanto nei nuovi dispositivi, ma nella capacità di trasformare l'intera infrastruttura tecnologica e industriale*» (Quantum Manifesto, 2016, p. 9). Quindi, le implicazioni di questo ecosistema vanno ben oltre la sfera della ricerca di base e in questo molti osservatori, vedono l'inizio di una nuova infrastruttura tecnologica

che è spesso presentata come capace di cambiare (tutto o in parte) il modo in cui produciamo, curiamo e gestiamo le risorse (Kaku, 2023). Nella Quantum Economy, il calcolo quantistico diventa una risorsa strategica per l'innovazione industriale e la competitività globale: consente di modellare con precisione atomica processi fisici e chimici complessi, di simulare reti logistiche, di ottimizzare le transazioni finanziarie e di sviluppare materiali e farmaci in ambienti computazionali che riproducono la dinamica reale della materia. L'economia quantistica è, in questo senso, un'economia della simulazione, capace di trasformare il calcolo in uno strumento diretto di progettazione e di governo della complessità. Nel campo della medicina, la capacità di elaborare simultaneamente stati multipli di sistemi molecolari apre la via a una nuova fase della ricerca biomedica. Le tecniche di simulazione quantistica permettono di prevedere il comportamento delle proteine e delle molecole con un grado di precisione irraggiungibile dai modelli classici, rendendo possibile una farmacologia personalizzata, predittiva e sostenibile. Questa prospettiva annuncia un cambiamento di paradigma nella cura: non sperimentazione *ex post*, ma progettazione preventiva delle interazioni tra farmaci e organismi, in una medicina fondata sul principio di previsione quantistica. Le stesse capacità di simulazione si applicano anche alle sfide energetiche e climatiche: la computazione quantistica permetterà di modellare con accuratezza i processi di fotosintesi, le reazioni nucleari e i sistemi atmosferici, aprendo la via allo sviluppo di energie pulite e alla comprensione fine delle dinamiche del clima globale.

È possibile che la seconda rivoluzione quantistica rappresenti davvero un passaggio di fase nella storia della conoscenza; e non solo per le applicazioni tecniche, ma per il modo in cui ci costringe a pensare diversamente il rapporto tra calcolo, materia e conoscenza. Ma la posta in gioco non è solo tecnica, anche politica ed etica: è necessario orientare questa potenza emergente entro orizzonti di responsabilità, equità e sostenibilità.

Le sfide globali della governance quantistica

Le tecnologie quantistiche sono forse uno degli esempi più chiari della condizione di emersione continua che caratterizza molte tecnologie emergenti: non solo per la loro novità, ma anche perché si trovano in uno stato costante di ridefinizione, in cui scienza, politica e cultura interagiscono nel determinare che cosa siano, a che cosa servano e quali valori (espliciti o impliciti) finiscano per incorporare.

Come approfondiremo nei capitoli successivi, le tecnologie emergenti si sviluppano attraverso promesse e visioni del futuro che non si limitano ad

anticipare l'innovazione, ma la orientano performativamente (Borup et al., 2006). In queste promesse e visioni, la distinzione tra potenzialità e realtà rimane deliberatamente sottile (Konrad, Böhle, 2019) e alimenta l'intreccio continuo tra ricerca scientifica, investimento politico e immaginazione collettiva (Felt, Wynne, 2007; Rip, 2018).

È proprio da questa condizione che deriva una delle sfide più urgenti di oggi: governare la complessità di un ecosistema in perenne trasformazione, che pone interrogativi etici, sociali e politici che travalicano la dimensione tecnica. Certamente, tali questioni non sono esclusive delle tecnologie quantistiche (come da tempo evidenziato dagli studi STS, ogni tecnologia incorpora valori, visioni e rapporti di potere) tuttavia, nel caso delle TQ, esse assumono una particolare intensità per almeno tre ragioni: l'elevata incertezza delle traiettorie di sviluppo, la forte concentrazione di competenze e infrastrutture e il potenziale impatto sistemico su sicurezza, economia e governance dell'informazione. In questo senso, non si tratta solo di capire cosa possono fare le TQ, ma come, per chi e con quali valori vengono prodotte.

Queste sfide si pongono in un contesto segnato da forti asimmetrie nell'accesso all'innovazione e da una crescente concentrazione del potere computazionale.

Secondo il Policy Framework for Quantum Technologies dell'OCSE (2024), le tecnologie quantistiche dovranno essere sviluppate «*in modo sicuro, trasparente e inclusivo, garantendo che i benefici derivanti dalla seconda rivoluzione quantistica siano distribuiti equamente*» (p. 7). Il documento individua tra le principali sfide la disuguaglianza nell'accesso alle risorse computazionali, tra cui la standardizzazione delle infrastrutture, la protezione dei dati sensibili e la gestione del dual use delle tecnologie.

Ma quelle stesse caratteristiche che rendono i computer quantistici strumenti straordinari possono, allo stesso tempo, generare rischi non trascurabili: potrebbero compromettere la sicurezza dei sistemi crittografici globali o ampliare le disuguaglianze economiche tra Paesi e attori in grado di sostenere i costi elevati dell'hardware e delle infrastrutture necessarie.

In questa prospettiva, l'OCSE richiama la necessità di una governance quantistica multilivello, in grado di bilanciare competitività e cooperazione, innovazione e la tutela dei diritti fondamentali.

Come osserva l'OCSE, la questione del rischio riguarda anche temi quali la fiducia pubblica, la distribuzione dei benefici e la possibilità per i cittadini di comprendere e orientare le scelte tecnologiche che incideranno sulla loro vita.

Come evidenzia Stilgoe (2023), ogni tecnologia emergente rinegozia il rapporto tra innovazione e regolazione, e le tecnologie quantistiche am-

plificano questa tensione perché si collocano in una zona di confine tra la ricerca di base e l'applicazione industriale, dove gli strumenti normativi attualmente esistenti risultano inadeguati. Stilgoe invita a non concepire la regolazione come un ostacolo, bensì come una forma di innovazione sociale: «ogni tecnologia emergente rinegozia il rapporto tra innovazione e regolazione [...] chi guida l'innovazione non è soltanto chi la produce, ma anche chi definisce i limiti e le condizioni del suo uso» (pp. 6-7).

In questa direzione, la *governance* delle tecnologie quantistiche deve assumere la forma di una meta-regolazione, capace di adattarsi all'evoluzione rapida del settore e di integrare in modo coerente competenze etiche, giuridiche e tecniche. Solo un approccio più flessibile e riflessivo può aiutare a garantire uno sviluppo delle TQ socialmente sostenibile, evitando che l'accelerazione scientifica superi la capacità collettiva di governarne gli effetti.

Anche la Royal Society (2020) ha sottolineato che il potenziale “*disruptive*” delle TQ richiede un approccio etico proattivo. I temi centrali sono tre: dual use, trasparenza e accountability.

Le TQ possono avere applicazioni civili e militari al contempo (basti pensare ai sistemi di navigazione quantistica o alle comunicazioni inviolabili) e ciò impone una riflessione sulla responsabilità dei ricercatori e delle imprese coinvolte. L'assenza di standard condivisi rischia di creare un far west tecnologico, in cui le scelte progettuali generano effetti sociali non deliberati.

Nel dibattito contemporaneo, le TQ vengono così lette come un caso emblematico di ciò che Owen et al. (2013) hanno definito *Responsible Research and Innovation* (RRI), ossia una modalità di produzione della conoscenza orientata ai principi di anticipazione, inclusione, riflessività e reattività. Applicare l'RRI alle TQ significa interrogarsi preventivamente sulle conseguenze sociali e ambientali dell'innovazione, coinvolgere gli attori non scientifici nei processi decisionali, promuovere una cultura della riflessività tra ricercatori e garantire la capacità istituzionale di correggere le traiettorie tecnologiche quando emergono nuovi rischi o valori. È un approccio che sposta il baricentro della governance dall'idea di controllo a quella di responsabilità condivisa, in cui la ricerca scientifica diventa un'impresa collettiva di costruzione del futuro. Questo tipo di governance richiama il concetto di “*anticipatory governance*” (Guston, 2014), spesso mobilitato nel dibattito sulle tecnologie emergenti. Secondo Guston, la governance anticipatoria indica la capacità delle società democratiche di “prepararsi e rispondere” a innovazioni in divenire attraverso strumenti di previsione, partecipazione e costruzione di capacità istituzionali. Tuttavia, l'applicazione di questo approccio al campo delle tecnologie quantistiche

solleva alcune criticità. L'elevata incertezza epistemica che caratterizza lo sviluppo delle TQ rende infatti problematica l'idea di anticipare in modo lineare traiettorie e impatti. Le pratiche di anticipazione incorporano infatti specifici immaginari sociotecnici e possono contribuire a stabilizzare specifiche visioni del futuro più che a mantenerle aperte al confronto pubblico (Jasanoff, 2015). In questo quadro, strumenti quali scenari, sandbox regolatorie o laboratori di policy sperimentale possono certamente favorire forme di apprendimento istituzionale e di coordinamento tra attori. Tuttavia, essi non eliminano le asimmetrie di potere né le incertezze strutturali che accompagnano l'emergere delle tecnologie quantistiche.

Non a caso, la Strategia Italiana per le Tecnologie Quantistiche (2025) insiste sulla necessità di affiancare all'eccellenza scientifica un'attenzione alla dimensione etica e sociale, richiamando temi quali la trasparenza dei partenariati pubblico-privati, la sostenibilità dei processi produttivi e la valorizzazione della diversità nei gruppi di ricerca. Tuttavia, più che risolvere tali tensioni, questa enfasi segnala il tentativo (tipico delle politiche dell'innovazione) di rendere governabile un campo tecnologico caratterizzato da elevata incertezza e forti asimmetrie di potere. In questa prospettiva, il lessico della responsabilità e dell'inclusione può essere interpretato non solo come un orientamento normativo, ma anche come parte di un più ampio lavoro di legittimazione e stabilizzazione delle traiettorie di sviluppo quantistico.

Letta in chiave analitica, questa enfasi segnala come la governance quantistica costituisca già oggi un campo di tensione tra obiettivi di competitività, esigenze di controllo e istanze di responsabilità pubblica.

In definitiva, la governance delle tecnologie quantistiche non riguarda soltanto la gestione dei rischi né può essere ridotta a un esercizio puramente anticipatorio. Piuttosto, essa si configura come un terreno intrinsecamente politico in cui si negoziano visioni del futuro, distribuzioni di potere e modelli di responsabilità. È proprio in questa tensione (non pienamente risolvibile) tra potenza tecnologica, incertezza e controllo che prende forma il significato politico e culturale della seconda rivoluzione quantistica.

Parte III
Metodi e pratiche di ricerca
per le tecnologie emergenti

7. *Regolamentare: politiche e governance*

di *Suania Acampa, Francesco Amato, Vincenzo Esposito*

Quadri normativi europei

L'immaginario di un'innovazione “umana e sostenibile”, al di là di ogni orpello retorico, costituisce l'orizzonte politico della traiettoria europea sulle tecnologie emergenti.

Tale visione propone l'anticipazione degli impatti sociali e ambientali delle TE, responsabilizzando progettisti e utilizzatori anziché inseguire ex post gli effetti collaterali (Stilgoe et al., 2013; von Schomberg, 2013). In questa direzione, negli ultimi anni l'UE ha costruito un ecosistema regolativo sempre più articolato, che intreccia norme orizzontali e interventi settoriali con strategie industriali, programmi di Ricerca e Innovazione e strumenti di monitoraggio. Il nucleo di questo ecosistema è composto da un insieme coerente, benché eterogeneo, di regolamenti, atti e policy framework che disciplinano ambiti ad alto impatto sociotecnico: intelligenza artificiale (con particolare attenzione ai modelli generativi e di uso generale, LLM/GPAI), gemelli digitali, sistemi di decisione automatizzata (ADM) e tecnologie quantistiche.

Il percorso normativo europeo si sviluppa dunque nell'interazione tra hard law, programmazione strategica e approcci data-driven e risk-based: dal recente AI Act al pacchetto dati (Data Governance Act e Data Act), dal Digital Services Act (DSA) e Digital Markets Act (DMA) al Cyber Resilience Act, fino al Chips Act e alla traiettoria – ancora in via di definizione – verso un futuro EU Quantum Act (Parlamento e Consiglio, 2022; 2023; 2024).

Considerate nel loro insieme, queste misure restituiscono un modello europeo di governance fondato su tre pilastri ricorrenti: responsabilità documentale, tracciabilità pubblica e riduzione delle asimmetrie informative. Sono questi i principi che, più di altri, funzionano da dispositivo unificante dell'intero quadro regolativo.

L'AI Act: un quadro orizzontale basato sul rischio

L'AI Act (Reg. UE 2024/1689) rappresenta il primo tentativo al mondo di disciplinare l'intelligenza artificiale in un quadro orizzontale. Entrato in vigore nel 2024, prevede un'applicazione graduale (6-36 mesi) e introduce un approccio rigorosamente risk-based: i sistemi sono classificati in quattro livelli (rischio inaccettabile, alto, limitato, minimo) e a ciascuno corrispondono obblighi proporzionati riguardo gestione del rischio, qualità e governance dei dati, tracciabilità, supervisione umana, valutazioni di conformità e trasparenza (European Parliament and Council, 2024a).

Rispetto alle altre norme, novità decisiva è il Capitolo V, dedicato ai modelli di uso generale, ovvero modelli adattabili a molte applicazioni diverse, i più noti esempi sono i grandi modelli linguistici e i modelli generativi avanzati (GPAI). Per questi modelli, considerati a “rischio sistemico”¹, sono previste misure quali: valutazioni di sicurezza, *adversarial testing*², gestione documentata delle vulnerabilità, e sintesi pubbliche delle basi di training (art. 53). Per facilitare l'applicazione pratica di queste norme, nel 2025 la Commissione ha pubblicato linee guida e un Codice di condotta, che forniscono modelli operativi e chiarimenti su temi come copyright, trasparenza e valutazione dei rischi (Commissione europea, 2025). Per gli ADM utilizzati in sanità, istruzione, occupazione, giustizia, servizi pubblici e credito, l'AI Act incide soprattutto attraverso l'inquadramento come “alto rischio” (Allegato III): la qualità dei dati, il *logging*³, la robustezza, la supervisione umana e gli audit diventano requisiti obbligatori. Nel caso dei LLM, il nodo più delicato è la documentazione tecnica e la tracciabilità degli input, condizione necessaria affinché autorità e ricercatori possano comprendere (e contestare) come un modello è stato sviluppato, addestrato, testato e immesso sul mercato. La scelta europea è dunque quella di ancorare l'innovazione alla responsabilizzazione documentale e al principio di verificabilità pubblica, in linea con la tradizione dell'innovazione responsabile (Stilgoe et al., 2013; Parlamento e Consiglio, 2024a). L'AI Act interviene sul “livello applicativo” dell'intelligenza artificiale: ciò che un sistema può fare, con quali requisiti e quali limiti. Ma questo livello è intrecciato con due dimensioni più strutturali: la disponibilità di dati e la configurazione dei mercati digitali. Senza un ecosistema dei dati aperto e interoperabile, e senza piattaforme responsabili e contendibili, la regola-

¹ Rischio che un modello molto potente, per diffusione o capacità, generi effetti negativi ampi sulla società o su interi mercati/servizi.

² Prove “avversarie” che cercano deliberatamente di far sbagliare il modello/sistema per scoprire vulnerabilità e comportamenti indesiderati.

³ Registrazione tecnica degli eventi (input, output, errori, decisioni) utile per audit, tracciabilità e ricostruzione ex post.

zione dell'IA resta monca. Di conseguenza, il livello applicativo regolato dall'AI Act deve essere letto insieme alle regole che definiscono l'ambiente informazionale in cui quei sistemi operano: disponibilità dei dati, accesso equo, trasparenza delle piattaforme e concorrenza dei mercati.

Ecosistema dei dati, piattaforme e mercati digitali

L'idea di fondo è che la regolazione dell'IA non può funzionare senza intervenire sulle condizioni materiali e informative che la rendono possibile: chi controlla i dati, come sono accessibili, quali incentivi governano le piattaforme e come sono strutturati i mercati digitali.

Dunque l'AI Act deve dialogare con un "ecosistema regolativo dei dati" basato sul Data Governance Act (DGA) (Reg. UE 2022/868) e sul Data Act (Reg. UE 2023/2854).

Il DGA promuove la condivisione sicura dei dati e introduce figure quali gli intermediari fiduciari e il data altruism, ossia la possibilità di mettere a disposizione dati personali e non personali per scopi di interesse generale.

Il Data Act riconosce diritti di accesso e riuso ai dati generati da prodotti connessi e servizi correlati: ossia, un'azienda o una PA che utilizza, ad esempio, macchinari dotati di sensori o dispositivi IoT può ottenere (a certe condizioni) l'accesso e la portabilità dei dati generati da tali dispositivi. Questo ha impatti diretti, ad esempio, sui gemelli digitali in cui la disponibilità e la portabilità dei flussi di dati dei sensori alimentano la creazione di gemelli digitali interoperabili, abilitano nuove catene del valore e riducono situazioni di blocco (*lock-in*⁴) e asimmetrie informative⁵ tra grandi player e utenti più piccoli (Parlamento e Consiglio, 2022c; 2023a; Richterich, 2018).

Accanto all'economia dei dati, l'UE interviene anche sul funzionamento delle piattaforme e dei mercati digitali.

Il Digital Services Act (DSA) (Reg. UE 2022/2065) introduce obblighi di gestione del rischio e di trasparenza, soprattutto per le piattaforme molto grandi (ad esempio sui sistemi di raccomandazione, sulla pubblicità e sull'accesso ai dati per i ricercatori), con effetti indiretti anche su ADM, che influenzano l'esposizione ai contenuti (Parlamento e Consiglio, 2022a).

Il Digital Markets Act (DMA) (Reg. UE 2022/1925) disciplina invece i comportamenti dei grandi operatori "gatekeeper" (ad esempio, le piattaforme che controllano l'accesso a mercati/utenti), imponendo obblighi e

⁴ Situazione in cui un utente o un'impresa non può cambiare fornitore per costi tecnici/contrattuali eccessivi o per mancanza di accesso ai propri dati.

⁵ Squilibri di potere dovuti al fatto che alcuni soggetti (es. grandi piattaforme) detengono molti più dati/conoscenza di altri.

divieti per mantenere i mercati concorrenziali: in concreto, significa più interoperabilità⁶, limiti all'auto-preferenza (*self-preferencing*)⁷ e condizioni più eque per chi vuole entrare o competere nell'ecosistema (Parlamento e Consiglio, 2022b). Per i LLM e i servizi generativi integrati nelle piattaforme, DSA e DMA, incidono sia sui canali di distribuzione (es. app store, search) sia sul regime di trasparenza e ricerca e, indirettamente, sull'accesso a dati utili per il loro addestramento e valutazione (Parlamento e Consiglio, 2022a; 2022b).

L'ecosistema dei dati, insieme alla disciplina delle piattaforme e dei mercati, agisce sul versante informativo della trasformazione digitale: chi controlla i dati, come sono condivisi, quali obblighi di trasparenza governano i servizi online.

Una volta chiarito chi detiene e gestisce i dati, occorre interrogarsi su un secondo livello altrettanto cruciale: quello infrastrutturale, ossia l'insieme delle condizioni tecniche che rendono possibile sviluppare, aggiornare e mettere in sicurezza le tecnologie che utilizzano tali dati.

Sicurezza e capacità materiali

Se la dimensione informazionale definisce “cosa” può essere sviluppato e condiviso, quella infrastrutturale determina “come” questo può essere fatto in modo sicuro e scalabile.

La dimensione della sicurezza-by-design emerge con forza nel Cyber Resilience Act (Reg. UE 2024/2847), che introduce requisiti orizzontali di cybersicurezza per tutti i prodotti con elementi digitali lungo l'intero ciclo di vita. I produttori devono prevenire vulnerabilità, garantire aggiornamenti tempestivi, documentare processi e rispondere legalmente per eventuali inadempienze (Parlamento e Consiglio, 2024b). Questo è decisivo per i gemelli digitali e l'IoT – che si appoggiano a catene hardware/software complesse – e per gli LLM inseriti in dispositivi e applicazioni industriali che rendono critico il controllo degli aggiornamenti (Commissione europea, 2024).

Parallelamente, il Chips Act (Reg. UE 2023/1781) affronta il nodo della capacità materiale, perché regolare l'IA serve a poco se mancano capacità di calcolo, memoria e componentistica. Il Chips Act è stato sviluppato a sostegno della progettazione e produzione europee di semiconduttori, cruciali per LLM, HPC, gemelli digitali complessi e (nel prossimo futuro) le tecnologie quantistiche (Parlamento e Consiglio, 2023b).

⁶ Capacità di sistemi e servizi di scambiare e usare reciprocamente dati e funzionalità senza barriere.

⁷ Pratica con cui un gatekeeper favorisce i propri servizi/prodotti rispetto a quelli dei concorrenti all'interno della propria piattaforma.

Sicurezza e capacità materiali diventano quindi due pilastri complementari: senza sicurezza-by-design e senza potenza di calcolo, l'intero impianto regolativo rischierebbe di essere inefficace o non implementabile.

Uno degli elementi chiave delle politiche europee è la coerenza tra strumenti orizzontali (AI Act, DGA, Data Act, DSA/DMA, CRA, Chips Act) e verticali (strategie settoriali, missions, programmi R&I) che converge in un'idea è semplice: unire regole di base comuni a interventi mirati dove necessario. Ad esempio: per LLM e ADM, l'allineamento tra AI Act e DSA opera su almeno tre piani: (a) più trasparenza e gestione del rischio per piattaforme molto grandi che usano sistemi di raccomandazione; (b) accesso a dati e documentazione quando i modelli toccano processi "civici" (disinformazione, pubblicità politica, tutela dei minori); (c) enforcement coordinato, ossia applicazione e vigilanza condivise tra autorità nazionali e Commissione/AI Office, per evitare "buchi" di controllo. Per i gemelli digitali, il Data Act incide perché vincola progettazione e contratti al diritto di accesso/portabilità dei dati generati dai dispositivi mentre il CRA – il regolamento europeo sulla "cybersicurezza dei prodotti digitali" – rende materialmente verificabili e controllabili gli aggiornamenti di sicurezza lungo la supply chain hardware/software che sostiene LLM e gemelli digitali (Parlamento e Consiglio, 2024b).

A questo quadro ex ante si affianca la revisione del regime di responsabilità civile, che completa la governance della sicurezza dal lato ex post. La Product Liability Directive (PLD) (Direttiva UE 2024/2853) è la direttiva che aggiorna la disciplina della responsabilità per il danno da prodotti digitali difettosi, riducendo l'onere probatorio e chiarendo le responsabilità dei produttori (Parlamento e Consiglio, 2024c). Che cosa significa in concreto? Che, se un prodotto digitale (da un semplice software a un sistema di IA integrato in un dispositivo) causa un danno perché difettoso, diventa più chiaro come far valere la responsabilità del produttore e, per il danneggiato, è meno gravoso provare il nesso tra difetto e danno, anche quando il problema deriva da aggiornamenti, dati o interazioni complesse del sistema. La combinazione tra AI Act (ex ante) e PLD riformata (ex post) crea così un doppio binario, basato sulla prevenzione e sulla responsabilità.

Dal quadro normativo alla capacità attuativa: strategie e coordinamento

Definite le condizioni minime di sicurezza e le basi materiali dell'autonomia tecnologica, emerge una domanda più ampia: come coordinare e orientare queste capacità nel tempo affinché producano valore pubblico, innovazione sostenibile e strumenti concreti per la ricerca, l'industria e le amministrazioni pubbliche?

La cornice regolativa europea è infatti sostenuta da strategie di lungo periodo che ne orientano l'attuazione.

La Digital Decade Policy Programme 2030 definisce obiettivi misurabili in materia di competenze, infrastrutture e digitalizzazione della PA e delle imprese, con un ciclo annuale di cooperazione e di monitoraggio tra gli Stati membri. La New European Innovation Agenda (NEIA) che orienta politiche e finanza verso il deep tech e mira a colmare lo *scale-up gap*, cioè la difficoltà strutturale delle imprese europee a passare dalla fase di *start-up* alla crescita su larga scala per scarsità di capitali, accesso a mercati e infrastrutture adeguate favorendo così la crescita di imprese innovative attraverso strumenti finanziari e infrastrutture dedicate (Commissione europea, 2022; 2023).

Queste strategie servono a “tradurre” le norme in capacità reali, allineando obiettivi, indicatori e risorse finanziarie per sostenere l’implementazione e la crescita. La linea rimane quella di procedere secondo la tabella di marcia, con possibili semplificazioni per le PMI e chiarimenti tecnici sulla documentazione, sulla gestione del rischio e sull’interoperabilità (Commissione europea, 2025). Perché tutto funzioni davvero, servono tre condizioni: (1) capacità amministrativa (autorità nazionali e AI Office in grado di vigilare e supportare); (2) standard tecnici maturi (modelli di documentazione condivisi, benchmark comuni, strumenti di audit); (3) leve finanziarie per il deep tech (ad esempio fondi per la competitività). In assenza di queste condizioni, la norma rischia di restare un’aspirazione e di non incidere sulla realtà: ad esempio, chi dispone di molta potenza di calcolo e di grandi basi dati può mantenere un vantaggio strutturale che rende i mercati poco contendibili (difficili da “scalfire” per nuovi entranti) e le decisioni automatizzate poco contestabili (cioè difficili da verificare o mettere in discussione da parte di utenti, giudici o autorità) (European Commission, 2025).

Un esempio di questa traduzione operativa è costituito da due iniziative che concretizzano le strategie europee in capacità applicative e casi d’uso reali: i gemelli digitali e il settore quantistico.

I gemelli digitali rappresentano uno dei casi d’uso più avanzati di integrazione tra dati regolati, AI, HPC e modelli predittivi in quanto rendono operativi (e mettono alla prova) il Data Governance Act e il Data Act per quanto riguarda l’accesso e il riutilizzo dei dati, e il Cyber Resilience Act per la sicurezza nella supply chain software. Per la ricerca sociale, i twin diventano strumenti di *policy intelligence*⁸: consentono valutazioni di impatto, scenari e processi di co-progettazione tra enti pubblici, imprese e comunità (Commissione europea, 2023-2025).

⁸ Uso di simulazioni, scenari e dati integrati per supportare decisioni pubbliche basate su evidenze e valutazioni d’impatto.

Sul fronte quantistico, il 2025 segna un punto di svolta con la Quantum Europe Strategy (Commissione europea, 2025), che tratta il quantum come infrastruttura strategica (calcolo, comunicazione, sensing, metrologia) e propone una governance integrata tra ricerca, politica industriale, sicurezza economica e catene del valore. La Quantum Europe Strategy del 2025 non rappresenta un punto di partenza, ma la formalizzazione politica di un percorso avviato con il Quantum Manifesto, che già nel 2016 aveva fissato obiettivi differenziati su breve, medio e lungo periodo: maturare le tecnologie abilitanti, costruire filiere industriali affidabili e sviluppare applicazioni scalabili integrate con HPC e reti quantistiche sicure. La Quantum Europe Strategy si limita a rendere operativo questo percorso, fornendo strumenti istituzionali, standard e incentivi.

A valle della Strategia, la Commissione sta raccogliendo contributi in vista di un EU Quantum Act previsto per il 2026. In termini semplici, l'Atto mira a ridurre la frammentazione, colmare il gap di capacità industriale e rafforzare la standardizzazione (cioè l'accordo su regole tecniche comuni per garantire compatibilità e qualità), anche attraverso linee pilota – piccole linee produttive dimostrative per portare la tecnologia dal laboratorio alla fabbricazione pre-industriale – e una design facility, ossia servizi condivisi per progettare chip, componenti e sistemi quantistici accessibili a università, PMI e industrie.

Il tutto si coordina con i programmi EuroHPC ed EuroQCI. EuroHPC (supercalcolo) è il programma europeo per supercomputer e calcolo ad alte prestazioni, mentre EuroQCI (comunicazioni quantistiche) è l'infrastruttura europea per reti ultra-sicure (Commissione europea, 2025). L'orientamento politico è chiaro: colmare il divario negli investimenti privati e creare condizioni normative, infrastrutturali e finanziarie affinché le tecnologie quantistiche passino dal laboratorio al mercato, tutelando al contempo diritti, sicurezza (ad esempio, in materia di crittografia) e il controllo delle esportazioni.

In tutte queste dimensioni ricorre un tratto comune: l'UE cerca di governare le TE non attraverso divieti o accelerazioni indiscriminate, ma tramite dispositivi che rendono visibili, documentabili e verificabili le scelte sociotecniche che strutturano l'innovazione.

Tirando le fila, i quattro ambiti tecnologici al centro di questo volume evidenziano la coerenza e le sfide della visione europea.

Per i LLM/GPAI, l'AI Act e le Linee Guida fissano una grammatica operativa della trasparenza (documentazione tecnica standardizzata, sintesi delle basi di training, gestione delle vulnerabilità, *incident reporting* e test proporzionati al rischio), aprendo spazi di valutazione pubblica e di ricerca indipendente (Parlamento e Consiglio, 2024a; Commissione europea, 2025).

Per gli ADM nei settori sensibili, l'inquadramento "alto rischio" estende le garanzie procedurali (qualità dei dati e *logging*, supervisione umana effettiva⁹, controllo delle decisioni), mentre il DSA rende più trasparenti i sistemi di raccomandazione (Parlamento e Consiglio, 2022a; 2024b). Per i digital twin, l'economia dei dati regolata (DGA/Data Act) e la cybersecurity-by-design (CRA) consentono gemelli interoperabili e sicuri, con catene di responsabilità chiare per produttori e operatori. Per le tecnologie quantistiche, la Quantum Europe Strategy e la consultazione sull'EU Quantum Act prefigurano una governance che integra politiche industriali, standardizzazione, investimenti pubblico-privati e sicurezza economica, con obiettivi di ibridazione tra HPC e Quantum e di scalabilità applicativa.

In sintesi, il quadro normativo e strategico europeo non isola l'innovazione dai vincoli sociali: la incanala entro una razionalità di responsabilità che pretende evidenze, tracciabilità e controllo pubblico (Stilgoe et al., 2013; von Schomberg, 2013).

Se l'immaginario di un'innovazione "umana e sostenibile" ne fornisce il lessico politico, sono i dispositivi giuridici e programmatici a renderlo operante. Qui i prossimi strumenti del capitolo (sandbox e framework sperimentali; accountability e trasparenza) troveranno il loro spazio operativo, traducendo l'ambizione regolativa in pratiche di prova, di adattamento e di apprendimento in contesti reali ad alto impatto sociale.

Uno strumento di regolamentazione e sperimentazione: le sandbox

La discussione sulle regulatory sandbox per l'intelligenza artificiale si iscrive in una trasformazione più ampia del diritto europeo della regolazione, sempre più orientato a intervenire in modo anticipatorio lungo il ciclo di sviluppo delle tecnologie. Le sandbox operano come strumenti attraverso cui le autorità possono osservare, accompagnare e orientare lo sviluppo tecnologico prima della piena immissione sul mercato.

È precisamente in questa prospettiva che si colloca la proposta di Truby et al. (2022), che interpretano la sandbox come componente strutturale di un'architettura normativa europea volta a governare applicazioni di IA ad alto rischio in modo armonizzato e prevedibile. La regolazione dell'IA incontra difficoltà che i modelli giuridici tradizionali faticano ad affrontare. La fault-based liability, che richiede la prova della colpa da parte del

⁹ Presenza di una persona in grado di comprendere, intervenire ed eventualmente "fermare" un sistema automatizzato.

soggetto danneggiato, si scontra con profonde asimmetrie informative, mentre la strict liability, che prescinde dall'accertamento della colpa, tende a trasferire integralmente su sviluppatori e produttori il peso della conformità e dell'incertezza. In presenza di sistemi opachi, adattivi e distribuiti, diventa centrale definire quali condizioni regolamentari debbano essere predisposte prima del danno, affinché il rischio possa essere osservato, misurato e corretto.

Gli autori cercano di mostrare come le sandbox svolgano una funzione insieme conoscitiva e normativa: conoscitiva perché permettono alle autorità di comprendere tecnologie difficili da disciplinare attraverso categorie astratte; normativa perché consentono una definizione progressiva e contestualizzata degli standard di sicurezza, trasparenza e tutela dei diritti.

Come mostrato da uno studio di Qiu et al. (2025), le esperienze maturate nel settore FinTech, in particolare i modelli della Financial Conduct Authority nel Regno Unito e della Monetary Authority of Singapore, hanno mostrato come sia possibile creare spazi in cui operatori innovativi testano prodotti in condizioni reali ma limitate, sotto supervisione regolatoria. Nel caso dell'IA, tuttavia, questi strumenti assumono una portata più ampia, incidendo su diritti fondamentali, processi decisionali e distribuzione del rischio. Una sandbox per l'IA non può dunque limitarsi a replicare il modello FinTech, ma richiede un progetto istituzionale capace di tenere insieme supervisione tecnica, coordinamento tra autorità e uniformità del quadro europeo. L'esperienza dell'AI Airlock britannico, sviluppata dalla Medicines and Healthcare products Regulatory Agency con il National Health Service, illustra come la sandbox possa servire a definire progressivamente criteri di valutazione e standard applicativi per tecnologie difficili da inquadrare attraverso i soli strumenti tradizionali (Qiu et al., 2025).

Implementazioni non coordinate a livello nazionale possono generare divergenze applicative e incentivi alla localizzazione strategica delle attività sperimentali. Il rischio di una pluralità non coordinata di sandbox nazionali non segnala soltanto inefficienza amministrativa, ma la necessità di ancorare questi strumenti a una logica autenticamente europea della regolazione, capace di produrre criteri condivisi, comparabilità tra pratiche e interoperabilità tra autorità.

La letteratura e le esperienze applicative (Truby et al., 2022; Qiu et al., 2025) convergono nel descrivere le sandbox come infrastrutture regolative distribuite lungo il ciclo di vita dei sistemi. Nella fase pre-applicativa, esse offrono uno spazio di orientamento anticipato nel quale imprese e autorità contribuiscono alla definizione delle metriche di valutazione. Nella fase di sperimentazione, strumenti e pratiche come i gemelli digitali e l'*auditing by design* consentono una sorveglianza dinamica capace di intercettare

criticità e permettere correzioni in corso d'opera. Nella fase di valutazione e uscita, emerge l'ipotesi di un modello di accreditamento integrato, sintetizzabile nell'idea di un passaporto europeo per l'IA, che possa associare parametri tecnici a indicatori etici, sociali e giuridici. Il monitoraggio post-sandbox, infine, prolunga nel tempo la logica regolativa della sperimentazione, trasformandole da episodio isolato a primo segmento di un dispositivo più ampio di sorveglianza adattiva.

Nel complesso, il significato delle regulatory sandbox nell'ambito dell'AI Act può essere colto adeguatamente se le si considera come strumenti di una regolazione non soltanto prescrittiva e sanzionatoria, ma anticipatoria, iterativa e capace di apprendere. La loro efficacia dipende dall'integrazione in un quadro europeo coordinato e dalla capacità di produrre conoscenza regolativa condivisa. È in questo senso che la sandbox assume rilievo come spazio in cui la regolazione non arriva dopo l'innovazione, ma si sviluppa con essa.

Anticipare la tecnologia per colmare il ritardo della regolazione

La produzione di regolamenti, policy e linee guida è frutto di complessi e intricati processi che coinvolgono attori istituzionali, esperti e stakeholders che richiede tempo, competenze specialistiche e risorse significative. Oggi, tuttavia, le TE pongono una sfida inedita a tali processi. I sistemi digitali e algoritmici, caratterizzati da una complessità crescente e da un'estrema opacità, non sono più entità tangibili e ispezionabili nel senso tradizionale: a differenza di una macchina o di un impianto industriale, non è possibile osservarne il funzionamento dall'esterno, smontarne i componenti o verificarne il comportamento attraverso un esame diretto. Non si tratta più soltanto di oggetti fisici, ma sempre più spesso di insiemi di codice, flussi di dati e sequenze di istruzioni che operano in modo dinamico, distribuito e spesso non deterministico, producendo effetti difficili da tracciare, attribuire o prevedere con gli strumenti concettuali tradizionali. Comprendere e analizzare questi sistemi richiede la padronanza di linguaggi tecnici e competenze interdisciplinari aggiornate, capaci di interpretare i principi logici e operativi che ne regolano il funzionamento.

È proprio questa complessità strutturale a rendere il processo regolatorio intrinsecamente lento e difficoltoso. La produzione di norme, standard e linee guida dipende da tempi istituzionali, negoziazioni politiche e procedure di consenso che difficilmente riescono a tenere il passo con la velocità dei processi tecnologici. Quando la tecnologia è ulteriormente protetta da barriere di accesso tecniche, giuridiche o proprietarie, come nel caso dei

sistemi di intelligenza artificiale basati su modelli chiusi o brevettati, la distanza tra ciò che accade e ciò che può essere regolato si amplifica in modo significativo.

Questa distanza temporale e conoscitiva non è solo un problema di efficienza amministrativa, ma rappresenta una questione di potere cognitivo, sociale, politico, tecnico ed economico. Chi arriva per primo nel campo dell'innovazione non solo definisce le regole del mercato, ma contribuisce anche a plasmare le categorie attraverso cui quella tecnologia viene compresa e valutata. Quando gli organismi regolatori intervengono tardi, si trovano ad operare all'interno di un orizzonte semantico e concettuale già costruito da altri, dagli attori economici e tecnologici che hanno promosso e guidato l'innovazione. Giungere in ritardo significa, in altri termini, adottare la visione di chi ha già vinto la corsa, accettando implicitamente la narrazione dominante sui benefici e i rischi del nuovo.

Regolamentare la tecnologia e il progresso tecnologico diventa dunque una sfida, un obiettivo verso cui puntare al fine di ridurre questa distanza e partecipare attivamente alla costruzione del futuro tecnologico. Non si tratta di un esercizio predittivo, ma di una forma di preparazione epistemica e politica, ovvero dotarsi progressivamente degli strumenti interpretativi necessari per comprendere il cambiamento mentre esso avviene, identificando precocemente le traiettorie emergenti, i segnali deboli e le aree di rischio. Agire in questa fase precoce non equivale a limitare o ostacolare lo sviluppo tecnologico; al contrario, significa riconoscere che la governance dell'innovazione deve diventare parte integrante del processo innovativo stesso, e non un intervento correttivo *ex post*. I metodi di ricerca anticipatoria, come la costruzione di scenari o la consultazione sistematica degli esperti, offrono in questo senso strumenti preziosi per intercettare le direzioni del cambiamento, esplorare le alternative possibili e sviluppare forme di governance più adattive e partecipate.

In questa prospettiva, regolamentare la tecnologia significa non lasciarsi trascinare dall'entusiasmo o dall'urgenza della novità, ma coltivare una capacità collettiva di visione, in grado di orientare il progresso tecnologico alla luce di specifici orizzonti normativi e visioni di giustizia sociale. In una società dove a prevalere è spesso il primo che arriva, costruire questa capacità di anticipazione rappresenta un atto di autonomia culturale e politica, ed è il modo più efficace per evitare che le traiettorie del futuro siano decise altrove.

8. *Immaginare: narrazioni, promesse e visioni del futuro*

di *Suania Acampa, Dario Chianese*

Gli immaginari sociotecnici come quadro teorico per indagare le TE

Le TE rappresentano uno dei principali terreni su cui oggi la società proietta le proprie aspettative, formula promesse, elabora timori e costruisce orientamenti per il futuro. Nel dibattito pubblico, esse compaiono come catalizzatori dell'immaginazione collettiva, poiché condensano visioni sul destino sociale, economico, ambientale e politico delle comunità. Questo fenomeno è attualmente molto evidente nell'intelligenza artificiale, ma emerge anche nelle tecnologie quantistiche, nei grandi sistemi di automazione, nelle infrastrutture di calcolo avanzate, e riguarda quindi l'intero dominio dell'innovazione contemporanea. Ciò che accomuna questi ambiti è il modo in cui diventano oggetti narrativi e dispositivi simbolici che pongono in evidenza quanto l'innovazione sia parte di un ampio processo, in cui aspettative, visioni e significati si intrecciano con le capacità tecniche e materiali, delineandone possibilità e limiti.

Questa prospettiva affonda le sue radici negli STS, che mostrano come scienza e tecnologia si sviluppino all'interno di configurazioni sociali complesse, in cui interessi istituzionali, significati culturali e rapporti di potere modellano gli artefatti e ne sono, al tempo stesso, modellati (Jasanoff, 2015; Law, 2008). In questo quadro si inserisce il concetto di co-produzione proposto da Jasanoff (2004), che permette di cogliere la natura reciproca dei processi sociotecnici: le tecnologie emergenti prendono forma attraverso norme, istituzioni, narrazioni e aspettative, e allo stesso tempo contribuiscono a rafforzare visioni collettive e priorità politiche. È su questo terreno che Jasanoff e Kim (2015) collocano la nozione di "immaginari sociotecnici", definendoli come visioni condivise di futuri ritenuti desiderabili (o da evitare), normative ma radicate nella materialità delle infrastrutture. Cantó-Milà e Seebach (2024) li interpretano come configurazioni relazionali di

immagini, contenuti e forme che producono aspettative e orientano ciò che una collettività ritiene possibile, mettendo in evidenza la natura dinamica e conflittuale di questi campi, attraversati da posizioni egemoniche e marginali. Taylor (2004) invita a distinguere tra immaginazione individuale e immaginari collettivi, intesi come insiemi condivisi di credenze su ciò che è e ciò che dovrebbe essere: una distinzione che chiarisce perché gli immaginari sociotecnici, pur radicati nella capacità proiettiva degli attori, assumano una dimensione istituzionale che eccede la sfera personale. In questa direzione, diversi autori (Beckert, 2016; Konrad, Böhle, 2019) hanno mostrato come le aspettative svolgano un ruolo costitutivo nella definizione dei percorsi di innovazione e nelle forme di coordinamento dell'azione collettiva, operando come dispositivi anticipatori che stabiliscono i parametri attraverso cui alcuni percorsi appaiono maturi, altri marginali o non allineati alle priorità istituzionali. La performatività degli immaginari deriva da questa capacità di orientare le traiettorie tecniche a partire dalle condizioni scientifiche e materiali del presente. Van Lente (2016) evidenzia come, in virtù del loro carattere anticipatorio, gli immaginari legittimino alcune direzioni di sviluppo e ne marginalizzino altre, rendendo certe priorità più visibili e altre meno. Il risultato è un campo di “futuri contestati”, come suggeriscono Brown, Rappert e Webster (2000), in cui visioni differenti coesistono e competono per imporsi, oppure un insieme di “futuri in divenire” (Halford, Southerton, 2023), continuamente modellati da decisioni, pratiche e innovazioni che si producono nel presente.

Un caso emblematico, utile a comprendere questi meccanismi, è la vicenda nucleare in Italia degli anni Ottanta. Prima dell'incidente di Chernobyl, il nucleare era al centro di narrazioni contrastanti: per una parte del mondo politico e industriale rappresentava una via verso l'autonomia energetica e la modernizzazione (Baracca, Ferrari, Renzetti, 2017; Senato della Repubblica, Dossier 273, 2011), mentre movimenti ambientalisti e settori dell'opinione pubblica lo descrivevano come un rischio difficilmente controllabile, in linea con quanto teorizzato da Douglas e Wildavsky (1982) sul ruolo culturale delle percezioni del rischio. L'evento del 1986 trasformò radicalmente il campo discorsivo. I media italiani dedicarono un'attenzione straordinaria alle conseguenze del disastro, alimentando la costruzione di un immaginario del nucleare come tecnologia precaria e potenzialmente devastante (Triandafyllidou, 1995; Corriere della Sera, 2022). Le narrazioni politiche si riconfigurarono rapidamente: alcuni attori – soprattutto nelle aree progressiste – enfatizzarono la necessità di ripensare l'intero modello energetico, mentre altri difesero l'impianto programmatico precedente, insistendo sulla sicurezza e sulla riduzione della dipendenza energetica (Corriere della Sera, 2023). Il referendum del 1987 avvenne all'interno di questa

riconfigurazione discorsiva: la moltiplicazione delle narrazioni negative, la forza simbolica di Chernobyl e la pressione esercitata da movimenti ambientalisti e attori collettivi (La Repubblica, 2022) favorirono una chiusura interpretativa a sfavore del nucleare. Il risultato, ossia la decisione di dismettere i programmi esistenti, fu il prodotto di un processo in cui l'evento, la sua mediatizzazione, la mobilitazione politica e la trasformazione dell'immaginario si intrecciarono fino a ridefinire completamente la traiettoria tecnologica nazionale (World Nuclear Association, Italy - Country Profile, 2024). In quella fase, l'intero dibattito pubblico si articolò attorno a un'infrastruttura di rappresentazioni, aspettative sociali e visioni del futuro nella quale – per dirla con Beck (1992) – il rischio tecnico si trasformò in rischio sociale, diventando un tema centrale di confronto politico e culturale.

Le narrazioni sono quindi il principale veicolo attraverso cui gli immaginari si formano e si consolidano. Ricerche recenti mostrano che le narrazioni contribuiscono a configurare la percezione collettiva delle tecnologie (Katzenbach, 2021; Bareis, Katzenbach, 2022), influenzano le priorità politiche (Macnaghten, Guivant, 2020) e condizionano le fasi di progettazione e implementazione (Rosyidah, Fajriyah, 2022; Guenduez, Mettler, 2023). Altri studi (Borup et al., 2006; Felt, 2009; Levidow, Papaioannou, 2013; Jasanoff, Kim, 2015) mostrano quanto anche la governance delle tecnologie emerge come un processo intrinsecamente narrativo: attraverso i racconti pubblici vengono selezionati temi e valori, articolate visioni normative e anticipate le possibili conseguenze dell'innovazione.

Le narrazioni svolgono così una funzione di riduzione della complessità, selezionando gli elementi ritenuti rilevanti, organizzandone il significato e collegandoli a possibili esiti futuri (Bazzani, 2023). Inoltre, la dimensione narrativa ha anche una valenza territoriale: Esposito et al. (2021) richiamano l'attenzione sui rischi connessi alle narrazioni standardizzate, che spesso proiettano la tecnologia come ineluttabile e globalmente uniforme, ignorando i contesti locali e le loro esigenze.

Le narrazioni che accompagnano le tecnologie emergenti oscillano sempre tra registri utopici e distopici. Nei discorsi istituzionali e industriali la tecnologia viene frequentemente presentata come soluzione lineare a problemi complessi, che vanno dalla crisi climatica alla sicurezza, alimentando ciò che Morozov (2013) definisce “soluzionismo tecnologico”. Allo stesso modo, Dourish e Bell (2014) parlano di “*solutionist imaginaries*”, per descrivere quelle visioni che attribuiscono alla tecnologia la capacità di risolvere quasi automaticamente problemi complessi, mentre Campolo e Crawford (2020) mostrano come tali discorsi possano assumere forme di vero e proprio “*enchanted determinism*”, in cui l'innovazione appare inevitabile e dotata di una forza quasi magica.

Questo processo è spesso mediato da metafore e figure retoriche che strutturano il modo in cui interpretiamo la tecnologia (Lakoff, Johnson, 2008; Wyatt, 2017), contribuendo alla costruzione di narrazioni moderniste di progresso (Brown, Rappert, 2017) e, talvolta, alla retorica del sublime tecnologico, in cui il potenziale trasformativo viene esaltato mentre tensioni e contraddizioni restano sullo sfondo (Nye, 1996; Marx, 2000). Queste narrazioni utopiche tendono spesso ad antropomorfizzare le tecnologie, attribuendo loro intenzioni, agency e coerenza interna, oscurandone le dinamiche e i rischi sociali, politici ed economici che ne guidano la costruzione (Natale, Ballatore, 2020).

Accanto a queste visioni utopiche, le narrazioni pubbliche possono assumere forme critiche o apertamente distopiche, mettendo in evidenza vulnerabilità, implicazioni democratiche e rischi inattesi. In questi casi, le narrazioni diventano strumenti di contestazione collettiva che contribuiscono a ridefinire l'immaginario associato a una tecnologia, come mostrano processi di partecipazione pubblica, consultazioni e pratiche deliberative orientate a esplorare alternative e a problematizzare i costi sociali dell'innovazione (Miller, 2005; Tan, 2022). Come osserva Schmidt (2008), la forza delle narrazioni risiede proprio nella loro capacità di creare connessioni tra istituzioni e comunità, offrendo chiavi interpretative utili ad affrontare l'incertezza e aprendo possibilità per immaginare futuri diversi.

Indagare le TE significa quindi analizzare i processi che le costituiscono come fenomeni sociotecnici complessi, in cui immaginari, narrazioni e definizioni contribuiscono a modellare orientamenti e quadri interpretativi, intrecciando dimensioni materiali, simboliche e istituzionali. Comprendere questi processi implica interrogarsi su come gli immaginari sociotecnici prendano forma attraverso discorsi, pratiche e infrastrutture; il passo successivo consiste dunque nel dotarsi di strumenti che permettano di osservare empiricamente tali articolazioni.

I paragrafi che seguono delineano alcune traiettorie metodologiche ed esempi di ricerca utili allo studio della produzione, della circolazione e della trasformazione delle narrazioni e degli immaginari futuri delle TE.

Gli immaginari dell'IA e i suoi futuri

Le TE sono tanto legate alle narrazioni che le riguardano quanto agenti produttivi di narrazioni su di esse e sul mondo sociale, è possibile intercettare le loro narrazioni sia esaminandole empiricamente che interagendo direttamente con le tecnologie in questione. In questo contesto l'IA rappresenta un caso interessante, poiché è al contempo un aggregato

di tecnologie, di nozioni teoriche e pratiche, e un fenomeno culturale su vasta scala.

Con alcuni sistemi di IA è possibile, in modo evidente, apprezzare il loro contributo alla formazione e rielaborazione di immaginari: talvolta in maniera reattiva, ossia rispondendo ai comandi di un agente umano, talvolta in maniera completamente autonoma, essi coproducono narrazioni, futuri possibili ed auspicabili, e rinforzano specifiche configurazioni narrative sul loro conto.

Mentre nel secolo scorso l'IA era raccontata e vissuta come un fenomeno avveniristico, futuribile, ma lontano dall'esperienza quotidiana, la diffusione di dispositivi informatici personali la rende oggi meno distante, ed è anzi frequentemente invocata nei discorsi pubblici come in quelli privati. L'immaginario dell'IA odierna si è ricostituito attorno alla cultura materiale e alle pratiche di produttori e utenti, ingaggiando tanto i valori, le credenze e le visioni future che ne hanno delineato a lungo la portata quanto le interazioni con essa che permeano, con diversi gradi di evidenza, il vissuto individuale.

Attraverso l'analisi di testi sia di finzione che di natura speculativa, Cave e Dihal (2019) hanno riconosciuto alcuni dei temi più rilevanti nelle narrazioni sull'IA, frutto della tensione fra visioni ottimiste e pessimiste del ruolo delle TE nei futuri immaginati:

Le applicazioni in ambito medico, potenzialmente rivoluzionarie per la salute e la longevità del genere umano, ma che al contempo minano l'integrità dei corpi; la liberazione dal lavoro grazie all'automazione, che può anche divenire una fonte di crisi laddove i lavoratori non avessero più le garanzie ed il sostentamento che il lavoro fornisce loro; la contrapposizione fra gratificazione dei bisogni e dei desideri e alienazione dalla natura umana e fra gli individui; infine, il dominio e il controllo resi possibili dall'utilizzo dell'IA, dinanzi al rischio di rivolte e azioni ostili da parte dell'IA stessa, mediante suoi sistemi, applicazioni o macchine sotto il suo controllo (Cave, Dihal, 2019).

La circolazione di questi temi nell'immaginario collettivo riecheggia poi nelle comunicazioni degli attori coinvolti nello sviluppo e utilizzo dei sistemi di IA e, in esse, le tensioni fra potenzialità e rischi sono mobilitate per dotare le tecnologie di una maggiore importanza, se non urgenza, e giustificare così l'attenzione nei loro confronti (Joyce et al., 2021).

Dall'analisi delle strategie nazionali – documenti ufficiali che delineano le azioni da compiere su un determinato tema – sull'IA dei governi di diversi paesi emergono retoriche simili rispetto all'inevitabilità dell'IA come aggregato di tecnologie dirompenti, nonché la necessità di adottare dette tecnologie per poter competere a livello internazionale e garantire be-

nefici alla cittadinanza; al contempo, gli immaginari e le promesse in essi dichiaratamente legate allo sviluppo dell'IA si declinano diversamente a seconda dei bisogni e dei desideri della nazione, proponendo visioni future distinte la cui stessa articolazione prefigura un restringimento dei futuri possibili, orientando le opzioni e le decisioni politiche (Halford, Southerton, 2023).

Attori istituzionali, politici ed industriali sono sovente latori di simili narrazioni soluzioniste nei riguardi dell'IA, tramite il cui progresso tecnologico si potrebbero arginare o risolvere le grandi crisi del nostro tempo, ottenere vantaggi strategici contro nazioni antagoniste e produrre nuovo valore a partire dalla mole di dati digitali e documenti esistente e in costante aumento (Crawford, 2021; Pink, 2025); queste ed altre ragioni supportano scelte di policy, investimenti e comportamenti individuali a diversi livelli, e perciò la copertura mediatica e la disseminazione di narrazioni sull'IA, al pari di altre TE, restano un obiettivo importante per il ricercatore interessato all'immaginario tecnologico e alla produzione concorrente di futuri.

Molti studi sull'IA come tema culturale partono, perciò, dall'analisi della narrativa di finzione e delle trattazioni speculative, dalle rilevazioni di opinioni o dall'analisi di documenti istituzionali. Nella tradizione STS, l'approccio etnografico allo studio delle TE continua ad essere estremamente proficuo, seppur costellato da difficoltà operative e metodologiche: la possibilità di fare esperienza diretta del lavoro di co-produzione di narrazioni ed immaginari, nonché delle pratiche situate che con essi interagiscono, permette di giungere a osservazioni circostanziate e precise rispetto al fenomeno in questione, ma deve misurarsi con la complessità degli assemblaggi attraverso cui si articola. Nel caso dell'IA, ciò vuol dire confrontarsi con un mastodontico apparato produttivo e di consumo, a partire dall'estrazione ed il trattamento di dati e risorse, passando per infrastrutture di calcolo, organizzazioni dedicate all'addestramento e al dispiegamento di modelli, e un vasto repertorio di interfacce di diverso tipo con la loro utenza (Aragona, Arvidsson, Felaco, 2020; Crawford, 2021). Una accurata selezione del campo deve tener conto del reticolo di interazioni che lascia inosservate, focalizzandosi su uno o più degli aspetti caratteristici del fenomeno, ad esempio l'interazione fra utenti e piattaforme, l'evoluzione normativa a disciplina dello sviluppo dei sistemi di IA, o la produzione mediatica a loro contorno.

Più di recente, grazie alla maturazione dell'interrogazione dei modelli di IA come approccio alla portata delle scienze sociali, sono i sistemi stessi ad essere ascoltati rispetto alle loro rappresentazioni dei temi più svariati, non ultimo proprio quello dell'IA. La necessità di porre sotto scru-

tinio le produzioni testuali dei LLM, oggi frequentemente utilizzati come supporti alla generazione di idee, testi e argomentazioni, è una delle ragioni per cui ricorrere all'interrogazione dei modelli. Questo è quantomai importante quando in gioco vi sono le narrazioni attorno al futuro del mondo contemporaneo, che operano nel presente ridefinendo le traiettorie possibili, plausibili o auspicabili di un dato fenomeno nel tempo (Pink, 2025). Interrogando uno o diversi modelli su specifici scenari futuri è possibile rilevarne indirettamente le propensioni verso determinate scelte terminologiche, argomenti preferiti e pratiche adeguate nella produzione di futuri.

In uno studio sulle rappresentazioni del futuro dei LLM, diversi modelli sono stati adoperati per generare scenari futuri mediante l'utilizzo di due metodi affermati nei future studies, la *"futures wheel"* ed il *"backcasting"*: questi due metodi, solitamente utilizzati con soggetti umani, richiedono loro di immaginare, nel caso del primo, le conseguenze di un determinato evento nel futuro, e le condizioni e le operazioni utili a giungere a quell'evento nel caso del secondo. Quando interrogati in questo modo su scenari legati al rapporto fra l'IA ed il futuro del lavoro, i modelli differiscono fra loro nelle risposte rispetto a diversi fattori; la focalizzazione tematica, selezionando o dando priorità diverse a temi quali la fiducia, l'innovazione tecnologica, l'implementazione di linee guida o di programmi di formazione; la lunghezza delle risposte, marcatamente diversa per alcuni modelli rispetto ad altri; la scelta terminologica, privilegiando talvolta un lessico ad un alto livello di generalità ed in altri casi adottando un gergo tecnico e specialistico; nella procedura di elaborazione degli scenari, operando indipendentemente scelte sul numero di effetti ed elementi da includere nella risposta o sull'estensione della finestra temporale da considerare (Chianese, 2026). In maniera simile, generando scenari sul futuro dell'Europa a partire da scenari esistenti, un altro studio ha evidenziato come gli scenari prodotti da LLM differissero marcatamente da quelli prodotti da esseri umani, più orientati a questioni tecniche e meno ad esperienze di vita vissuta, ma al contempo, quando esaminati alla cieca da esperti del settore tramite un panel Delphi, fossero difficili da differenziare da scenari di origine umana (Kozachek, 2026).

Grazie a queste interrogazioni, un elemento ulteriore giunge all'attenzione del ricercatore: le disposizioni e le inclinazioni dei sistemi permettono di studiarne l'agency e arricchiscono il quadro teorico degli studi sulle tecnologie, che ora si può dotare della relazione fra la proiezione di specifiche narrazioni ed immaginari sullo sviluppo dell'IA e la loro espressione, diretta e indiretta, in quegli stessi sistemi in cui risiedono.

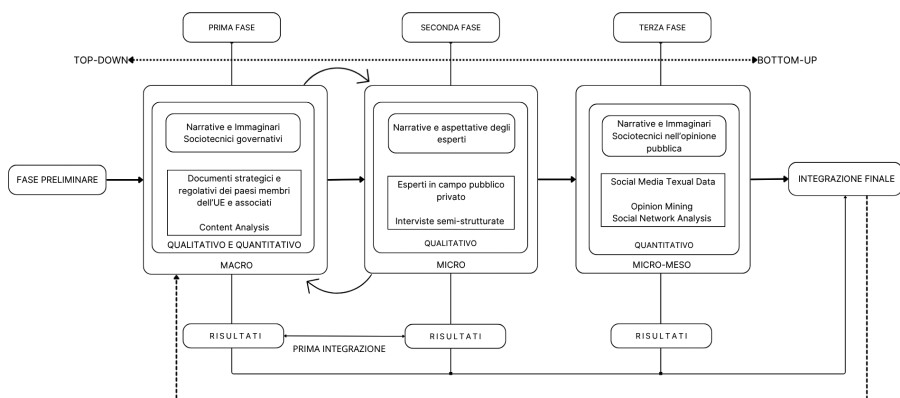
Indagare gli immaginari sociotecnici delle tecnologie quantistiche: un caso studio

Così come per l'IA, studiare come nascono, circolano e si trasformano le narrazioni e gli immaginari che accompagnano le tecnologie quantistiche richiede particolare attenzione alle loro caratteristiche di tecnologie ad alta incertezza epistemica: un insieme di soluzioni non ancora mature, in fase di sviluppo, caratterizzate da un'elevata complessità scientifica e da un grado di incertezza che rende difficile anticiparne usi, tempi e impatti. Le promesse, le visioni e i futuri immaginati rappresentano vere e proprie forze che agiscono nel presente e plasmano il modo in cui la tecnologia si sviluppa. Quando si indagano oggetti di questo tipo, la ricerca deve attrezzarsi per cogliere sia la componente discorsiva e anticipatoria, sia il modo in cui queste narrazioni vengono prodotte, tradotte e riconfigurate da attori diversi. Non esiste un'unica prospettiva in grado di restituire da sola la complessità del fenomeno: è necessario integrare strumenti e mettere in dialogo approcci diversi. L'integrazione metodologica è necessaria affinché sia possibile ricostruire la dimensione discorsiva dell'emergere delle TQ e coglierne le dinamiche di trasformazione attraverso una pluralità di fonti e attori, condizione necessaria per comprendere la natura intrinsecamente ibrida e performativa delle TQ.

In un progetto di ricerca¹⁰ condotto da chi scrive, che qui utilizzeremo come esempio di buona pratica di ricerca sulle TE, è stata adottata proprio questa prospettiva integrata, trattando le TQ come un vero e proprio oggetto sociotecnico: un oggetto composto da discorsi, attori, infrastrutture materiali e aspettative sociali. Un oggetto che cambia nel tempo e che richiede strumenti capaci sia di seguirne le trasformazioni, sia di renderle comparabili. In termini generali, il lavoro mira a costruire un quadro metodologico in grado di identificare e analizzare in modo sistematico gli immaginari sociotecnici che si articolano attorno alle TQ, mostrando come tali visioni collettive del futuro vengano costruite, stabilizzate e trasformate attraverso diversi livelli della sfera pubblica e diversi tipi di attori. L'intento è colmare il divario tra un uso prevalentemente teorico del concetto di immaginario e la sua traduzione in una categoria empirica analizzabile, proponendo strumenti e passaggi operativi per tracciarne la produzione e la circolazione. Più nello specifico, la ricerca mirava a una serie di obiettivi

¹⁰ Progetto "Tracing Sociotechnical Imaginaries: A Methodological Framework to Study Emerging Technologies" Dipartimento di Eccellenza SCDT Southern Centre for Digital Transformation (CUP: E63C22003740006) - Dipartimento di Scienze Sociali, Università degli Studi di Napoli Federico II

analitici interconnessi: in primo luogo, intendeva individuare quali immaginari sociotecnici emergono dalle narrazioni che circondano le TQ in diversi ambienti discorsivi; in secondo luogo, si proponeva di ricostruire quali attori partecipassero alla co-costruzione di tali immaginari e quali ruoli, posizionamenti e risorse di potere assumono. Un terzo obiettivo riguardava l'analisi dei legami tra immaginari e governance: comprendere in che modo quadri istituzionali, strategie nazionali ed europee, discorsi esperti e percezioni pubbliche contribuiscano a definire le priorità di intervento, i modelli regolatori, le modalità di legittimazione delle TQ e le forme di inclusione/esclusione rispetto ai possibili futuri tecnologici. Il disegno di ricerca costruito per rispondere a questi obiettivi (figura 1) si articolava quindi in tre piani della sfera pubblica, istituzionale (macro), esperti (micro) e piattaforme digitali (meso/micro), che consentono di osservare come le narrazioni si generino, si sedimentino e si riconfigurino man mano che attraversano livelli diversi del discorso. La prospettiva multilivello consente di mettere a fuoco tanto i processi top-down, legati alle strategie e alle politiche pubbliche, quanto quelli bottom-up, che emergono dalle pratiche discorsive diffuse e dalle dinamiche delle piattaforme digitali.



concettuali, gli attori, le strategie e le visioni che i decisori pubblici utilizzano per immaginare e governare le TQ. La raccolta dei documenti si è basata su procedure rigorose: identificazione delle fonti primarie (portali istituzionali europei, ministeri, agenzie per l'innovazione di ogni paese europeo), utilizzo di combinazioni di parole chiave per intercettare le strategie nazionali e le politiche dedicate alle TQ e applicazione di criteri di inclusione volti a selezionare documenti coerenti con l'obiettivo della ricerca, per natura programmatica, rilevanza tematica e collocazione temporale. Una volta raccolto il corpus di documenti, è stato analizzato mediante la costruzione di una scheda di Analisi del Contenuto, così da integrare la profondità qualitativa della lettura con la possibilità di aggregare i risultati in matrici e visualizzazioni comparabili. Le schede comprendevano cinque aree principali: 1) contesto (anno di pubblicazione del documento, titolo, autore, numero di pagine); 2) definizioni della tecnologia contenute nel documento; 3) obiettivi della tecnologia; 4) pratiche di governance; 5) attori coinvolti nell'implementazione. Questo approccio quanti-qualitativo ha permesso di condurre analisi di corrispondenza, *network analysis* e analisi tematiche, mettendo in luce differenze nazionali, convergenze, modelli emergenti e tensioni tra le diverse visioni governative ed economiche nel contesto europeo. I documenti istituzionali si sono rivelati vere e proprie infrastrutture discorsive: artefatti che contribuiscono a dare forma al campo, definendo priorità, ruoli, investimenti e rischi (Aragona, Acampa, 2025). La seconda fase ha spostato l'attenzione sugli esperti, attori fondamentali per comprendere come le narrazioni sulle TQ si costruiscano nel vivo dei processi di ricerca e innovazione. Gli esperti agiscono come mediatori tra scienza, industria e policy: raccolgono questioni e aspettative da questi ambiti e, al contempo, li influenzano attraverso le proprie visioni e competenze. Per questo motivo, le interviste semi-strutturate sono state una componente centrale del disegno. La selezione degli interlocutori è avvenuta tramite campionamento a valanga e ha permesso di includere ricercatori, professionisti dell'industria, e tecnici del settore pubblico e privato. Le interviste sono state condotte online e hanno seguito una traccia articolata in otto dimensioni, che riprendevano e ampliavano i risultati della fase precedente: definizione tecnologica, ricerca e innovazione, competenze, governance, rischi, impatto sociale, visioni del futuro, etica e regolazione. L'analisi tematica ha messo in evidenza convergenze e scarti rispetto alle narrative istituzionali, oltre a far emergere con chiarezza il ruolo delle aspettative, delle incertezze e delle promesse nella costruzione della tecnologia. La terza fase ha infine affrontato la dimensione più fluida e dinamica delle narrazioni: la loro circolazione nelle piattaforme digitali. I social media rappresentano un ambiente privilegiato per osservare come

gli immaginari si trasformano quando entrano in contatto con attori diversi, forme di engagement specifiche e logiche algoritmiche che influenzano la visibilità e la rilevanza dei contenuti. L'analisi ha riguardato più di un milione di contenuti pubblicati su Facebook e Instagram tra il 2018 e il 2024. L'arco temporale è stato scelto per coprire l'intero ciclo regolatorio recente: dall'entrata in vigore del GDPR fino alla crescente attenzione verso i modelli generativi e alle prime strategie dedicate alle TQ. Per evitare di appiattire le differenze tra ecosistemi comunicativi molto diversi tra loro, le analisi sono state condotte separatamente per piattaforma. Questo ha permesso di leggere con precisione le specificità degli ambienti – dalla lunghezza dei testi al linguaggio visivo, dalla presenza di attori istituzionali alla logica delle interazioni – senza forzare generalizzazioni improprie. Il corpus è stato sottoposto a un accurato lavoro di pre-processing (data cleaning, lemmatizzazione, tokenizzazione) e, successivamente, alla ricostruzione degli attori attraverso un workflow di matching e linkage che ha combinato dizionari, regole lessicali, metadati e validazioni iterative. Questo passaggio è stato cruciale per collegare i discorsi che emergono dalle piattaforme agli attori già individuati nelle prime due fasi, consentendo di osservare convergenze discorsive, fratture e fenomeni di polarizzazione o di ibridazione e rendendo possibile una comparazione multilivello. L'analisi computazionale ha previsto una doppia strategia: da un lato, una Semantic Network Analysis longitudinale per osservare l'evoluzione delle narrazioni, individuare nodi semantici centrali e analizzare l'emergere e la trasformazione dei cluster tematici; dall'altro, lo sviluppo di un modello di machine learning basato su BERT per classificare la polarità e le categorie narrative. La costruzione del training set (composto da 6.000 post selezionati con campionamento stratificato e codificati manualmente) ha garantito coerenza con gli insight emersi dalle fasi precedenti, rendendo il modello coerente con la struttura concettuale complessiva del progetto. Le metriche di performance del modello sono state utilizzate per valutarne l'affidabilità e guidare le fasi successive di miglioramento. L'uso di modelli transformer ha permesso di cogliere sfumature semantiche difficilmente identificabili con strumenti NLP tradizionali e ha mostrato come le narrazioni cambino in relazione agli attori, alle piattaforme e ai momenti storici. In questa fase, un disegno di ricerca così costruito consente di riconoscere il ruolo degli algoritmi, delle affordance e delle logiche di visibilità nella produzione e circolazione degli immaginari, superando la distinzione tradizionale tra "discorsi online" e "discorsi offline".

L'integrazione delle tre fasi, dunque, ha consentito di ricostruire un quadro articolato e coerente della produzione, della circolazione e della trasformazione delle narrazioni sulle TQ. L'analisi dei documenti istituzio-

nali mostra come la tecnologia venga immaginata e governata; le interviste esplorano le aspettative e le tensioni percepite dagli esperti; le piattaforme digitali rivelano come le narrazioni si diffondono, si ibridano o si polarizzano nel discorso pubblico. Questa triangolazione consente di mettere in relazione le dimensioni macro, micro e meso del fenomeno, riduce i rischi di letture parziali e offre un quadro interpretativo più robusto. L'esperienza di ricerca qui descritta beneficia di un impianto metodologico plurale, iterativo e multimodale, capace di coniugare profondità e ampiezza a diversi livelli discorsivi e immaginativi. Il disegno di ricerca si colloca all'interno di una prospettiva metodologica adattiva e situata, capace di affrontare la complessità delle tecnologie emergenti nei diversi contesti in cui prendono forma. L'ibridazione tra analisi del contenuto, interviste e metodi computazionali consente di coniugare profondità interpretativa e ampiezza empirica, e permette di leggere le TQ non solo come oggetti tecnologici, ma come costruzioni sociotecniche modellate da discorsi, attori, politiche e infrastrutture. Le narrazioni che le accompagnano non anticipano semplicemente un futuro possibile, ma contribuiscono a costruirlo, orientando i processi di innovazione, di governance e di investimento che sostengono l'emergere dei futuri sociotecnici.

9. *Valutare: modelli e pratiche di assessment*

di *Francesco Amato, Dario Chianese*

L'innovazione e le sue implicazioni

L'innovazione tecnologica costituisce un fattore rilevante della trasformazione contemporanea, capace di incidere sulle modalità di produzione, di comunicazione e di interazione sociale. Le tecnologie emergenti possono generare opportunità, ampliando le possibilità di accesso alla conoscenza e trasformando i processi produttivi, con risultati variabili in termini di efficienza e di personalizzazione. Allo stesso tempo, l'innovazione introduce rischi e implicazioni sociali che richiedono attenzione e capacità di analisi critica.

Ogni progresso tecnologico produce effetti che travalicano la dimensione tecnica, incidendo sulle pratiche quotidiane, sui modelli organizzativi e sulle strutture di potere. Le innovazioni modificano i criteri di accesso alle risorse, trasformano le forme di lavoro, alterano la percezione della privacy e ridefiniscono il rapporto tra autonomia individuale e controllo sociale. Questi processi non si manifestano in modo uniforme, ma generano impatti differenziati in base ai contesti, alle competenze digitali, alle condizioni economiche e alle forme di vulnerabilità degli attori coinvolti. In questo senso, l'innovazione non è mai neutrale, ma riflette valori, scelte politiche e interessi che si traducono in configurazioni materiali e simboliche capaci di influenzare la vita collettiva.

Le implicazioni sociali dell'innovazione si manifestano spesso in forme sottili o non immediatamente visibili. Le trasformazioni riconducibili alle tecnologie digitali tendono, infatti, a operare in modo silenzioso, incorporandosi nelle infrastrutture e nei processi decisionali fino a diventare parte integrante dell'ambiente sociale. Ciò rende complesso riconoscerne gli effetti nel lungo periodo, in particolare quelli legati alla concentrazione del

potere informativo, alla riproduzione delle disuguaglianze o alla perdita di agency individuale nei confronti dei sistemi automatizzati.

Per queste ragioni, la riflessione sull'innovazione non può limitarsi a una valutazione dei suoi vantaggi economici o funzionali, ma deve includere un'analisi delle sue ripercussioni etiche, politiche e sociali. La valutazione delle tecnologie, intesa come attività sistematica di osservazione, interpretazione e giudizio, diventa così uno strumento essenziale per comprendere in che modo le innovazioni trasformano la società, chi ne trae beneficio e chi ne subisce gli effetti (Grunwald, 2009; Stilgoe et al., 2013). Comprendere e valutare l'innovazione significa, dunque, andare oltre la logica dell'efficienza e considerare criticamente le implicazioni che le tecnologie producono sui diritti, sulle relazioni sociali e sulle forme di vita collettiva. È in questa prospettiva che si colloca l'attenzione crescente verso i modelli e le pratiche di assessment delle tecnologie emergenti, volti a bilanciare le promesse di progresso con la necessità di una governance consapevole e responsabile del cambiamento tecnologico.

Nel trattare il tema della valutazione, è possibile distinguere due vie complementari. La prima è quella istituzionale, che comprende l'insieme delle pratiche di valutazione promosse da attori pubblici e privati operanti a livello nazionale e sovranazionale. Essa si concentra sulla definizione di quadri normativi, linee guida e strumenti di analisi finalizzati a comprendere e gestire le implicazioni etiche, sociali e ambientali delle TE. È il livello in cui la valutazione assume un valore di governance, orientando le politiche dell'innovazione e la responsabilità collettiva nei confronti del progresso tecnico.

La seconda via è quella organizzativa, che riguarda le modalità con cui imprese, enti e istituzioni incorporano la valutazione nei propri processi decisionali e operativi, facendone uno strumento di gestione interna utile a identificare rischi, opportunità e criteri di responsabilità lungo l'intero ciclo di vita delle tecnologie. Essa consente di tradurre principi generali in pratiche concrete, coinvolgendo, quando possibile, gli stakeholder interessati e favorendo scelte più consapevoli, sostenibili e coerenti con i valori dell'organizzazione e della società civile.

La discussione di queste due prospettive, quella istituzionale e quella organizzativa, permette di cogliere la pluralità dei significati e delle pratiche che accompagnano la valutazione delle tecnologie emergenti e costituisce il punto di partenza per le sezioni successive, dedicate ai modelli e alle pratiche di valutazione delle tecnologie emergenti.

La via istituzionale, tra valutazione delle implicazioni sociali e pratiche di assessment

La valutazione delle tecnologie emergenti rappresenta oggi una delle questioni più rilevanti nell'ambito della governance dell'innovazione. I rapidi sviluppi della ricerca scientifica e dell'industria tecnologica impongono agli attori pubblici e privati di dotarsi di strumenti capaci di identificare, analizzare e gestire le implicazioni sociali, economiche e ambientali delle tecnologie emergenti. In questo contesto, negli ultimi anni, la ricerca sociale ha avuto un ruolo decisivo nell'evidenziare le implicazioni sociali dell'innovazione tecnologica e nel fornire le basi teoriche e metodologiche per la costruzione di strumenti di valutazione. Le scienze sociali consentono infatti di cogliere le implicazioni manifeste e latenti dei processi innovativi, spesso trascurati da un approccio puramente tecnico, e di restituire una visione complessiva che integri dimensioni materiali, normative e simboliche.

I framework di valutazione sviluppati da istituzioni pubbliche, organismi sovranazionali e consorzi industriali rappresentano l'esito più strutturato di questo sforzo. Essi mirano a tradurre concetti complessi, come equità, trasparenza o sicurezza, in indicatori operativi che consentano una valutazione sistematica delle tecnologie emergenti. Si tratta di strumenti che integrano criteri di analisi tecnica con considerazioni etiche, giuridiche e sociali, cercando di bilanciare le esigenze di innovazione con quelle di tutela dei cittadini e dei diritti fondamentali.

Questi strumenti si caratterizzano per la capacità di rendere misurabile ciò che, nell'innovazione, tende a restare opaco. Indicatori dedicati alla rilevazione dei bias, alla tracciabilità dei processi decisionali automatizzati, o alla misurazione dell'impatto socioeconomico consentono di portare alla luce criticità che altrimenti rimarrebbero invisibili. In questo modo, i framework di valutazione si configurano non solo come strumenti di analisi, ma anche come dispositivi di responsabilizzazione, capaci di favorire trasparenza e fiducia tra cittadini, istituzioni e imprese.

A livello istituzionale, la valutazione assume forme differenti in base ai contesti nazionali e sovranazionali (Jasanoff, 2005). In Europa, ad esempio, il dibattito si è concentrato su principi come affidabilità, centralità dell'essere umano e accountability, che orientano la costruzione di linee guida e standard comuni per la gestione delle tecnologie emergenti. Tuttavia, le differenze tra i sistemi normativi e gli approcci politici dei vari Paesi generano una molteplicità di strumenti e pratiche, riflesso delle diverse priorità degli attori coinvolti. Alcuni quadri di valutazione privilegiano gli aspetti etici o la protezione dei dati, altri si focalizzano sulla sicurezza informatica, sull'impatto ambientale o sulla sostenibilità economica.

Questa eterogeneità di approcci non rappresenta un limite, ma piuttosto la testimonianza della complessità del fenomeno. L'innovazione tecnologica è un processo intrinsecamente situato, e la sua valutazione deve adattarsi a contesti d'uso, settori industriali e culture istituzionali differenti. In questo senso, i framework di valutazione non costituiscono strumenti neutri, ma riflettono le priorità, le sensibilità e gli interessi delle istituzioni che li elaborano. Ogni framework, nel definire ciò che è misurabile e ciò che non lo è, contribuisce a stabilire implicitamente quali aspetti dell'innovazione debbano essere considerati rilevanti e quali possano essere trascurati. A tal proposito, un ulteriore elemento di rilievo riguarda i trade-off tra criteri di valutazione, che emergono in modo ricorrente nella letteratura e nelle pratiche istituzionali. Migliorare la trasparenza di un sistema, ad esempio, può entrare in tensione con le esigenze di sicurezza informatica; aumentare la precisione di un algoritmo può ridurre la protezione della privacy; potenziare i controlli etici può rallentare il ritmo dell'innovazione. La valutazione, in questo senso, non è mai un esercizio puramente tecnico, ma una negoziazione di valori che riflette il modo in cui una società bilancia progresso e responsabilità.

Negli ultimi anni, le principali organizzazioni internazionali, tra cui OCSE e Commissione Europea, hanno promosso standard e linee guida condivise per orientare la valutazione delle tecnologie emergenti, con l'obiettivo di costruire un linguaggio comune e ridurre la frammentazione normativa. In ambito europeo, questa traiettoria si è tradotta anche in uno strumento giuridicamente vincolante: l'AI Act (Regolamento (UE) 2024/1689), che istituzionalizza un approccio di governance e valutazione basato sul rischio, imponendo obblighi differenziati per fornitori e utilizzatori, con requisiti specifici per i sistemi ad alto rischio e una logica di conformità verificabile lungo il ciclo di vita del sistema (Unione Europea, 2024).

Accanto alla dimensione regolativa, gli standard e i framework internazionali stanno incorporando criteri che eccedono la sola performance tecnica e chiedono di rendere valutabili anche gli impatti distributivi, le discriminazioni e l'esclusione. In questa direzione vanno, per esempio, i principi dell'OCSE sull'IA, che includono la tutela dei diritti e valori democratici e l'obiettivo di una crescita inclusiva (OECD, 2019), nonché le indicazioni della Commissione Europea sulla *"diversity, non-discrimination and fairness"* come requisito della fiducia nei sistemi di IA (Commissione Europea, 2019) e lo strumento ALTAI come supporto operativo di auto-valutazione (Commissione Europea, 2020). In parallelo, l'UNESCO propone un impianto normativo globale che lega esplicitamente etica dell'IA e principi di equità e inclusione (UNESCO, 2021), mentre il NIST AI RMF for-

malizza, in un quadro volontario, la gestione dei rischi includendo anche la mitigazione di bias dannosi e gli impatti su persone e società (Tabassi, 2023).

Tuttavia, proprio perché molti di questi strumenti restano non vincolanti o dipendono da capacità organizzative diseguali, i framework realmente “socio-centrati” e orientati in modo robusto a equità, diversità e inclusione non risultano ancora uniformemente istituzionalizzati né adottati in modo omogeneo tra settori e paesi; i rapporti OCSE sull’implementazione mostrano, infatti, una diffusione discontinua e differenze marcate tra approcci nazionali, priorità e strumenti effettivamente messi in opera (OECD, 2021; OECD, 2023). Anche i framework più avanzati restano oggetto di riflessione critica. Molti di essi si limitano a fornire linee guida generali o procedure di autovalutazione che si basano sulle dichiarazioni dei soggetti valutati, senza prevedere un controllo indipendente né un’ispezione diretta dei sistemi tecnologici. In numerosi casi, la valutazione viene condotta attraverso questionari o moduli compilativi che invitano le organizzazioni a descrivere le proprie pratiche di conformità, senza tuttavia verificare empiricamente gli effetti reali delle tecnologie o la coerenza tra i principi dichiarati e i comportamenti effettivi. Ne deriva che il processo valutativo rischia di assumere un carattere formale, più vicino a una dichiarazione di principio che a un’analisi sostanziale. Questa impostazione riflette in parte le difficoltà operative e politiche che accompagnano la valutazione istituzionale. Da un lato, la complessità tecnica delle tecnologie emergenti rende difficile accedere ai loro meccanismi interni e valutare in modo indipendente la loro effettiva funzionalità; dall’altro, la tutela della proprietà intellettuale e la riservatezza dei dati costituiscono barriere ulteriori alla trasparenza. In questo quadro, le istituzioni tendono a privilegiare modelli di valutazione impostati secondo una logica di classificazione e gestione del rischio, strettamente connessa agli obblighi di conformità, lasciando sullo sfondo dimensioni più refrattarie alla formalizzazione, come gli effetti sociali indiretti, le trasformazioni culturali o i mutamenti nei rapporti di potere che le innovazioni possono produrre; l’AI Act, infatti, struttura l’intervento regolatorio attraverso una tassonomia di livelli di rischio e, per i sistemi ad alto rischio, prescrive requisiti organizzativi e tecnici e procedure di valutazione della conformità, orientando la valutazione verso criteri verificabili e documentabili (Unione Europea, 2024). Il risultato è una tensione strutturale tra la crescente esigenza di accountability e la limitata capacità degli strumenti esistenti di cogliere gli effetti imprevisti, emergenti o latenti dell’innovazione. Tali effetti non si manifestano necessariamente nell’immediato, ma possono divenire evidenti solo nel tempo, incidendo su pratiche sociali, relazioni lavorative, percezioni del rischio o

forme di fiducia collettiva. È proprio su questo piano che la ricerca sociale offre un contributo insostituibile: attraverso metodi qualitativi, osservativi e interpretativi, essa consente di individuare le conseguenze non intenzionali e le asimmetrie di potere che sfuggono alle metriche standardizzate dei framework istituzionali. Queste riflessioni hanno condotto parte della letteratura scientifica più recente a riconsiderare la natura stessa della valutazione, spostandola da un approccio puramente tecnico verso una prospettiva più relazionale e contestuale, in cui i processi di valutazione non misurano soltanto l'efficienza o la conformità, ma anche la capacità delle tecnologie di integrarsi in modo equo e sostenibile nei sistemi sociali. In questa prospettiva, la valutazione diventa un processo di apprendimento collettivo che coinvolge diversi attori, istituzioni, cittadini, esperti, imprese, e riconosce la dimensione politica implicita in ogni decisione tecnologica.

In conclusione, la via istituzionale della valutazione rappresenta oggi un campo dinamico e in continua evoluzione, in cui principi etici, interessi economici e responsabilità politiche si intrecciano nella definizione di nuovi strumenti di regolazione dell'innovazione.

La valutazione come pratica organizzativa

Le trasformazioni nell'uso delle TE, spesso descritte nei termini di balzi improvvisi o cambiamenti radicali, pure dipendono strettamente da processi di integrazione efficaci all'interno delle organizzazioni e degli ambienti socio-istituzionali (Edquist, 2009). Questi sforzi di integrazione creano incertezza e richiedono una valutazione, che diviene una componente centrale in particolare quando le TE sono oggetto di politiche pubbliche, di cambiamenti organizzativi o di questioni di pubblico interesse.

In questo senso, la valutazione è un processo formativo che, al di là delle disposizioni normative, può aiutare le organizzazioni a comprendere il valore, i rischi e le implicazioni dell'adozione di una data tecnologia. Può servire a verificare la conformità a standard o protocolli prefissati, monitorare la gestione dei sistemi impiegati e dei servizi erogati, rilasciare certificazioni, stimare l'efficacia di azioni correttive, rilevare competenze e consapevolezza del pubblico e costruire graduatorie (Aragona, 2022).

L'intelligenza artificiale, oggi, permette di inquadrare chiaramente queste dinamiche: la sua diffusione è vincolata dal modo in cui aziende, istituzioni pubbliche e intermediari integrano infrastrutture dati, sistemi algoritmici e meccanismi di governance nelle pratiche esistenti. Poiché le pratiche di valutazione spesso richiedono interventi dall'esterno delle organizzazioni, sia per via della necessità di attori terzi nella presa di decisioni che

per la possibile mancanza delle competenze necessarie, queste a loro volta si presentano come momenti di collaborazione e partenariato fra diverse organizzazioni, pubbliche e private. Così, entità diverse, mediante l'offerta di supporto infrastrutturale, pratico e di consulenza, fungono da intermediari dell'innovazione, fornendo metodi che aiutano le organizzazioni a comprendere le tecnologie emergenti, offrendo strumenti come valutazioni di fattibilità, studi sugli utenti e linee guida normative (Howells, 2006; Caloffi et al., 2023). Questi strumenti, a loro volta, co-producono i percorsi di innovazione, sia quelli esistenti, che quelli soltanto previsti o auspicabili. Le politiche pubbliche formalizzano questi processi di valutazione, creando requisiti che spingono le organizzazioni ad adottare metodi di valutazione spesso basati sulle scienze sociali, in particolare quelli che affrontano l'equità, la responsabilità e la supervisione umana. Anche i governi supportano la valutazione nel mondo dell'impresa attraverso infrastrutture orientate all'innovazione come banche di prova, centri di competenza e hub di innovazione digitale. Questi intermediari offrono consulenza metodologica, accesso ad ambienti sperimentali e competenze normative.

Diversi approcci alla valutazione traggono linfa dai metodi delle scienze sociali, e possono essere distinti in base al grado di intrusività e al livello di accesso richiesto all'assemblaggio sociotecnico. Per grado di intrusività si intende l'influenza esercitata dal ricercatore sui soggetti o sul fenomeno durante la rilevazione: può essere molto alta, come nell'osservazione partecipante palese, oppure nulla, come nelle raccolte di dati destinate a un'analisi secondaria. I metodi non intrusivi sono non reattivi e non modificano il comportamento dei soggetti; di conseguenza, non consentono neppure di raccogliere le loro opinioni sull'oggetto di studio. Al contrario, molte ricerche centrate sulle tecnologie emergenti permettono ai partecipanti di descrivere come hanno progettato il codice, a partire dagli obiettivi, dalle tecnologie utilizzate, dalle pratiche adottate e dai vincoli incontrati. Le survey si servono di questionari standardizzati per raccogliere dati omogenei e analizzarli statisticamente, così da evidenziare le differenze tra i soggetti. Le tecnologie digitali facilitano in modo significativo questi metodi e, unite alla statistica testuale, rendono più gestibili anche le inchieste con domande aperte, facendo emergere più agevolmente il loro valore informativo (Aragona, 2022).

Vi è anche una grande rilevanza dei metodi qualitativi nell'analisi di casi di studio, ampiamente utilizzata per esaminare il funzionamento delle TE in contesti organizzativi reali. Questi studi documentano spesso discrepanze tra ipotesi tecniche e pratiche locali, soprattutto quando la qualità dei dati, le competenze della forza lavoro o le routine istituzionali influenzano i risultati. Gli approcci etnografici e in generale le tecniche di osser-

vazione rendono visibile il lavoro implicito coinvolto nella preparazione dei dati, nel monitoraggio dei modelli e nell'adeguamento dei flussi di lavoro, e aiutano a identificare dipendenze nascoste o impatti sociali che altrimenti potrebbero essere trascurati (Suchman, 2007).

L'adozione di un approccio etnografico permette di indagare il più vasto assemblaggio sociotecnico di cui una TE è parte, e può riguardare tanto le fasi di sviluppo quanto quelle di impiego. Ciò implica osservare anche le esperienze di coloro che, volontariamente o meno, sono esposti alle pratiche di adozione delle tecnologie. Christin (2020) individua tre prospettive per lo studio etnografico degli algoritmi, che ben si attagliano ai sistemi di IA odierni e richiama le condizioni tipiche della valutazione delle TE odierne, digitali, in rete e in aggiornamento, esemplificandone i diversi contesti applicativi di un approccio etnografico al loro studio: la rifrazione algoritmica, ossia l'insieme degli effetti che un algoritmo produce sulla rete sociale e sulle strutture organizzative in cui viene adottato; il confronto algoritmico, ovvero l'analisi delle somiglianze e delle differenze nelle modalità d'uso di un algoritmo in diversi contesti, come differenti organizzazioni; la triangolazione algoritmica, cioè l'impiego di un algoritmo già in funzione in un determinato ambiente sociale per facilitare la ricerca etnografica, ad esempio seguendo i contenuti che esso raccomanda o individuando soggetti potenziali da intervistare, supportando così la saturazione empirica, il posizionamento dei ricercatori e il loro successivo disimpegno dal contesto studiato.

In questi contesti, vengono usate estensivamente interviste e mappature degli stakeholder, che consentono alle organizzazioni di integrare prospettive diverse nei processi di valutazione ed aiutano aziende e autorità pubbliche a identificare chi è interessato dalle applicazioni delle TE e come emergono nella pratica preoccupazioni relative a fiducia, responsabilità ed equità.

Il caso che segue esemplifica come queste pratiche di ricerca sociale possano fornire ben più della semplice descrizione dei processi osservati, e riguarda un'osservazione partecipante condotta all'interno di una software house italiana impegnata nello sviluppo di un sistema di AI commissionato da un soggetto esterno (Amato, 2025). In un contesto simile il ricercatore non si limita a registrare pratiche e interazioni, ma agisce anche come dispositivo di attivazione riflessiva, contribuendo a far emergere criticità, responsabilità e implicazioni del lavoro tecnologico all'interno dell'organizzazione. La sua presenza consente infatti di portare alla luce tensioni latenti nel contesto studiato, siano le logiche di consegna, i vincoli di costo o le richieste del committente; possono emergere, così, la scarsa trasparenza degli algoritmi forniti, la presenza di bias, la difficoltà di attribuire in mo-

do chiaro le responsabilità decisionali e, in generale, le implicazioni sociali dell'interazione delle TE con gli utenti finali (Amato, 2025).

Alcuni di questi metodi integrano tecniche di interrogazione diretta delle tecnologie, come nel caso degli esperimenti digitali, particolarmente promettenti perché, una volta predisposte basi di dati adeguate, consentono di controllare ipotesi mirate (Aragona, 2022).

Le procedure di valutazione delle tecnologie emergenti più sofisticate, perlopiù alla portata delle grandi organizzazioni e degli enti pubblici, oppure fornite dagli intermediari dell'innovazione, spesso combinano metriche quantitative di performance con evidenze qualitative in un approccio di metodi misti. Le pratiche di validazione quantitative, ad esempio, generano indicatori statistici o metriche che le analisi qualitative aiutano a interpretare, situandole nel contesto di applicazione. Di converso, le valutazioni di impatto normativo integrano sempre più analisi quantitative del rischio con metodi partecipativi che sollecitano il contributo della società civile, di esperti del settore e degli utenti (OCSE, 2019).

Di rilievo è l'utilizzo di pratiche di co-design e design partecipativo: l'intento è quello di coinvolgere gli individui e i gruppi coinvolti direttamente o indirettamente dall'applicazione delle TE fin dall'inizio del loro ciclo di vita, inquadrandone i desideri e i bisogni e incorporandone i suggerimenti e le proposte. Attraverso la partecipazione di questi stakeholder, la co-produzione delle TE si può arricchire del loro contributo, che può essere significativo in virtù della conoscenza palese o tacita sul contesto di destinazione che recano, nonché delle disposizioni etiche e pratiche che li orientano (Steen, 2013); la conoscenza e il processo creativo alla guida della progettazione vengono democratizzate, emancipando gli stakeholder e rendendoli parte attiva della messa in campo delle TE (Bannon, Ehn, 2012). Questi approcci alla progettazione ben si sposano con la concezione degli assemblaggi sociotecnici già discussa, per cui le organizzazioni, nel dinamico e reticolare ciclo di vita delle TE, possono trarre beneficio dall'attivazione di voci esterne ma profondamente legate al contesto dell'innovazione.

La valutazione a livello organizzativo può dotarsi quindi di un arsenale piuttosto vasto di approcci, che possono essere integrati per poter trattare, al contempo, le prestazioni tecniche richieste dalla normativa vigente, il funzionamento desiderato dall'organizzazione, l'aderenza ai principi etici e al contesto sociotecnico di applicazione. Essa rappresenta per l'organizzazione l'occasione di una presa di coscienza critica rispetto alle sue pratiche e alle TE con cui opera, sollecitando la sensibilizzazione e la riflessività e fornendo una leva di innovazione organizzativa attraverso l'emersione di nuove domande, nuove responsabilità e nuove forme di consapevolezza nei processi di progettazione e sviluppo (Amato, 2025).

10. Interrogare: interazione e prompting

di *Dario Chianese*

Introduzione

I modelli linguistici di grandi dimensioni (LLM) hanno assunto un ruolo centrale nel funzionamento delle società digitali contemporanee. Addestrati su quantità di dati senza precedenti e raffinati con tecniche di ottimizzazione e allineamento, essi producono testi coerenti e contestuali in risposta a una grande varietà di stimoli, gestiscono conversazioni estese e svolgono compiti complessi in molti settori disciplinari (Brown et al., 2020; Bubeck et al., 2023). La loro crescente presenza nei servizi digitali e nelle interfacce quotidiane rende l'interazione con questi sistemi un'esperienza diffusa e socialmente rilevante. Dal punto di vista delle scienze sociali i LLM possono essere studiati come artefatti culturali che incorporano forme di conoscenza derivate dai corpora di addestramento e dalle procedure di fine tuning, e come interlocutori capaci di sostenere scambi linguistici organizzati in turni conversazionali. Questa duplice prospettiva richiama da un lato l'attenzione degli studi sociali della tecnologia e delle scienze (Woolgar, 1985; Suchman, 2007; Joyce et al., 2021) sugli assemblaggi sociotecnici in cui i sistemi intelligenti sono prodotti e utilizzati, e dall'altro la tradizione di ricerca che analizza il linguaggio come forma di azione sociale. Il comportamento dei modelli dipende dai dati e dalle procedure che li hanno generati, ma anche dalle condizioni di accesso, dalle interfacce e dalle istruzioni operative che mediano ogni interazione (Devlin et al., 2019; Mitchell et al., 2019).

Gli LLM riorganizzano le informazioni apprese durante l'addestramento e rispondono agli stimoli degli utenti producendo testi che esprimono rappresentazioni, classificazioni e valutazioni provenienti dai contesti culturali in cui sono stati costruiti. Tali rappresentazioni riflettono sia proprie-

tà statistiche dei dati sia scelte progettuali, regole di sicurezza e interventi umani successivi alla fase di addestramento. La loro struttura consente inoltre di generare nuove configurazioni concettuali e di combinare elementi provenienti da domini eterogenei, con effetti che possono essere difficili da prevedere a priori. L'estensione delle capacità conversazionali e la disponibilità di interfacce generaliste, come nel caso di chatbot che integrano diversi modelli e modalità di interazione, favoriscono un impiego quotidiano dei LLM come mediatori nella ricerca di informazioni, come supporto alla scrittura e come strumenti di interpretazione di problemi complessi. In queste pratiche d'uso l'esito delle loro generazioni testuali contribuisce alla formazione di opinioni, giudizi e aspettative, e partecipa quindi alla definizione della conoscenza sociale. Ciò rende necessario indagare con attenzione quali configurazioni di conoscenza i modelli mobilitano e come le rappresentazioni incorporate emergano nel corso di interazioni con utenti differenti.

L'interrogazione sistematica dei LLM può facilitare l'osservazione delle categorie con cui essi organizzano il mondo, dei loro schemi narrativi e dei loro criteri di inferenza. In questo senso lo studio sociologico dei modelli si concentra sulle condizioni tecniche, organizzative e culturali che ne orientano il funzionamento e sugli effetti che le loro produzioni generano nei contesti sociali in cui sono inseriti. L'obiettivo non consiste nell'attribuire ai modelli proprietà antropomorfe, ma nel comprendere come le loro risposte discorsive si formino all'interno di un assemblaggio sociotecnico che coinvolge infrastrutture, dati, sviluppatori e utenti finali (Aragona, Felaco, 2018; Chianese, 2026).

Ispezionare i modelli linguistici: vincoli epistemologici e aspetti di metodo

L'ispezione dei modelli linguistici solleva due questioni fondamentali: quali forme di conoscenza sono incorporate nei modelli e quali metodi risultano adeguati a rilevarle. I LLM vengono costruiti attraverso procedure di apprendimento statistico basate su grandi insiemi di testi e su interventi umani che plasmano il comportamento del modello attraverso il fine tuning e l'allineamento (Mitchell et al., 2019). Le risposte che producono non derivano da un accesso diretto al mondo, ma da regolarità linguistiche e rappresentazionali presenti nei dati di addestramento. Questo vincolo impone un'attenzione metodologica specifica: ogni rilevazione riguarda configurazioni discorsive generate da correlazioni statistiche e dalle istruzioni operative che strutturano l'interazione.

Dal punto di vista epistemologico i LLM possono essere analizzati come parti di un assemblaggio sociotecnico composto da infrastrutture computazionali, procedure di selezione dei dati, criteri di filtraggio e comportamenti degli utenti (Aragona, Felaco 2018; Joyce et al., 2021). Le rappresentazioni prodotte dai modelli emergono da questo insieme di elementi, e la loro osservazione richiede di tenere conto sia delle condizioni materiali di produzione, sia dei vincoli imposti dalle interfacce attraverso cui si accede al sistema. L'interazione non avviene mai in modo neutrale: la formulazione dei prompt, i parametri di campionamento, la presenza o assenza di un contesto conversazionale e le politiche di sicurezza influenzano sulla forma dell'output (Domínguez Olmedo et al., 2024; Zhou, Zhang, 2024). Per questa ragione i metodi delle scienze sociali possono essere adattati allo studio dei modelli, ma richiedono una riformulazione esplicita dei propri assunti. Gli strumenti di misurazione che presuppongono esperienze vissute, intenzioni o preferenze non sono immediatamente trasferibili, mentre risultano più adeguati i metodi che si concentrano su stimoli standardizzati e verificabili, come questionari chiusi, compiti di ordinamento, test basati su vignette e tecniche di analisi del contenuto (Cai et al., 2024; Chianese, 2026; Jiang et al., 2024). L'utilizzo di interrogazioni ripetute e variate permette inoltre di osservare la stabilità delle risposte e di distinguere tra tendenze regolari e variabilità stocastica dovuta al processo di generazione. Lo studio sociologico dei LLM consiste quindi nell'esaminare le configurazioni discorsive che essi elaborano e nel collegare tali configurazioni ai processi tecnici e sociali che ne hanno orientato la formazione. L'obiettivo non riguarda l'attribuzione di stati mentali, ma la ricostruzione delle condizioni che producono determinati esiti linguistici. Questo approccio consente di utilizzare i LLM come indicatori indiretti delle rappresentazioni incorporate nei dati e nelle procedure che li sostengono e di valutare, tramite metodi controllati, in quali circostanze tali rappresentazioni emergano o vengano attenuate.

L'analisi empirica dei LLM dipende strettamente dalle condizioni di accesso al modello. Ogni modalità di interazione, infatti, introduce vincoli specifici che incidono sulla replicabilità delle osservazioni e sulla qualità dei dati raccolti. Le differenze principali riguardano tre aspetti: il tipo di interfaccia, i parametri tecnici disponibili e la trasparenza delle procedure che mediano la generazione delle risposte. L'accesso può avvenire tramite interfacce conversazionali, API o, più raramente, modelli completamente scaricabili. Le interfacce conversazionali, oggi ampiamente diffuse, integrano spesso funzioni di moderazione dei contenuti, pre-prompt invisibili e procedure di *routing* verso modelli diversi a seconda del compito richiesto. Questi elementi introducono livelli di opacità difficili da controllare e spie-

gano perché studi dedicati alla riproduzione dell'esperienza d'uso preferiscano talvolta parlare di approcci black box (Zhou, Zhang, 2024; Urman, Makhortykh, 2025). L'interazione via API offre un controllo maggiore sui parametri, ma anche in questo caso possono intervenire filtri o versioni del modello non documentate. La variabilità introdotta dai parametri di generazione influisce in modo significativo sui risultati. La temperatura, ad esempio, regola il livello di esplorazione del modello e condiziona l'ampiezza dello spazio delle risposte possibili (Qu, Wang, 2024). Il campionamento multiplo è spesso necessario per ottenere una stima stabile della distribuzione delle risposte quando non sono disponibili controlli diretti sui parametri o quando i modelli introducono variabilità stocastica anche a temperatura zero. In diversi studi è stato osservato che repliche numerose del medesimo stimolo sono utili per distinguere variazioni strutturali da rumore casuale (Liu et al., 2025; Domínguez Olmedo et al., 2024).

Un ulteriore elemento di vulnerabilità riguarda l'evoluzione e la trasformazione dei modelli. Le piattaforme aggiornano frequentemente i modelli senza documentare in modo esaustivo le modifiche intervenute, e questo può produrre variazioni nelle prestazioni nel tempo, anche a parità di stimoli (McIntosh et al., 2024). Per questa ragione le ispezioni basate su serie temporali o su confronti tra modelli devono registrare con precisione la versione utilizzata, l'orario della somministrazione e il canale di accesso. Senza questa documentazione, la comparabilità dei dati risulta compromessa. Infine, le politiche di sicurezza influenzano sia la disponibilità delle risposte sia il contenuto generato. I modelli possono rifiutarsi di rispondere a determinati stimoli o generare produzioni testuali fortemente filtrate. Diversi studi hanno mostrato come le condizioni di rifiuto possano variare in base alla lingua, alla provenienza della richiesta o al contesto di utilizzo (Motoki et al., 2025). Per questo motivo la rilevazione deve includere anche la misura dei rifiuti e l'analisi delle condizioni in cui si verificano.

Le ispezioni dei LLM richiederebbero perciò, idealmente, ambienti di somministrazione stabili, controllo dei parametri tecnici e registrazione sistematica delle condizioni di accesso. Solo in questo modo è possibile garantire comparabilità tra modelli, replicabilità delle analisi e interpretazioni coerenti delle variabili osservate. L'applicazione dei metodi delle scienze sociali ai LLM richiede un adattamento delle tecniche tradizionali, poiché le condizioni di osservazione e le proprietà dei modelli differiscono da quelle dei soggetti umani. Le possibilità offerte dalla generazione stocastica, dalla ripetibilità delle interrogazioni e dalla standardizzazione degli stimoli permettono di estendere o riformulare molte pratiche consolidate. In questa sezione vengono delineati gli approcci quantitativi, qualitativi e misti più rilevanti.

I metodi quantitativi si basano su interrogazioni standardizzate, ripetute e comparabili. Risultano particolarmente adatti quando l'obiettivo è misurare tendenze ricorrenti nelle risposte, valutare la coerenza del modello o confrontare comportamenti tra diversi sistemi. La possibilità di replicare la somministrazione più volte consente di stimare la distribuzione delle risposte e di isolare la componente stocastica dal comportamento stabile del modello (Liu et al., 2025). Gli approcci quantitativi richiedono attenzione alla costruzione della base empirica. Le variazioni minime nel testo delle domande, nell'ordine delle opzioni o nella formulazione delle istruzioni possono produrre cambiamenti significativi nei risultati (Ceron et al., 2024; Domínguez Olmedo et al., 2024). Per questo motivo i protocolli quantitativi includono spesso controlli di invarianza, parafrasi controllate e test di perturbazione. Quando le risorse lo permettono, l'impiego di campionamenti numerosi amplia la validità delle inferenze.

I metodi qualitativi si concentrano sulle produzioni discorsive generate dai modelli. Le interrogazioni aperte, le vignette, le sequenze argomentative e le richieste di spiegazioni permettono di analizzare narrazioni, giustificazioni e schemi interpretativi attivati dal modello in risposta agli stimoli (Fan et al., 2019). Questi materiali possono essere trattati mediante analisi tematiche, codifiche manuali o strumenti di analisi del contenuto. Le tecniche qualitative risultano altrettanto utili ad analizzare queste produzioni testuali, nonché ad osservare le configurazioni concettuali elaborate dai modelli e per identificare orientamenti impliciti, schemi narrativi ricorrenti o formulazioni che riflettono regolarità apprese durante l'addestramento. Tuttavia, richiedono particolare cautela interpretativa. Le spiegazioni prodotte dai modelli non corrispondono a processi mentali interni, ma derivano da strutture linguistiche apprese e possono costituire razionalizzazioni post hoc (Turpin et al., 2023). I metodi qualitativi devono tenere conto anche delle modalità con cui le politiche di sicurezza influenzano il comportamento del modello. I rifiuti di rispondere, le formulazioni neutre o le strategie di mitigazione linguistica rappresentano elementi osservabili che possono orientare l'analisi (Motoki et al., 2025).

Gli approcci misti combinano i vantaggi dei metodi quantitativi e qualitativi. Una strategia comune consiste nel somministrare domande chiuse e, successivamente, chiedere al modello di spiegare la risposta fornita. Questa doppia rilevazione consente di osservare sia la selezione dell'opzione sia i criteri discorsivi attivati durante la spiegazione. Un'altra forma di approccio misto prevede l'integrazione tra analisi del contenuto e codifica algoritmica. I testi generati possono essere trattati come grandi corpora, analizzati con tecniche di statistica testuale o classificati attraverso modelli linguistici addestrati separatamente (Dunlap et al., 2025). Tuttavia, quando

i LLM vengono impiegati come strumenti di annotazione o valutazione, si corre il rischio di introdurre circolarità, poiché modelli simili possono condividere bias o pattern linguistici comuni.

Tecniche di interrogazione per i modelli linguistici

La ricerca sui sistemi algoritmici offre un insieme di tecniche che possono essere adattate allo studio dei LLM, pur con le cautele necessarie dovute alla loro natura linguistica e alla variabilità introdotta dai parametri di generazione (Hartmann et al., 2023; Cai et al., 2024). Per orientare in modo più sistematico l'osservazione dei comportamenti dei modelli, risulta utile partire dalla classificazione dei principali tipi di input impiegati nelle pratiche di ispezione. Ogni classe di interrogazione corrisponde a specifiche forme di rilevazione e comporta determinati vantaggi e limitazioni.

Interrogazioni a risposta chiusa

Le interrogazioni a risposta chiusa includono domande a scelta multipla, item di scala, compiti di ordinamento o classificazione, nonché richieste di output strutturati in tabelle, elenchi o formati nidificati, eventualmente anche in formati di file specifici come il JSON (Guo, 2023; Li et al., 2024). Questi strumenti sono utilizzati quando occorre un controllo elevato sulla forma dell'output o quando la comparabilità con dati ottenuti da esseri umani è cruciale. L'impiego di inventari psicometrici, scale Likert o questionari politici consente di valutare stabilità, coerenza e orientamenti espressi dal modello attraverso risposte uniformi e facilmente aggregabili (Jiang et al., 2023; Motoki et al., 2025). Altre applicazioni sono invece peculiarmente adatte ai LLM, come il mascheramento linguistico, che consiste nel chiedere al modello di prevedere una parola o una sequenza all'interno di un testo parzialmente oscurato. Storicamente impiegato per valutare modelli di tipo BERT (Devlin et al., 2019), rimane utile come forma di interrogazione chiusa che permette di osservare associazioni statistiche, sensibilità semantica e capacità predittive. Nell'ambito dei LLM generativi, questo metodo funziona come forma di test mirato che riduce la variabilità dell'output. Dal punto di vista metodologico, le interrogazioni a risposta chiusa favoriscono la replicabilità, ma richiedono attenzione alle variazioni introdotte dai parametri di generazione e dall'ordine delle alternative. Studi recenti mostrano che inversioni di etichette, parafrasi o modifiche minime possono influire sull'esito delle risposte e devono essere incluse nei protocolli di robustezza (Ceron et al., 2024; Domínguez Olmedo et al., 2024).

Interrogazioni a risposta aperta

Le domande aperte, che possono produrre come risposte testi liberi brevi o estesi, sono utilizzate per esplorare forme di ragionamento, giustificazioni, narrazioni e spiegazioni generate dal modello (Fan et al., 2019). Sono adatte per analizzare la capacità del modello di sostenere sequenze argomentative, rispondere a vignette, simulare dialoghi multilivello o commentare output precedenti. Le richieste di traduzione sono un caso particolare di interrogazioni a risposta aperta e riguardano la mappatura tra lingue diverse e possono includere testi brevi, documenti lunghi o traduzioni da modalità audio e visive. Le valutazioni cross linguistiche hanno mostrato che i modelli presentano differenze marcate in base alla lingua di input, rivelando bias culturali e divergenze significative nelle risposte (Zhou, Zhang, 2024; Durmus et al., 2024). Questo rende la traduzione una tecnica utile per individuare variazioni culturali incorporate e per testare la coerenza semantica interlinguistica. L'approccio richiede protocolli di analisi del contenuto o tecniche di codifica manuali e automatizzate. Le domande aperte forniscono più informazioni sulle strutture discorsive dei modelli, ma amplificano la variabilità e necessitano di campionamenti più numerosi per individuare tendenze stabili nelle risposte.

Generazione libera di testi

La generazione libera comprende richieste di sintesi, trasformazione di dati in testo, produzione di scenari o scritture condizionate (Dunlap et al., 2025; Breazu, Katsos, 2024). Questo tipo di interrogazione permette di valutare la capacità del modello di espandere contenuti, creare situazioni ipotetiche, variare stile e tono, oppure elaborare materiali provenienti da altre modalità. È particolarmente utile per analizzare schemi narrativi ricorrenti, preferenze implicite e ragionamenti tipici del modello. Ciascuno di questi tipi di input può essere integrato con tecniche di audit e con test di perturbazione controllata. L'audit permette di valutare sistematicamente il comportamento del modello attraverso batterie di interrogazioni comparabili, mentre le perturbazioni servono a misurare la stabilità delle risposte di fronte a variazioni minime nelle formulazioni (Scherrer et al., 2023; Ceron et al., 2024).

Elementi critici nell'ispezione dei modelli linguistici

L'interrogazione dei LLM richiede di considerare altre proprietà specifiche che influenzano in modo diretto le tecniche di rilevazione. Questi modelli producono testi sulla base di correlazioni statistiche apprese du-

rante l'addestramento e non dispongono di stati interni osservabili o rappresentazioni accessibili in modo diretto. L'opzione più promettente per analizzarne il comportamento, e ancor più per quanto riguarda modelli proprietari come quelli utilizzati dalla maggior parte degli utenti, consiste quindi nell'esaminare le generazioni testuali prodotte in risposta a stimoli controllati, registrando, per quanto possibile, le condizioni tecniche che hanno determinato tali risposte (Mitchell et al., 2019).

Un primo aspetto riguarda la struttura *stateless* delle interazioni. Ogni istanza del modello elabora l'input sulla base del testo fornito e del contesto conversazionale precedente, senza mantenere una memoria stabile da un'interazione all'altra. Questa caratteristica consente di somministrare interrogazioni ripetute, indipendenti o variate, e di raccogliere campioni di risposte comparabili senza il rischio di interferenze dovute all'esperienza pregressa all'interno della stessa sessione. Un secondo elemento riguarda la generazione stocastica. Anche a parità di stimolo e di parametri, il modello può produrre output differenti, e questa variabilità deve essere trattata come una proprietà metodologica strutturale. La ripetizione delle somministrazioni è quindi necessaria per ricostruire la distribuzione delle risposte e distinguere le tendenze ricorrenti dal rumore generato dal processo di campionamento (Liu et al., 2025; Qu, Wang, 2024).

I LLM sono inoltre sensibili alla formulazione degli input. Piccole variazioni nelle istruzioni, nell'ordine delle opzioni o nella lingua di somministrazione possono produrre cambiamenti significativi nel contenuto delle risposte (Ceron et al., 2024; Zhou, Zhang, 2024). Questa sensibilità rende indispensabile l'uso di stimoli equivalenti e la valutazione dell'invarianza degli output rispetto a parafrasi e inversioni controllate. Un ulteriore elemento consiste nella capacità del modello di generare spiegazioni, motivazioni o narrazioni sulle proprie risposte. Questi testi non derivano da processi riflessivi interni, ma da schemi discorsivi appresi durante l'addestramento. Le spiegazioni possono essere utili per comprendere i criteri che il modello tende a mobilitare in determinate situazioni, ma non garantiscono trasparenza sui meccanismi che generano l'output (Turpin et al., 2023). La loro interpretazione richiede quindi particolare cautela. Infine, i modelli incorporano tracce delle pratiche di filtraggio e delle politiche di sicurezza. Gli episodi di rifiuto o di autocensura fanno parte del comportamento osservabile e influiscono sulla rilevazione. La registrazione sistematica dei rifiuti e delle condizioni che li determinano rappresenta una componente importante del disegno di ricerca, poiché tali fenomeni riflettono sia vincoli tecnici sia scelte normative del sistema (Motoki et al., 2025). Nel complesso, queste proprietà definiscono il quadro entro cui è possibile applicare metodi delle scienze sociali ai LLM. L'analisi deve te-

nere conto della natura statistica e regolata delle risposte, della possibilità di ottenere grandi volumi di output omogenei e della sensibilità dei modelli a variazioni minime degli stimoli. Questo insieme di caratteristiche richiede protocolli accurati e una documentazione dettagliata delle condizioni di somministrazione.

I modelli linguistici come strumenti di ricerca sociale

La diffusione degli LLM ha comportato un ampliamento significativo delle modalità con cui le scienze sociali interrogano le tecnologie. La possibilità di ottenere volumi elevati di output standardizzabili, di riprodurre le somministrazioni in modo controllato e di trattare il modello come in una sorta di ambiente sperimentale ha trasformato le pratiche di ispezione rispetto a quelle tradizionalmente utilizzate nello studio di piattaforme digitali e sistemi algoritmici. In passato molti approcci si concentravano sull'analisi dei contesti d'uso o delle affordance delle interfacce, come avviene nel walkthrough method, che osserva in modo sistematico come una tecnologia orienta l'azione dell'utente attraverso i passaggi e le configurazioni dell'interfaccia (Light, Burgess, Duguay, 2018). L'applicazione di questi approcci ai LLM rimane utile per analizzare le condizioni di accesso, i filtri di moderazione e le istruzioni implicite che guidano il comportamento del modello attraverso l'interfaccia conversazionale. Tuttavia, i LLM permettono anche una forma di interrogazione diretta che prescinde dal contesto d'uso e opera attraverso la generazione controllata di dati testuali.

Una prima area di applicazione innovativa riguarda il cosiddetto silicon sampling. I modelli vengono interrogati per simulare popolazioni, atteggiamenti o distribuzioni di risposte a partire da prompt che descrivono condizioni demografiche o socioculturali. Questa pratica viene utilizzata sia per il pre-testing di questionari sia per esplorare scenari sociali difficilmente osservabili o costosi da rilevare empiricamente (Giorgi et al., 2024; Qu e Wang, 2024). Sebbene questa strategia richieda cautela nella valutazione della rappresentatività, essa fornisce uno strumento rapido per analizzare tendenze, verificare la coerenza di strumenti e generare ipotesi.

Una seconda area riguarda le simulazioni multi-agente. Diversi studi hanno mostrato che più istanze di un modello, eventualmente condizionate da ruoli o identità diverse, possono essere coinvolte in compiti di cooperazione, negoziazione o conflitto, come nel caso di dibattiti politici o giochi strategici (Guo, 2023; Taubenfeld et al., 2024). Queste simulazioni rivelano pattern di comportamento emergente e permettono di esplorare dinamiche

sociali che, pur semplificate, suggeriscono criteri di coerenza interna e schemi decisionali tipici del modello.

Un terzo ambito di sviluppo riguarda gli studi prospettici. I LLM possono essere impiegati per generare scenari futuri, identificare catene causali o mappare conseguenze a partire da cambiamenti ipotetici, come nelle applicazioni che utilizzano metodi ispirati a backcasting o futures wheel (Chianese, 2026). Questi strumenti consentono di esplorare come il modello rappresenti l'evoluzione possibile di fenomeni sociali e tecnologici, evidenziando gli schemi narrativi e i criteri interpretativi che orientano le previsioni generate.

Conclusioni

L'ispezione dei modelli linguistici rappresenta un ambito di ricerca in rapida evoluzione. Le possibilità di interrogazione offerte dai LLM permettono di applicare con relativa facilità molte tecniche delle scienze sociali, ma mettono anche in evidenza limiti e ambiguità che richiedono un approccio prudente. Le risposte prodotte dai modelli dipendono da infrastrutture opache, da dati di cui non è possibile conoscere la composizione completa e da parametri tecnici che influenzano profondamente l'esito della generazione. Questi vincoli suggeriscono che ogni osservazione debba essere trattata come il risultato di un'interazione situata, e non come evidenza stabile di proprietà interne del modello. Gli studi presenti in letteratura indicano che i LLM sono sensibili a variazioni minime degli input, alle lingue utilizzate, alle versioni del modello e ai controlli introdotti dalle interfacce (Zhou, Zhang, 2024; McIntosh et al., 2024). Le analisi che non documentano queste condizioni rischiano di sovrapporre fenomeni diversi o di attribuire al modello caratteristiche che emergono soltanto in un contesto specifico. Per questo motivo la replicabilità rappresenta una sfida centrale. Le piattaforme aggiornano i modelli con frequenza e modificano i filtri di moderazione senza pubblicare documentazione esaustiva, e ciò rende necessario un registro accurato dei parametri e delle configurazioni utilizzate in ogni somministrazione. Queste considerazioni mostrano l'importanza di un approccio sperimentale fondato sui test. La standardizzazione delle interrogazioni, la costruzione di set di input equivalenti, l'uso di repliche numerose e la verifica sistematica della robustezza delle risposte costituiscono strumenti indispensabili per distinguere tra variabilità stocastica e tendenze regolari. I metodi delle scienze sociali possono essere adattati allo studio dei LLM solo se i loro assunti vengono resi espliciti e se le procedure vengono riformulate per tenere conto delle proprietà dei

modelli. Ciò implica un investimento nella progettazione di protocolli comparabili, nella produzione di benchmark personalizzati e nella condivisione di strumenti che facilitino il confronto tra studi diversi.

L'estensione delle possibilità di interrogazione non elimina la necessità di prudenza interpretativa. Le risposte dei modelli non rendono conto di intenzioni, preferenze o credenze nel senso attribuito agli attori umani. Sono invece esiti di processi statistici e di configurazioni discorsive apprese durante l'addestramento. Un'interpretazione eccessivamente antropomorfica può condurre a conclusioni fuorvianti, mentre un'analisi centrata sulle condizioni tecniche e socioculturali di produzione permette di situare con maggior precisione il comportamento osservato.

Il consolidamento delle pratiche di ricerca e lo sviluppo di una metodologia condivisa rappresentano una sfida rilevante per lo sviluppo della ricerca. La documentazione scrupolosa degli esperimenti, la costruzione di archivi pubblici di prompt e risposte, la collaborazione tra gruppi di ricerca e l'integrazione tra approcci qualitativi e quantitativi possono favorire l'emergere di un quadro più robusto e cumulativo. Nel lungo periodo, l'obiettivo non consiste nel fissare una volta per tutte metodi definitivi, ma nello sviluppare strumenti flessibili, capaci di adattarsi a modelli in continua trasformazione. Questa opportunità richiede responsabilità metodologica, attenzione critica e consapevolezza dei limiti delle tecniche adoperate. L'adozione di un approccio sperimentale basato su test di ipotesi, controlli rigorosi e procedure precise, di rado alla portata delle scienze sociali ma in questo caso molto più accessibile alla ricerca, costituisce quindi la via più promettente per costruire metodi adeguati allo studio dei modelli linguistici e per contribuire alla comprensione delle loro implicazioni sociali.

11. Co-creare l'innovazione: attori, reti e infrastrutture

di *Mattia De Angelis*

Innovazione come processo sociotecnico

Nelle scienze sociali contemporanee il concetto di innovazione è progressivamente mutato, passando da una visione lineare e sequenziale, tipica della tradizione economica neoclassica, a un approccio relazionale e processuale, in cui la dimensione co-creativa assume un ruolo centrale (Chesbrough, 2003). Lungi dall'essere il risultato dell'azione di un singolo inventore o dell'implementazione di un'idea tecnica isolata, l'innovazione viene interpretata come un fenomeno sociotecnico complesso, che emerge dall'interazione continua tra molteplici attori, istituzioni, visioni e infrastrutture materiali e simboliche (Geels, 2004; Fuenfschilling, Truffer, 2014). Questo spostamento di prospettiva, ormai consolidato negli studi di sociologia dell'innovazione, nelle policy studies e nelle Science and Technology Studies (STS), consente di riconoscere come la produzione di novità tecnologiche non sia mai un atto puramente tecnico, ma il risultato di una costruzione collettiva che coinvolge una pluralità di contesti e livelli, generando dinamiche di co-produzione intrinsecamente ibride (Jasanoff, 2004; Kanger, Schot, 2019).

Le prime teorizzazioni sistemiche dell'innovazione, a partire dagli anni Ottanta, hanno introdotto il concetto di National Innovation System, mostrando come la capacità innovativa non dipenda esclusivamente dalla ricerca scientifica o dall'intraprendenza delle imprese, ma da un ecosistema più ampio composto da politiche industriali, università, capitale umano, cultura economica, regolazione e infrastrutture (Lundvall, 1992; Nelson, 1993). In questa prospettiva, l'innovazione è un processo ancorato a reti istituzionali e organizzative, il cui funzionamento è modellato da norme, aspettative e investimenti collettivi. L'enfasi si sposta così dal prodotto finale ai processi di apprendimento condiviso, agli scambi tra attori etero-

genei e alle relazioni di lungo periodo che sostengono l'emergere di nuove soluzioni. L'innovazione è, in questa lettura, sempre un fenomeno sistemico e relazionale (Edquist, 2009).

Il modello della Tripla Elica, proposto da Etzkowitz e Leydesdorff (2000), ha rappresentato un ulteriore passo in questa direzione, offrendo una chiave interpretativa per comprendere come università, imprese e istituzioni pubbliche co-producono innovazione attraverso forme di cooperazione, negoziazione e allineamento reciproco. L'università non appare più confinata al ruolo tradizionale di generatore di conoscenza scientifica, ma diviene un attore coinvolto direttamente nei processi di trasferimento tecnologico, nell'imprenditorialità accademica, nella progettazione collaborativa e nella produzione di capitale umano specializzato (Etzkowitz, 2003). Il governo, dal canto suo, abbandona il ruolo di regolatore distaccato per assumere quello di facilitatore, stratega e co-investitore, mentre le imprese riconoscono di dipendere da una rete di attori istituzionali e sociali più ampia rispetto ai soli rapporti di mercato. L'innovazione emerge così come il risultato di interazioni cooperative e competizioni controllate tra attori che condividono responsabilità e risorse (Ranga, Etzkowitz, 2013).

L'estensione alla Quadrupla e alla Quintupla Elica ha arricchito ulteriormente questa prospettiva, integrando l'importanza della società civile, delle comunità locali, delle ONG, degli utenti e, più recentemente, delle dimensioni ambientali ed ecologiche (Carayannis, Campbell, 2009). L'innovazione non è più solo un fatto economico o tecnico, ma coinvolge sensibilità culturali, forme di partecipazione dal basso e immaginari collettivi che orientano ciò che una società considera desiderabile, accettabile o sostenibile (Kuhlmann, Rip, 2018). Allo stesso tempo, la crescente rilevanza delle sfide climatiche e della sostenibilità ha imposto di considerare il rapporto tra innovazione e ambiente non come un vincolo esterno, ma come un elemento strutturale dei processi co-creativi. La Quintupla Elica, in questo senso, rappresenta un tentativo di riconoscere come la natura stessa diventi un attore, o quantomeno un vincolo agente, nella definizione delle traiettorie tecnologiche (Carayannis et al., 2012).

Parallelamente, gli studi sulla governance multilivello hanno mostrato come i processi innovativi contemporanei siano profondamente influenzati da attori sovranazionali e da dinamiche locali. Nel contesto europeo, ad esempio, la Commissione, le agenzie UE, i programmi di finanziamento come Horizon e i dispositivi regolatori nel campo del digitale, della sostenibilità o dell'intelligenza artificiale forniscono quadri istituzionali che orientano l'innovazione attraverso investimenti, standard, linee guida e visioni del futuro (Borrás, Edler, 2014; Kuhlmann, Rip, 2018). Le città, i governi regionali e gli Stati nazionali si muovono così all'interno di uno

spazio istituzionale stratificato, in cui attori dotati di risorse e poteri differenti co-determinano la direzione del cambiamento tecnologico. Ne deriva un campo di interazione complesso, marcato da forme di collaborazione, ma anche da tensioni e asimmetrie.

Le asimmetrie non sono una distorsione della co-creazione, ma uno dei suoi elementi costitutivi. In qualunque processo innovativo, alcuni attori possiedono maggiori risorse materiali, politiche, simboliche e cognitive (Callon, 1984). Le imprese globali possono fissare standard tecnici e influenzare scelte politiche; gli Stati definiscono incentivi, norme e infrastrutture; le istituzioni sovranazionali orientano interi settori industriali; le comunità locali e gli utenti contribuiscono invece con conoscenze contestuali, pratiche situate e forme di appropriazione che possono rafforzare o deviare gli obiettivi iniziali. La co-creazione dell'innovazione, se osservata empiricamente, è dunque un processo in cui simmetria e asimmetria convivono, la collaborazione non implica necessariamente equità, e la distribuzione del potere influenza in modo rilevante ciò che viene considerato innovazione e ciò che invece rimane marginalizzato.

In questo quadro, la componente immaginaria gioca un ruolo determinante. Gli studi sugli immaginari sociotecnici sviluppati nell'ambito delle STS, hanno evidenziato come visioni del futuro, proiezioni desiderabili, narrazioni pubbliche e aspettative collettive agiscano come dispositivi orientativi (Borup et al., 2006). L'innovazione è anche l'esito di un lavoro interpretativo che definisce cosa un sistema considera un problema, quali soluzioni appaiono possibili e quali traiettorie vengono percepite come credibili. L'immaginazione tecnologica, dalle smart cities all'intelligenza artificiale, dai gemelli digitali alle tecnologie quantistiche, modella le preferenze di attori pubblici e privati, orienta gli investimenti, stabilisce priorità politiche e influenza la legittimazione sociale di nuovi strumenti. Gli immaginari operano come meccanismi di coordinamento, talvolta impliciti, attraverso cui la società costruisce anticipazioni condivise (Jasanoff, Kim, 2009).

Questa natura socio-relazionale e immaginaria dell'innovazione consente di comprendere meglio anche la sua diffusione. La teoria classica di Rogers ha descritto la diffusione delle innovazioni attraverso una curva a S, identificando categorie di adottanti (innovatori, early adopters, maggioranza precoce e tardiva, ritardatari) e mettendo in luce il ruolo delle reti sociali nel facilitare il passaggio da nicchie sperimentali a pratiche consolidate (Rogers, 2003). Tuttavia, questa prospettiva, pur utile, tende a rappresentare la diffusione come un processo in cui la società accoglie un'innovazione già formata, limitando il ruolo degli utenti alla semplice adozione.

Gli sviluppi più recenti, soprattutto nell'ambito delle STS e della media anthropology, hanno invece mostrato come l'innovazione non venga mai semplicemente "adottata", ma venga addomesticata. Il concetto di addomesticamento, elaborato da Silverstone e Hirsch sottolinea come i soggetti reinterpretino, trasformino, ri-configurino e talvolta resistano alle innovazioni tecnologiche, integrandole nelle proprie pratiche e nelle proprie identità. L'addomesticamento implica un lavoro attivo di negoziazione del significato, di adattamento materiale, di ridefinizione degli usi e di integrazione simbolica. Le tecnologie non vengono dunque accolte così come progettate, ma vengono tradotte e re-significate in funzione dei contesti d'uso (Silverstone, Hirsch, 1992).

Questa lettura si collega direttamente alla prospettiva co-creativa. Se l'innovazione è costruita nel dialogo tra attori molteplici, allora la sua diffusione non può essere vista come un semplice trasferimento, ma come una fase ulteriore del processo co-produttivo. Gli utenti diventano co-autori dell'innovazione nella misura in cui i loro modi d'uso, le loro invenzioni quotidiane, le loro pratiche situate contribuiscono a stabilizzare o a trasformare il valore e la funzione delle tecnologie. La curva di diffusione descrive una dinamica aggregata, mentre il concetto di addomesticamento restituisce l'eterogeneità delle pratiche reali, mettendo in luce come l'innovazione continui a essere negoziata anche dopo il suo ingresso nella società.

Di conseguenza, la distinzione tra produzione e diffusione dell'innovazione tende a dissolversi. La diffusione diviene un luogo di co-creazione tanto quanto le fasi di progettazione e sperimentazione. Le tensioni tra simmetrie e asimmetrie osservate nella fase di sviluppo si ripresentano nelle fasi di appropriazione e uso, dove differenze di capitale culturale, potere economico, risorse digitali e supporto istituzionale influenzano la capacità di gruppi e individui di integrare le innovazioni nelle proprie pratiche. L'innovazione, in questa lettura, appare come un processo aperto, relazionale e negoziato, attraversato da immaginari, relazioni di potere e forme di appropriazione che ne definiscono di volta in volta la direzione e il significato.

Investigare la co-creazione: metodi, reti, infrastrutture e dati

Se l'innovazione, come discusso nella parte teorico-epistemologica, non può essere intesa come il risultato lineare della progettazione tecnica, ma come un processo co-creativo che emerge dall'interazione dinamica di attori eterogenei, immaginari sociotecnici, regimi istituzionali e in-

infrastrutture materiali e digitali, allora il lavoro metodologico deve essere ripensato in profondità. La scelta dei metodi non può prescindere dalle caratteristiche ontologiche dell'oggetto: ciò che si presenta come un fenomeno reticolare, situato e multilivello richiede strumenti capaci di renderlo empiricamente visibile nelle sue dimensioni processuali, relazionali e infrastrutturali (Suchman, 2007; Latour, 2005). Non si tratta semplicemente di "applicare metodi diversi", ma di assumere un orientamento epistemico che consideri la co-creazione come una condizione costitutiva del conoscere, oltre che dell'innovare (Jasanoff, 2004).

In primo luogo, la natura situata e pratica della co-creazione impone l'adozione di metodologie che seguano gli attori nel loro fare, poiché è nelle pratiche quotidiane che l'innovazione prende forma. Le etnografie socio-tecniche sviluppate nell'alveo delle Science and Technology Studies, dalle osservazioni dei laboratori di Suchman alle etnografie delle infrastrutture di Star e Bowker, mostrano come ogni innovazione sia un assemblaggio di gesti, artefatti, documenti, vincoli tecnici e interpretazioni contingenti (Suchman, 2007; Star, 1999; Star, Bowker, 2010). L'etnografia consente di cogliere il lavoro di coordinamento, di manutenzione e di traduzione attraverso cui gli attori legano insieme frammenti di conoscenza e di materiale, dando vita a configurazioni sociotecniche. In un processo di co-creazione, non esiste innovazione senza un continuo lavoro di negoziazione: osservare da vicino riunioni progettuali, sessioni di prototipazione, incontri tra stakeholder e interazioni con sistemi digitali permette di rendere empiricamente visibile la dimensione emergente del fenomeno (Star, Ruhleder, 1996; Koch, 2017).

Accanto alle etnografie, una seconda linea metodologica riguarda le tecniche orientate a seguire il movimento degli attori e delle loro relazioni. L'approccio dell'Actor-Network Theory, più che una tecnica delimitata, costituisce una sensibilità metodologica che invita il ricercatore a non separare a priori l'elemento sociale da quello tecnico, a non attribuire ruoli fissi agli attori e a trattare materiali, standard, documenti, algoritmi e dati come elementi attivi della rete (Callon, 1984; Latour, 2005; Venturini, 2010). Seguire le traiettorie degli attori, istituzioni, comunità, imprese, artefatti, significa individuare punti di traduzione, momenti in cui gli interessi vengono allineati o trasformati, rotture che ridistribuiscono le posizioni. Il vantaggio di questa prospettiva è la sua capacità di trattare la co-creazione non come un risultato consensuale, ma come un processo segnato da attriti, competizioni interpretative e ricomposizioni continue.

Tuttavia, analizzare le relazioni solamente sul piano qualitativo rischia di non cogliere la struttura più ampia delle reti in cui gli attori sono immersi. Le tecniche derivate dall'analisi delle reti sociali permettono

di rappresentare queste relazioni in modo formalizzato, rendendo visibili configurazioni di collaborazione, centralità strategiche, forme di intermediazione e cluster di competenze (Wasserman, Faust, 1994; Freeman, 1978). In un contesto di innovazione co-creata, la Social Network Analysis non è utilizzata per individuare “responsabili” o “driver” dell’innovazione, ma per comprendere la configurazione reticolare che rende possibile o più difficile la generazione collettiva di conoscenza (Powell, Grodal, 2006). La triangolazione tra osservazione etnografica e analisi delle reti permette di collegare ciò che accade nelle interazioni micro con dinamiche organizzative e istituzionali di livello superiore, evitando sia la riduzione dell’innovazione a infrastruttura tecnica sia la sua riduzione a sola interazione sociale (Emirbayer, Goodwin, 1994; Nicolini, 2012).

Un ulteriore asse metodologico riguarda lo studio delle infrastrutture che sostengono l’innovazione. Come mostrato da Star e Ruhleder, le infrastrutture non sono semplici “sfondi” neutri dell’azione: condizionano ciò che può essere fatto, immaginato, negoziato (Star, Ruhleder, 1996). Studiare l’innovazione co-creata implica rendere empiricamente leggibili componenti che normalmente operano in modo silenzioso, procedure burocratiche, piattaforme digitali, dataset, standard di interoperabilità, algoritmi, sistemi di valutazione, interfacce, dimensioni che Bowker e Star mostrano essere costitutive dei processi sociotecnici (Bowker, Star, 1999). Qui le tecniche di *infrastructure ethnography* diventano decisive: seguire un documento lungo il suo ciclo di vita, analizzare come un dataset viene generato, trasformato e mobilizzato, osservare come un prototipo passa da oggetto di sperimentazione a componente stabilizzata sono tutti modi attraverso cui il ricercatore rende visibili gli strati materiali del processo innovativo (Hull, 2012).

Un tema spesso trascurato, ma metodologicamente cruciale nelle innovazioni contemporanee, è quello dei dati come attori. Nelle prospettive STS, i dati non sono considerati meri input tecnici, ma elementi attivi che partecipano alla configurazione dei processi sociotecnici, si accumulano, circolano, vincolano, orientano decisioni e rendono possibili determinate azioni a scapito di altre (Latour, 1999; Gitelman, 2013). Nei processi di co-creazione basati su piattaforme digitali, intelligenza artificiale, sistemi predittivi o gemelli digitali, i dataset non sono semplici input, ma vere e proprie entità che configurano il campo delle possibilità. Studiare l’innovazione implica dunque osservare come i dati vengono raccolti, puliti, selezionati, classificati, contestati e resi interoperabili; come vincolano o abilitano determinate interpretazioni; come orientano il lavoro degli attori umani; e come partecipano alla costruzione di immaginari tecnologici. Le tecniche che permettono di seguire i dati, come document analysis,

file genealogy, audit dei dataset, analisi dei log digitali, sono ormai parte integrante del toolkit metodologico delle scienze sociali dell'innovazione (Gitelman, 2013; Ruppert et al., 2017; Kitchin, 2014).

Una componente essenziale dei processi co-creativi è la produzione di prototipi, mock-up, artefatti intermedi ed esperimenti. Lungi dal costituire una fase preparatoria, questi elementi assumono una funzione epistemica: rendono possibile mettere alla prova visioni, esplorare scenari alternativi e rendere esplicite aspettative che resterebbero altrimenti implicite. Come mostrato da Rosner (2020), la prototipazione costituisce un processo di conoscenza in cui gli artefatti in continua trasformazione guidano la discussione, orientano le decisioni e generano nuove possibilità d'azione. Osservare come un prototipo viene costruito, discusso, modificato, abbandonato o integrato significa cogliere il processo di innovazione nel suo farsi. Le tecniche che consentono di analizzare prototipi, osservazione partecipante, videoanalisi delle sessioni di lavoro, analisi materiale degli artefatti, rivelano la dimensione esperienziale e riflessiva della co-creazione, mostrando come il sapere si generi attraverso tentativi materiali e non attraverso sole idee (Rosner, 2020).

I metodi partecipativi, ereditati dal design partecipativo e dalle pratiche di co-design, costituiscono un altro passaggio fondamentale. Non sono strumenti di consultazione, ma tecniche di indagine e produzione di conoscenza attraverso cui gli attori esplicitano bisogni, valori, visioni e vincoli. Nei workshop, nelle sessioni di mappatura delle relazioni, nei co-prototipi emergono conflitti, allineamenti, resistenze e possibilità interpretative che difficilmente affiorerebbero in contesti meno strutturati (Simonsen, Robertson, 2013). L'uso di strumenti visuali, mappe di attori, scenari futuri, storyboard e simulazioni consente di studiare l'innovazione come un processo immaginativo e anticipatorio, in cui il futuro viene negoziato collettivamente (Björgvinsson et al., 2012).

Infine, accanto all'osservazione delle pratiche e delle reti, è necessario ricostruire la dimensione temporale dell'innovazione. L'innovazione co-creata non segue sequenze ordinate, ma si sviluppa attraverso retroazioni, improvvise accelerazioni, ritorni a fasi precedenti e ridefinizioni dei problemi, secondo traiettorie sociotecniche non lineari (Pinch, Bijker, 1984; Bijker, 1997; Garud et al., 2013). Ricostruire questi percorsi consente di collegare dinamiche micro, meso e macro, mostrando come immaginari istituzionali, politiche sovranazionali, infrastrutture e pratiche quotidiane si combinino nel generare configurazioni sociotecniche nuove.

Questi diversi approcci non costituiscono un repertorio da cui selezionare strumenti alternativi, ma una costellazione di tecniche che, combinate in modo situato, permettono di rendere empiricamente visibile la comples-

sità dell'innovazione co-creata. È la natura stessa dell'oggetto che richiede pluralità metodologica, l'innovazione non è mai un'entità unica, ma un processo che attraversa livelli diversi di realtà, in cui attori, infrastrutture e immaginari si co-determinano. Studiare tale processo implica dunque una riflessione metodologica che tenga conto dell'ibridità dell'innovazione e della necessità di far emergere ciò che, nelle sue forme quotidiane, tende a rimanere opaco o invisibile.

Trasformare la ricerca sociale in design e innovazione

Riconoscere la natura co-creativa dell'innovazione implica riconsiderare anche il rapporto tra ricerca, progettazione e implementazione tecnologica. Se l'innovazione emerge dall'interazione situata di attori, infrastrutture, immaginari e pratiche, allora le tecniche impiegate per studiarla non possono restare confinate all'analisi *ex post*: devono essere integrate fin dalle prime fasi dei processi di sviluppo. I metodi della ricerca sociale non servono soltanto a osservare l'innovazione, ma a orientarla, dotandola di forme di riflessività coerenti con ciò che la letteratura scientifica mostra come determinante nei processi di co-creazione (Stilgoe et al., 2013; von Schomberg, 2013).

In questa prospettiva, le conoscenze prodotte attraverso etnografie sociotecniche, analisi di rete, studio delle infrastrutture e metodi partecipativi non sono meri strumenti descrittivi. Diventano dispositivi che aiutano progettisti, organizzazioni e istituzioni a riconoscere i molteplici elementi che concorrono alla costruzione dell'innovazione, gli allineamenti precari tra attori, la materialità dei prototipi, i vincoli infrastrutturali, le aspettative implicite, le forme di contestazione, i processi di addomesticamento che si attivano quando un sistema entra nell'uso quotidiano. Trasferire questi saperi nei contesti di sviluppo significa inserire nel design momenti di attenzione alle pratiche reali degli utenti, all'emergere di interpretazioni divergenti, alle dinamiche istituzionali che modellano l'adozione.

Un ambito emblematico, ad esempio, è quello dell'intelligenza artificiale e, in particolare, delle pratiche di esplicazione. Gli approcci tradizionali alla explainable artificial intelligence (XAI), ossia l'insieme di metodi volti a rendere comprensibili e interpretabili i sistemi di intelligenza artificiale, tendono a concepire la spiegazione come un prodotto informativo che chiarisce il funzionamento interno di un modello. Ma se si assume il quadro della co-creazione e dell'addomesticamento, la spiegazione diventa un processo relazionale, deve rendere visibile non solo l'aspetto tecnico, ma l'intera rete sociotecnica che sostiene l'algoritmo, i modi in cui dati, attori

e istituzioni hanno contribuito alla sua configurazione e le assunzioni che ne orientano l'uso (Abdul et al., 2018; Binns et al., 2018). Una spiegazione efficace non “describe” semplicemente il modello: crea le condizioni affinché gli utenti possano negoziare significati, appropriarsi dell'artefatto e reinterpretarne le implicazioni (Miller, 2019). La XAI, in questa visione, non è una tecnica di trasparenza, ma uno strumento di co-adattamento tra sistemi e contesti (Amato et al., 2023).

In modo analogo, checklist (Madaio et al., 2020), standard e protocolli possono diventare strumenti per tradurre nel design ciò che la ricerca mette in luce sui processi co-creativi.

I datasheets per i dataset, le model cards o le etichette algoritmiche possono essere ampliate per includere non solo parametri tecnici, ma anche informazioni sulle reti di attori coinvolte, sugli immaginari che hanno orientato la progettazione, sulle pratiche di interpretazione e manutenzione e sulle controversie emerse (Mitchell et al., 2019; Gebru et al., 2021). Questo tipo di standard non ha lo scopo di fissare una forma definitiva dell'innovazione, ma di mantenere aperta la possibilità di rinegoziare la tecnologia alla luce dei contesti d'uso.

Ne risulta che integrare i saperi della ricerca sociale nei processi di sviluppo tecnologico significa riconoscere che l'innovazione è sempre un processo collettivo. Rendere visibili attori, infrastrutture, dati e pratiche permette non solo di comprendere meglio come nasce un'innovazione, ma di costruirla in modo più trasparente, riflessivo e sensibile ai contesti. In questo senso, la co-creazione non è soltanto un oggetto della ricerca, ma un principio guida per progettare tecnologie che possano essere comprese, negoziate e integrate nei mondi sociali che contribuiscono a generarle.

12. Anticipare l'innovazione: consultare gli esperti, costruire scenari

di Vincenzo Esposito, Cristiano Felaco

Perché il metodo Delphi

La sfida della governance dell'innovazione è semplice da formulare, ma difficile da praticare: accorciare il divario tra dinamiche tecnologiche in atto e capacità delle istituzioni di comprenderle e orientarle. Pur riconoscendo che anticipare l'innovazione in condizioni di scarsità informativa e in contesti in continua trasformazione è un compito intrinsecamente arduo, resta una delle questioni centrali della ricerca quella di elaborare dispositivi concettuali e metodologici capaci di governare il mutamento, trasformando l'incertezza in conoscenza operativa e cercando di orientare l'azione pubblica con sufficiente tempestività, affinché l'innovazione non colga impreparate né le istituzioni né i cittadini (Di Zio et al., 2023).

In questa prospettiva, e soprattutto nei contesti in cui le basi empiriche sono ancora fragili o incomplete, una via percorribile per anticipare il cambiamento consiste nell'interrogare sistematicamente gli esperti, traducendo la loro conoscenza tacita in evidenza utilizzabile attraverso dispositivi rigorosi di elicitazione e validazione. Il ricorso a comunità epistemiche eterogenee per domini, ruoli e provenienze consente infatti di studiare fenomeni complessi, identificare punti di fragilità, prefigurare traiettorie plausibili e costruire ipotesi operative per i futuri possibili dei processi di innovazione, riducendo l'incertezza decisionale anche in assenza di dati consolidati.

In questo quadro, la scelta della tecnica di ascolto e consultazione degli esperti non è neutra. Essa determina quali saperi riescono a emergere, come vengono aggregati e con quale grado di affidabilità possano poi orientare eventuali decisioni. Scegliere un metodo significa quindi decidere non soltanto come raccogliere informazioni, ma anche quale rapporto instaurare tra expertise, incertezza e costruzione di scenari. Ciò implica definire con precisione il disegno di campionamento degli esperti, il formato dell'e-

licitazione, le procedure di mitigazione dei bias, i meccanismi di aggregazione dei risultati e la tracciabilità complessiva del processo.

La scelta metodologica si gioca, in termini generali, lungo due assi interdipendenti: il tipo di disegno da adottare e la tipologia di esperti da coinvolgere. In contesti ad alta incertezza, come quelli dell'intelligenza artificiale, i questionari standardizzati consentono di produrre misure comparabili e rapidamente aggregabili, utili per una ricognizione trasversale delle opinioni, ma tendono a restituire soprattutto una fotografia del presente, risultando meno efficaci quando si tratta di esplorare futuri possibili o di far emergere conoscenze tacite su rischi sistemici, effetti di lungo periodo e dinamiche di lock-in tecnologico (Steenbergen, Marks, 2007). Le interviste agli esperti, dal canto loro, permettono di approfondire il ragionamento individuale e i nessi causali che sorreggono i giudizi, ma lasciano quasi interamente al team di ricerca il lavoro di mediazione e sintesi, con il rischio di introdurre filtri interpretativi legati al background dei ricercatori (Meuser, Nagel, 2009).

Il metodo Delphi nasce precisamente per attenuare tali limiti e si presta in modo particolare allo studio delle nuove tecnologie e dell'intelligenza artificiale. Attraverso round iterativi, feedback controllato e anonimato tra pari nelle fasi valutative, esso favorisce l'interazione cognitiva tra esperti senza riprodurre le distorsioni tipiche del confronto faccia a faccia. Al tempo stesso, consente di raccogliere contenuti qualitativi profondi, anche mediante domande aperte, orientati al futuro e alla prefigurazione di traiettorie plausibili (Sablatzky, 2022). In assenza di serie storiche affidabili o di dati osservativi sul lungo periodo – condizione tipica dei fenomeni emergenti – il Delphi permette di simulare scenari regolativi e sociotecnici, esplorare ipotesi controfattuali e valutare ex ante impatti potenziali e trade-off.

Altrettanto decisiva è la scelta degli esperti. In una prospettiva graduata dell'expertise, è opportuno distinguere tra competenze distribuite, come quelle degli utenti intensivi di piattaforme digitali o dei lavoratori esposti all'automazione, e competenze professionali stabilizzate, come quelle di sviluppatori, responsabili IT, policy maker, autorità di regolazione e rappresentanti delle parti sociali. I questionari risultano appropriati per campionamenti ampi e differenziati, mentre le interviste possono coinvolgere anche soggetti portatori di expertise situata. Il Delphi, per contro, trae il massimo beneficio dall'ingaggio di esperti professionali con esperienza continuativa nel dominio di riferimento, poiché la procedura iterativa e la richiesta di formulare giudizi condizionati su sviluppi futuri presuppongono familiarità con meccanismi, interdipendenze e incertezze strutturali del campo (Collins, Evans, 2019; Sablatzky, 2022).

L'impostazione di un Delphi efficace muove dunque dalla definizione del perimetro analitico: occorre precisare con rigore il dominio oggetto d'indagine, l'orizzonte temporale di interesse e le domande guida della consultazione. Questa chiarificazione iniziale non è un adempimento formale, ma la condizione che assicura comparabilità tra i contributi degli esperti e coerenza nella successiva fase di sintesi (Culot et al., 2020). È parimenti cruciale esplicitare il livello di astrazione richiesto – descrizioni di tendenze, ipotesi condizionali, valutazioni di impatto e probabilità – e il lessico operativo condiviso, così da ridurre ambiguità interpretative già in fase di arruolamento.

In termini applicativi, il Delphi risulta particolarmente adatto a individuare segnali deboli e aree di rischio legate all'adozione di nuove tecnologie di IA, a stimare plausibilità e impatto di sviluppi tecnologici in assenza di basi dati storiche solide, a ordinare per priorità interventi di policy e requisiti regolativi, nonché a generare assunzioni di scenario coerenti da impiegare nella successiva costruzione di scenari regolativi, organizzativi o sociotecnici. In questo senso, il Delphi non pretende di predire il futuro delle tecnologie emergenti, ma rende esplicite e verificabili le ragioni degli esperti sui futuri possibili dell'intelligenza artificiale, riducendo il bias di gruppo e fornendo output immediatamente utilizzabili per la governance.

Il disegno empirico

A fini esemplificativi, la riflessione metodologica può essere illustrata attraverso un caso di studio dedicato alla governance dei sistemi di intelligenza artificiale. In questo caso, il Delphi non è stato impiegato per esporre risultati sostantivi in senso stretto, ma per mostrare come un disegno di consultazione strutturata possa essere utilizzato in un contesto caratterizzato da elevata incertezza e rapida innovazione tecnologica.

Il disegno empirico è stato costruito in tre round, sviluppati nell'arco di tre mesi, secondo un approccio progressivo volto a elaborare un consenso informato su alcuni snodi strategici dello sviluppo e della regolazione dell'IA nel quinquennio successivo. Il primo round ha avuto natura esplorativa e qualitativa. È stato somministrato un questionario con otto domande aperte, articolate attorno a due nuclei tematici principali. Il primo ha riguardato dimensioni critiche dell'ecosistema dell'IA, come l'opacità delle tecnologie e dei processi decisionali automatizzati, la distribuzione delle responsabilità tra produttori e utilizzatori, e gli impatti sul lavoro in termini di trasformazione dei modelli organizzativi, riconfigurazione delle competenze e dinamiche occupazionali. Il secondo ha approfondito possi-

bili evoluzioni del quadro regolativo europeo, con particolare attenzione ai dispositivi utili a sostenere innovazione responsabile, gestione dei rischi, accountability e flessibilità regolatoria.

Al termine della raccolta, ciascun esperto ha ricevuto una sintesi strutturata delle risposte aggregate, così da favorire una riflessione informata e il confronto tra prospettive disciplinari e professionali differenti, in coerenza con la natura iterativa del metodo. Nei round successivi si è proceduto a una progressiva strutturazione del materiale emerso. Gli item derivati dal primo round sono stati trasformati in affermazioni valutabili con scale numeriche o, quando opportuno, in stime probabilistiche. Le strategie ricavate dalle risposte aperte sono state organizzate e sintetizzate per essere presentate nel secondo round, in conformità a standard metodologici consolidati, così da garantire coerenza interna, completezza e rappresentatività del ventaglio di posizioni.

Nel secondo round, tali strategie sono state sottoposte a una doppia elicitazione: una sezione quantitativa, in cui i partecipanti hanno espresso un giudizio su scala da 1 a 10, e una sezione qualitativa, destinata a raccogliere considerazioni sulle condizioni di applicazione e sulle implicazioni etico-sociali delle opzioni prospettate. È stata adottata una soglia di consenso pari ad almeno 7,5/10, in linea con la prassi Delphi, utile a distinguere con chiarezza le aree di accordo robusto da quelle che richiedevano ulteriori approfondimenti. L'integrazione di misure numeriche e motivazioni testuali ha consentito di selezionare un primo nucleo di linee guida operative, ancorando la prioritarizzazione a ragioni esplicite e a condizioni d'implementazione realistiche nei diversi contesti sociotecnici, giuridici ed economici.

Il terzo round ha svolto una funzione di affinamento e consolidamento. Le strategie e le linee guida, riformulate alla luce dei riscontri quantitativi e delle riflessioni qualitative emerse nel round precedente, sono state nuovamente sottoposte agli esperti per una valutazione finale orientata a migliorarne precisione concettuale, chiarezza espositiva e applicabilità operativa. L'intero percorso ha preservato l'anonimato tra pari e ha impiegato un feedback controllato di natura descrittiva, basato su statistiche sintetiche e su estratti anonimi delle giustificazioni più rilevanti, così da alimentare la revisione riflessiva dei giudizi senza introdurre dinamiche persuasive.

La qualità degli esiti dipende, come segnala la letteratura, in misura decisiva dalla composizione del panel. In coerenza con le indicazioni metodologiche più diffuse, il disegno di ricerca ha privilegiato una eterogeneità mirata di competenze, combinando profili con technical expertise nel campo dell'intelligenza artificiale e, ove opportuno, soggetti con interactional expertise, capaci di illuminare pratiche d'uso, vincoli organizzativi e interfacce regolative. La selezione degli esperti ha seguito criteri di per-

tinenza tematica e continuità dell'esperienza maturata nel dominio dell'IA e della sua governance. Nel settore pubblico e in quello privato sono stati inclusi profili che, oltre a lavorare in modo continuativo sul tema, presentavano una produzione scientifica o tecnico-professionale pertinente all'oggetto dell'indagine. In particolare, sono stati privilegiati esperti attivi in laboratori universitari e privati, centri di ricerca, commissioni, organismi istituzionali e reti professionali direttamente coinvolti nello sviluppo, nella valutazione o nella regolazione dei sistemi di IA. La costruzione del panel è partita da una lista iniziale di 523 profili, 318 appartenenti al settore pubblico e 205 al settore privato, ed è avvenuta per fasi successive. Un primo nucleo di contatti consolidati è stato impiegato per il pre-test di chiarezza e pertinenza dello strumento; successivamente il bacino è stato ampliato mediante una ricognizione nel settore pubblico – università, centri di ricerca e istituzioni nazionali coerenti con l'oggetto d'indagine, tra cui CNR, ISTAT, GSSI, Istituto di BioRobotica e Istituto Piaggio – e nel settore privato, attraverso reti professionali, comunità tematiche su piattaforme digitali, newsletter specialistiche ed eventi nazionali. La mappatura ha incluso inoltre figure istituzionali coinvolte nella strategia nazionale, contattando anche esperti nominati dalla Presidenza del Consiglio dei Ministri e dal Ministero dello Sviluppo Economico. Tutti i profili individuati sono stati sottoposti a verifica qualitativa tramite banche dati bibliometriche, al fine di accertare continuità e pertinenza della produzione e riconoscimento comunitario. Il reclutamento è avvenuto mediante comunicazioni personalizzate via e-mail e messaggistica diretta, corredate da una presentazione della ricerca e dal calendario delle attività. La dimensione finale del panel, pari a 45 esperti, di cui 23 del settore pubblico e 22 del privato, è stata mantenuta volutamente contenuta e bilanciata per domini e provenienze, in linea con le buone pratiche Delphi. Nelle fasi di consultazione sono stati garantiti anonimato tra pari, feedback descrittivo non persuasivo e tracciabilità delle decisioni metodologiche.

Interrogare sistematicamente gli esperti costituisce, in questo quadro, un dispositivo epistemico e operativo di primaria utilità per anticipare l'innovazione. Sul piano conoscitivo, l'expertise consente di accedere a saperi taciti, segnali deboli e regolarità emergenti che raramente compaiono nei dati disponibili o nei modelli retrospettivi. Sul piano decisionale, la consultazione strutturata, specie quando organizzata in forme iterative come il Delphi, trasforma intuizioni individuali in giudizi argomentati e aggregati, fornendo priorità comparabili, ipotesi condizionali e indicatori sentinella che riducono l'incertezza e orientano in tempo utile la definizione di politiche, standard e sperimentazioni. Ne deriva un vantaggio di timing: le istituzioni possono intervenire prima che le traiettorie si cristallizzino,

testando opzioni e costruendo capacità adattive invece di limitarsi a correzioni ex post.

Questa utilità non è tuttavia incondizionata. Senza adeguate garanzie metodologiche – eterogeneità mirata del panel, anonimato tra pari, feedback controllato, tracciabilità delle assunzioni – l’elicitazione rischia di amplificare bias di status, conformismo o narrazioni dominanti. Inoltre, il sapere esperto non sostituisce l’evidenza empirica: la sua forza risiede nell’essere complementare a dati sperimentali, progetti pilota, benchmark aperti e monitoraggi continuativi. È precisamente la combinazione di consultazione esperta e cicli brevi di verifica a rendere possibile una forma di apprendimento organizzativo capace di tenere il passo con l’accelerazione tecnologica, legittimando le scelte attraverso processi trasparenti e mantenendo aperto lo spazio delle alternative.

Costruire scenari

È nella natura intrinseca del metodo Delphi trasformare giudizi esperti in scenari plausibili. La consultazione iterativa, l’anonimato tra pari e il feedback controllato consentono infatti di far emergere assunzioni condivise, dissensi argomentati e condizioni abilitanti, organizzandoli in traiettorie alternative del futuro (Renzi, Freitas, 2015). In questa prospettiva, lo scenario non è un esercizio narrativo o descrittivo, ma un dispositivo di coordinamento che permette di tradurre valutazioni esperte in architetture decisionali.

Nel caso della nostra ricerca sulla regolazione dei sistemi di intelligenza artificiale, le opinioni del panel sono state tradotte in tre scenari possibili, che condensano le valutazioni su traiettorie probabili, auspicabili o temute del quadro normativo e di governance dell’IA nei prossimi anni. Lo scenario indicato come più realistico è quello definito *Balanced Tech*: un modello in cui il progresso tecnologico è accompagnato da una cornice regolativa solida, capace di garantire trasparenza, tracciabilità e tutela dei diritti senza soffocare l’innovazione. In questa traiettoria, anche alla luce dell’AI Act, l’Unione europea tende a fungere da *standard setter*; si rafforzano obblighi di documentazione, valutazione dei rischi ed esplicabilità e si affermano meccanismi di certificazione ibrida che integrano profili tecnici, quali robustezza, affidabilità e sicurezza, con dimensioni etico-sociali, come equità, non discriminazione e accountability. Nei domini sensibili, come sanità, giustizia, lavoro e credito, l’IA rimane prevalentemente uno strumento di supporto decisionale, con responsabilità finale in capo all’ope-

ratore umano e gestione trasparente dei bias tramite reporting periodico e sorveglianza continua.

Accanto a questa rotta centrale, il panel ha delineato due alternative meno probabili nel medio periodo ma particolarmente istruttive. Lo scenario Tech-Led enfatizza la spinta autonoma dell'industria: standard de facto guidati da big tech e startup, priorità alla velocità e alla performance anche a costo di minore trasparenza ed esplicabilità, con mitigazione dei bias prevalentemente reattiva e post hoc. Lo scenario Public-First rappresenta invece il controcanto: forte controllo pubblico, investimento in modelli open-source, divieti mirati per pratiche ad alto rischio e partecipazione civica strutturata nei progetti di IA. Si tratta di un'impostazione che massimizza tutela e accountability, ma con possibili rallentamenti, costi maggiori e rischio di burocratizzazione.

La costruzione di scenari consente di rendere esplicite le assunzioni che sorreggono le scelte pubbliche – condizioni abilitanti, rischi, trade-off – e di legarle a indicatori sentinella che possono essere monitorati nel tempo. Al tempo stesso, organizza l'incertezza in traiettorie alternative coerenti, permettendo di preparare ex ante pacchetti di policy differenziati: regole proporzionate, requisiti di trasparenza graduati, criteri di valutazione del rischio e piani di adeguamento tecnologico. In questo senso, gli scenari non prevedono il futuro, ma aiutano a governarlo in modo informato.

L'efficacia degli scenari per la governance risiede anche nella loro capacità di informare la gestione di un portafoglio di interventi. Collegando a ciascuna traiettoria un insieme di opzioni low-regret e misure condizionate, le istituzioni possono allocare risorse in modo dinamico, testare su scala ridotta le soluzioni più promettenti e ampliare o ritirare interventi alla luce dei segnali osservati. Laddove il quadro normativo evolve rapidamente, come nel caso dell'IA, gli scenari possono inoltre supportare il raccordo tra norme orizzontali e specificità settoriali, favorendo coerenza interagenzia e prevedibilità per gli operatori economici. La trasparenza del percorso che dal Delphi conduce agli scenari accresce, inoltre, la legittimazione delle scelte, poiché rende visibili motivazioni, condizioni e assunzioni che portano a ritenere preferibile una linea d'azione rispetto a un'altra.

Esistono tuttavia anche elementi problematici. Gli scenari hanno una dimensione performativa: una volta pubblicati, possono orientare aspettative e investimenti, riducendo la varietà delle opzioni considerate o alimentando un eccesso di fiducia su una traiettoria centrale. Se il panel è poco eterogeneo o se le assunzioni non vengono aggiornate, gli scenari rischiano di riflettere la visione dominante più che il ventaglio delle possibilità. Anche la cosiddetta fatica da scenario è reale: senza chiari punti di aggancio a strumenti regolativi, budget e calendari di revisione, il prodotto rischia di

restare descrittivo e poco operativo. Per massimizzarne il valore, gli scenari vanno trattati come asset vivi, non come verità definitive. In condizioni di incertezza, la loro funzione è di orientamento: offrono mappe per muoversi tra futuri possibili, senza pretendere di esaurirne la complessità.

Delphi e valore della co-creazione nel processo innovativo

Alla luce della prospettiva teorico-metodologica qui adottata, che interpreta l'innovazione come processo sociotecnico relazionale, multilivello e attraversato da immaginari, asimmetrie e pratiche di traduzione, il metodo Delphi può essere letto come qualcosa di più di una semplice tecnica di consultazione. Esso rappresenta piuttosto un dispositivo capace di rendere empiricamente osservabile il carattere co-creativo dell'innovazione, cioè come le traiettorie tecnologiche non emergano in modo lineare da una sequenza di invenzione, sviluppo e diffusione, ma prendono forma attraverso interazioni continue tra attori eterogenei, portatori di saperi, interessi, vincoli e aspettative differenti.

In questa prospettiva, l'innovazione non coincide mai con il solo avanzamento tecnico di un artefatto. Essa è sempre anche il risultato di processi di interpretazione, selezione, adattamento e negoziazione, nei quali ciò che viene riconosciuto come utile, accettabile, rischioso o desiderabile dipende dalla relazione tra sviluppatori, regolatori, utilizzatori, organizzazioni intermedie e cornici istituzionali. La regolazione, di conseguenza, non interviene semplicemente “dopo” l'innovazione, come momento esterno di controllo, ma partecipa essa stessa alla sua configurazione. Definire criteri di rischio, standard di trasparenza, soglie di accettabilità o forme di responsabilità significa contribuire a plasmare il tipo di innovazione che diventa possibile, sostenibile e legittima.

La letteratura sulla co-creazione consente di precisare ulteriormente questo punto. Frow et al. (2015) mostrano come la co-creazione non sia un esito spontaneo della collaborazione, ma un processo che richiede design, orchestrazione e capacità strategica di mettere in relazione attori diversi attorno a problemi comuni di innovazione. In modo analogo, Hodapp, Hawlitschek e Kramer (2019), studiando ecosistemi nascenti di piattaforma nel contesto dell'Internet of Things proprio attraverso il metodo Delphi, evidenziano come la co-creazione del valore emerga in ambienti caratterizzati da forte incertezza, interdipendenza tra attori e assenza di assetti consolidati, ossia in condizioni molto simili a quelle che segnano oggi i processi di sviluppo e regolazione dell'intelligenza artificiale. Anche Chuang (2023), pur in un diverso ambito applicativo, insiste sul fatto che

le piattaforme intelligenti non producono valore in modo unilaterale, ma lo generano attraverso pratiche di interazione, partecipazione e ridefinizione reciproca delle aspettative tra soggetti coinvolti. Trasposto sul terreno della governance dell'IA, ciò significa riconoscere che anche la regolazione produce valore non soltanto perché limita o abilita comportamenti, ma perché costruisce le condizioni entro cui innovazione, fiducia, legittimità e coordinamento possono essere co-generati.

Da questo punto di vista, il Delphi consente di cogliere l'innovazione nel suo farsi, poiché mette in relazione soggetti che occupano posizioni differenti nell'ecosistema sociotecnico e li coinvolge in un processo iterativo di esplicitazione, confronto e rielaborazione dei giudizi. Gli esperti non sono chiamati soltanto a “fornire risposte”, ma a partecipare alla costruzione progressiva di una rappresentazione condivisa, seppur sempre provvisoria e contestata, dei problemi rilevanti, delle priorità emergenti e delle traiettorie future plausibili. Il valore del metodo risiede proprio qui: non tanto nel produrre una fotografia statica delle opinioni, quanto nel rendere visibile il processo attraverso cui certe interpretazioni acquistano stabilità, altre vengono marginalizzate e alcune opzioni si consolidano come più realistiche, desiderabili o urgenti.

In questo senso, la co-creazione non va intesa in termini ingenui o puramente consensuali. Il contributo dei diversi attori non si colloca su un piano simmetrico, né il processo di interazione conduce necessariamente a un'armonia tra prospettive. Al contrario, la consultazione strutturata degli esperti rende spesso visibili le asimmetrie di potere, di risorse, di linguaggio e di capacità di definire i problemi che attraversano i processi innovativi. Vi sono attori che dispongono di maggiore autorità epistemica, altri che controllano infrastrutture tecniche, altri ancora che incidono sulle regole del gioco o sulle condizioni materiali di implementazione. La co-creazione dell'innovazione è dunque anche uno spazio di tensione, in cui si confrontano visioni concorrenti del futuro, modelli diversi di sviluppo tecnologico e differenti modi di distribuire rischi, costi e benefici.

Proprio per questo, un metodo come il Delphi presenta un interesse particolare. Grazie all'anonimato tra pari e al feedback controllato, esso attenua almeno in parte il rischio che le gerarchie reputazionali o istituzionali schiaccino il confronto, offrendo uno spazio in cui il peso degli argomenti può prevalere, più di quanto avvenga in contesti deliberativi tradizionali, sul peso delle posizioni. L'interazione non si svolge come negoziazione immediata tra attori forti e deboli, ma come processo riflessivo in cui ciascuno è messo nelle condizioni di riconsiderare i propri giudizi alla luce di sintesi aggregate, motivazioni altrui e configurazioni alternative del problema. Ne risulta una forma di co-produzione della conoscenza che non

elimina il conflitto, ma lo rende più leggibile, più tracciabile e più utile ai fini della governance.

Questa dimensione è particolarmente rilevante nello studio dell'intelligenza artificiale, dove l'incertezza non riguarda soltanto gli esiti futuri delle tecnologie, ma anche il significato stesso delle categorie con cui le si descrive. Concetti quali rischio, accountability, trasparenza, affidabilità, autonomia, controllo umano o impatto sociale non sono dati una volta per tutte, ma vengono continuamente ridefiniti nei processi di interazione tra comunità tecniche, giuridiche, politiche ed economiche. In tale quadro, il Delphi non si limita a rilevare opinioni preesistenti, ma contribuisce a stabilizzare temporaneamente un lessico comune, rendendo comparabili posizioni che altrimenti resterebbero disperse in linguaggi disciplinari e professionali differenti. Anche questo è un effetto di co-creazione: non solo la produzione di contenuti, ma la costruzione delle categorie stesse attraverso cui quei contenuti diventano intelligibili e utilizzabili.

Un ulteriore aspetto riguarda il rapporto tra co-creazione e anticipazione. Se l'innovazione è un processo aperto, non lineare e distribuito, allora anticiparne gli sviluppi non può significare predirne meccanicamente l'esito, ma costruire condizioni cognitive e istituzionali che rendano più governabili le alternative possibili. In questo senso, la consultazione esperta non serve soltanto a raccogliere previsioni, ma a rendere esplicite le assunzioni su cui si fondano certe traiettorie, i trade-off che esse implicano, i segnali deboli che meritano attenzione e le soglie oltre le quali una determinata configurazione può diventare problematica o desiderabile. L'anticipazione, quindi, non è l'opposto dell'incertezza: è una modalità di lavoro dentro l'incertezza, che cerca di organizzarla in forme utili all'azione pubblica.

Il passaggio dalla consultazione alla costruzione di scenari mostra con particolare chiarezza questa funzione. Gli scenari non costituiscono una mera traduzione narrativa dei risultati, ma il punto in cui la co-creazione della conoscenza si trasforma in co-produzione di orientamenti per la governance. Attraverso gli scenari, le valutazioni degli esperti vengono ricomposte in configurazioni più ampie, nelle quali si intrecciano ipotesi sullo sviluppo tecnologico, opzioni regolative, assetti istituzionali, modelli di mercato e aspettative sociali. In tal modo, il futuro non viene trattato come un oggetto da prevedere, ma come uno spazio di possibilità da organizzare, discutere e rendere politicamente trattabile.

Il tema indagato – l'anticipazione della governance dell'innovazione in un campo ad alta complessità e rapida evoluzione come quello dell'intelligenza artificiale – richiede strumenti capaci non soltanto di descrivere l'esistente, ma di raccogliere e ordinare conoscenze ancora instabili, disperse e spesso implicite. In questo senso, la ricerca non produce semplicemente

un insieme di opinioni esperte, ma costruisce un'infrastruttura cognitiva che rende discutibili e comparabili problemi, priorità e futuri possibili. Il suo contributo non consiste quindi solo nei contenuti emersi, ma nella possibilità di offrire una base riflessiva e metodologicamente trasparente per l'orientamento delle decisioni pubbliche.

Il metodo adottato è parte integrante di questo valore. L'impiego del Delphi, infatti, non ha una funzione meramente tecnica, ma consente di trattare l'expertise come una risorsa collettiva da organizzare, mettere in dialogo e sottoporre a revisione iterativa. Ciò è particolarmente importante in un ambito come quello dell'IA, in cui i saperi rilevanti sono distribuiti tra attori differenti e in cui nessun punto di osservazione individuale è sufficiente, da solo, a cogliere la complessità dell'oggetto. La forza di questo approccio sta allora nel trasformare la pluralità in confronto strutturato e l'incertezza in configurazioni argomentate, senza cancellare divergenze e conflitti, ma anzi rendendoli parte dell'analisi.

Ne deriva una riflessione finale più generale: in contesti di innovazione accelerata, il valore della ricerca non si misura soltanto nella capacità di spiegare ciò che è già avvenuto, ma anche nella sua attitudine a rendere il futuro più intelligibile e, per questa via, più governabile. Consultare sistematicamente gli esperti, costruire scenari e leggere questi processi nella chiave della co-creazione significa assumere che la conoscenza utile alla governance non sia mai già disponibile in forma compiuta, ma debba essere prodotta attraverso dispositivi capaci di connettere saperi, interessi e visioni del futuro. È in questa capacità di organizzare il dubbio, senza semplificarlo artificialmente, che risiede il contributo principale della ricerca qui proposta.

Conclusioni. Ricerca sociale e tecnologie emergenti: verso il pluralismo metodologico

di *Enrica Amaturò, Biagio Aragona*

Nel corso del volume abbiamo mostrato come le tecnologie emergenti (TE) non possano essere comprese come semplici strumenti tecnici, ma vadano lette come fenomeni sociodigitali che coinvolgono persone, organizzazioni e pratiche sociali. In questa prospettiva, emerge un progressivo intreccio di azione umana e non umana e una crescente traduzione della socialità in forme di dati. Allo stesso tempo, anche le pratiche di ricerca si trasformano, sfumando i confini disciplinari e aprendo a nuove modalità di indagine. Per comprendere questi processi, abbiamo adottato una pluralità di prospettive teoriche e metodologiche, raccogliendo contributi che analizzano le tecnologie emergenti e le trasformazioni interconnesse tra vita sociale e pratiche di ricerca.

I contributi raccolti nei diversi capitoli, pur nella loro eterogeneità di approcci, oggetti e strategie metodologiche, convergono nel mostrare come queste tecnologie ridefiniscano simultaneamente le pratiche sociali, le forme della conoscenza e i dispositivi stessi della ricerca.

Il percorso del volume si è mosso lungo traiettorie diverse ma intrecciate, che qui riprendiamo sinteticamente articolandole lungo tre assi principali: epistemologico, tecnologico e metodologico, a cui si intrecciano le questioni della governance e della regolamentazione. Nella prima parte, la riflessione epistemologica sulle TE come fenomeni sociodigitali performativi ci ha portati a riconoscere che la conoscenza è sempre situata e co-prodotta, che l'ingiustizia epistemica attraversa il modo in cui alcuni attori vengono riconosciuti come soggetti credibili mentre altri restano ai margini, e che le tecnologie dell'umiltà impongono di rendere esplicite le scelte ontologiche e politiche implicate nei disegni di ricerca. Ne è derivato il riconoscimento delle TE non tanto come oggetti o come strumenti, quanto come mediatrici epistemiche, entità che modellano simultaneamente ciò che conosciamo e il modo in cui lo conosciamo, e che

richiedono un ethos metodologico fondato sulla riflessività e sulla responsabilità epistemica.

Su queste basi teoriche, la seconda parte del volume ha riguardato l'analisi di quattro traiettorie tecnologiche, ciascuna approfondita in uno specifico capitolo, relative ai modelli linguistici di grandi dimensioni (LLM), ai sistemi di decisione automatizzata (SDA), ai gemelli digitali e alle tecnologie quantistiche (TQ), che, negli ultimi anni, hanno assunto una crescente centralità nel dibattito pubblico e, soprattutto, per la ricerca sociale. Infatti, le TE sono rilevanti nel discorso sociologico non solo per le implicazioni che producono sulla vita degli individui e sulle dinamiche collettive, ma anche per il loro potenziale nel trasformare e ampliare le stesse pratiche di indagine, aprendo nuove possibilità metodologiche e conoscitive. Gli LLM, lungi dall'essere semplici strumenti di calcolo, si sono rivelati esiti di processi di estrazione, selezione e classificazione che ereditano e riproducono le strutture sociali da cui originano, chiamando in causa forme di governance multilivello tra ricercatori, aziende, istituzioni pubbliche e cittadini. I SDA sono stati letti a partire dalla loro non neutralità algoritmica, attraversando bias, sorveglianza e forme di discriminazione, e mostrando come la ricerca sociale sugli SDA agisca da ponte tra analisi teorica e pratiche di analisi applicata, contribuendo al contempo ad accrescere la consapevolezza critica nella vita collettiva. I gemelli digitali urbani hanno offerto un terreno in cui mediazione algoritmica, standard, disuguaglianze socio-spaziali e problemi di privacy si intrecciano, sollecitando spazi di *participatory steering* e di contestabilità dei modelli, e interrogando chi abbia il potere di modificare parametri e contestare scenari. Le TQ, a loro volta, hanno incarnato nel modo più intenso la condizione di emersione continua delle TE rendendo evidente la necessità di una meta-regolazione flessibile.

Nella terza parte del volume, queste questioni sono state affrontate attraverso una pluralità di pratiche di ricerca che articolano in modo diverso il rapporto tra conoscenza, azione e governance delle tecnologie emergenti. Si è cercato di mostrare che è solo attraverso la regolamentazione, la valutazione delle implicazioni e la co-creazione che le tecnologie emergenti potranno davvero apportare benefici alla società. Attraverso la co-creazione, le persone vengono coinvolte tramite l'impegno piuttosto che l'adempimento, generando forti sinergie; questo processo dovrebbe essere alimentato da evidenze sul funzionamento delle tecnologie in diversi contesti. Per questo la sfida principale non è di natura tecnologica, bensì sociale e politica: si tratta di garantire che le tecnologie emergenti siano in linea con i valori umani e i principi etici. Sul piano della governance e della valutazione, abbiamo messo in luce come la complessità e l'opacità dei sistemi digitali

rendano i processi regolatori strutturalmente lenti rispetto all'accelerazione tecnologica, al punto che regolamentare diventa oggi una forma di preparazione epistemica e politica, un atto di autonomia culturale volto a evitare che le traiettorie del futuro siano decise altrove. In questa direzione, lo studio degli immaginari sociotecnici delle TQ, condotto attraverso un disegno multilivello che intreccia documenti istituzionali, interviste semi-strutturate e analisi computazionali sulle piattaforme digitali, ha mostrato come queste tecnologie vadano lette non come meri oggetti tecnici ma come costruzioni sociotecniche modellate da discorsi, attori, politiche e infrastrutture.

La valutazione delle TE si è articolata, infine, lungo una via istituzionale di quadri normativi e linee guida e una via organizzativa con cui imprese ed enti incorporano l'assessment nei propri processi decisionali, restituendo la pluralità dei significati e delle pratiche che oggi accompagnano questa attività. Sul piano metodologico, un passaggio ha riguardato l'interrogazione dei modelli linguistici come pratica di ricerca, dove le risposte dei LLM dipendono da infrastrutture opache, dati non interamente conoscibili e parametri tecnici, e richiedono dunque un approccio sperimentale fondato su documentazione rigorosa, replicabilità e prudenza interpretativa nei confronti di letture antropomorfiche dei comportamenti osservati. Abbiamo poi ricostruito la natura co-creativa dell'innovazione attraverso le teorie della tripla, quadrupla e quintupla elica, in cui asimmetrie, immaginari sociotecnici e pratiche di addomesticamento concorrono a definire le traiettorie tecnologiche, e in cui etnografie sociotecniche, analisi delle reti, studio delle infrastrutture e metodi partecipativi si offrono come dispositivi non solo descrittivi, ma anche orientativi, capaci di entrare nei processi di sviluppo fin dalle loro prime fasi.

A chiudere il percorso, l'anticipazione è stata affrontata attraverso la consultazione strutturata degli esperti e la costruzione di scenari, dove il metodo Delphi permette di attenuare le asimmetrie reputazionali e di trasformare la pluralità in un confronto argomentato, assumendo l'anticipazione non come opposto dell'incertezza ma come una modalità di lavoro al suo interno che cerca di organizzarla in forme utili all'azione pubblica.

In tutti i capitoli si è voluto evidenziare il contributo che la ricerca sociale può dare alla comprensione e valutazione delle TE nella società digitale. Spesso, la ricerca sociale si è concentrata prevalentemente sulla critica alle tecnologie, ma questa prospettiva rischia di incidere solo marginalmente sul modo in cui tali tecnologie vengono sviluppate e implementate. Viceversa, immaginari utopistici vengono portati avanti supportati da interessi e opportunità specifiche. Riteniamo, invece, che a partire dalle critiche bisogna sviluppare un sistema adeguato di regolamentazione e valutazione. In questo processo, un ruolo fondamentale spetta alla ricerca so-

ziale. La ricerca sociale è spesso considerata troppo lenta per stare al passo con l'innovazione tecnologica, ma è solo con una progressiva accumulazione di ricerche, dati e risultati sul funzionamento delle tecnologie emergenti nelle società che sarà possibile riequilibrare le traiettorie di sviluppo delle tecnologie verso forme più inclusive, democratiche ed eque.

Sono almeno tre i motivi per cui la ricerca sociale non può rimanere in disparte. In primo luogo, come abbiamo già evidenziato, i ricercatori sociali hanno gli apparati concettuali necessari per affrontare le sfide delle TE. Accantonando trattazioni epocaliste, utopiche o distopiche che siano, e adottando prospettive teoriche pluraliste, post-antropocentriche e relazionali, le scienze sociali hanno dimostrato di poter fornire riferimenti preziosi per leggere e interpretare le trasformazioni in atto. In secondo luogo, gli apparati metodologici, o meglio i dispositivi, per adottare un linguaggio caro agli studi STS, degli scienziati sociali possono mettere in luce le negoziazioni, i vincoli e i contrasti che co-producono le tecnologie emergenti, aiutando nella ridefinizione e nell'aggiustamento continuo di obiettivi, priorità e traiettorie. La performatività delle narrazioni, le pratiche situate di uso e sviluppo che intrecciano materialità e discorsività delle tecnologie sono il focus su cui bisogna concentrare l'attenzione di tutti gli scienziati, non solo di quelli sociali. Noi abbiamo il compito di mettere a disposizione il nostro bagaglio metodologico, costituito da tecniche che da sempre hanno avuto un'origine interdisciplinare, e che hanno prodotto importanti risultati, all'interno, e al di fuori delle scienze sociali.

Le tecniche presentate in questo volume, che intrecciano analisi documentale, analisi del discorso, metodi sperimentali, osservazione etnografica, computational social science e human computer interaction, possono essere considerate cerniere tra confini epistemologici.

Infine, un'altra ragione che ci spinge a continuare a studiare le tecnologie emergenti è che la tecnologia è da sempre caratteristica imprescindibile di ogni metodo, anche di quello della ricerca sociale. Studiare le TE attraverso le tecniche della ricerca sociale significa interrogarsi anche sul ruolo che vogliamo attribuire alla tecnica all'interno del nostro metodo. Esistono oramai numerose pratiche di ricerca che si avvalgono di tecnologie, e l'analisi delle TE serve anche a comprendere cosa può significare, in prospettiva, aumentarne progressivamente l'uso. In che modo stanno cambiando i processi di costruzione del dato? In che modo si analizzano empiricamente i nostri fenomeni di ricerca? È solo studiando le TE che si può davvero comprendere l'uso delle TE nelle pratiche di ricerca.

La nostra proposta, come Osservatorio F.A.S.T. è un approccio pluralista alla conoscenza delle TE. Per pluralismo intendiamo una variante più ampia dei mixed methods (Creswell, Plano Clark, 2007; Hesse-Biber,

Johnson, 2013) che sono ampiamente praticati tra i ricercatori sociali. La tradizione dei metodi misti ha avuto il merito di legittimare l'integrazione tra approcci qualitativi e quantitativi all'interno di disegni paralleli o sequenziali, ma li ha mantenuti come strategie distinte e separate. Il pluralismo che proponiamo si colloca su un piano differente, che non riguarda solo il modo in cui assemblare metodi pre-esistenti, bensì una riconsiderazione più profonda dell'organizzazione stessa delle pratiche di ricerca, che può essere affrontata solo a partire dalle questioni epistemologiche. La differenza, apparentemente sottile, ha implicazioni di rilievo, perché sposta l'attenzione dalla complementarità delle tecniche alla costituzione stessa del rapporto tra il ricercatore, i dispositivi di ricerca, i fenomeni indagati e il contesto.

Il pluralismo nasce dal riconoscimento che le TE non sono semplici strumenti tecnici, ma configurazioni sociodigitali, in cui codice, dati, infrastrutture, attori e narrazioni si intrecciano e co-evolvono. La riflessione epistemologica condotta nella prima parte del volume ci ha portati a riconoscere che la conoscenza è sempre situata e co-prodotta, e che le TE non sono tanto oggetti o strumenti quanto mediatrici epistemiche, entità che modellano simultaneamente ciò che conosciamo e il modo in cui lo conosciamo. Da questa premessa deriva un'esigenza precisa: i metodi non possono essere selezionati in astratto rispetto al proprio oggetto, ma devono essere modulati insieme alla sua emersione. La pluralità dei dispositivi della ricerca non è quindi una scelta tattica di robustezza analitica, ma la conseguenza di un ethos metodologico fondato sulla riflessività e sulla responsabilità epistemica, che chiede di rendere esplicite le scelte ontologiche, politiche e tecniche implicate nei disegni di ricerca.

In tal senso, il primo tratto distintivo del pluralismo è la capacità di adattare continuamente i disegni di ricerca alla natura sociodigitale degli oggetti studiati, riconoscendo che dati, metodi e teorie co-evolvono insieme alle tecnologie indagate. Secondo questa prospettiva, adattare significa accettare che la scelta degli strumenti non sia un punto di partenza ma un esito provvisorio del confronto con il campo, soggetto a revisione man mano che il fenomeno si rivela. Lo studio degli immaginari sociotecnici delle TQ è, in questo senso, esemplare. Come abbiamo visto nella terza parte del volume, la proposta di un disegno di ricerca multilivello capace di intrecciare documenti istituzionali, interviste semi-strutturate e analisi computazionali sulle piattaforme digitali propone una strategia di ricerca che non si limita a giustapporre fonti e tecniche, ma le tiene in dialogo lungo un percorso in cui ciascun piano dell'analisi orienta la formulazione del successivo, restituendo come le TQ siano modellate dall'azione congiunta e intrecciata di discorsi, attori, politiche e infrastrutture.

Entro questo perimetro va riconosciuto poi il ruolo del ricercatore che, attraverso il proprio posizionamento, opera esso stesso come dispositivo di articolazione tra livelli eterogenei dell'analisi, mettendo in relazione pratiche discorsive, configurazioni tecniche e contesti istituzionali. In questo processo, tuttavia, egli contribuisce anche a delimitare ciò che diventa empiricamente accessibile, selezionando, in modo situato, quali pratiche possano essere rese osservabili e quali restino sullo sfondo. Ne deriva la configurazione di un campo epistemologico entro cui l'analisi può prendere forma, che il pluralismo metodologico non assume come dato, ma riconosce come esito contingente del processo di ricerca, richiedendo una riflessività costante sulle condizioni di produzione della conoscenza.

Da questa impostazione adattiva discende un repertorio ampio di tattiche di ricerca volte a integrare prospettive e punti di vista differenti, che spaziano dal metodo etnografico all'interazione uomo-macchina. Le etnografie sociotecniche, l'analisi delle reti, lo studio delle infrastrutture e i metodi partecipativi, già richiamati più volte nel corso del volume, si offrono come dispositivi non solo descrittivi ma anche orientativi, capaci di entrare nei processi di sviluppo delle TE fin dalle loro prime fasi e di rendere visibili le asimmetrie, gli immaginari sociotecnici e le pratiche di adomesticamento che concorrono a definire le traiettorie tecnologiche. Dove la co-creazione, intesa come coinvolgimento tramite l'impegno piuttosto che l'adempimento, diventa per il pluralismo non soltanto una pratica di valutazione o di governance, ma una postura di ricerca in cui l'intreccio tra strategie metodologiche, attori economici e istituzionali e voci della società civile può configurarsi come parte integrante del processo conoscitivo.

A questo si aggiungono, sul versante dell'anticipazione, la consultazione strutturata degli esperti e la costruzione di scenari, dove il metodo Delphi permette di attenuare le asimmetrie reputazionali e di trasformare la pluralità in confronto argomentato, assumendo l'anticipazione non come opposto dell'incertezza ma come modalità di lavoro al suo interno. Ecco, quindi, come le già citate sandbox partecipano a questa logica come ambienti protetti in cui ricercatori, imprese, governi e cittadini testano e sperimentano questioni critiche prima del rilascio sul mercato entro spazi definiti dove la regolazione si sviluppa con l'innovazione anziché arrivare dopo, e in cui la valutazione si articola lungo la duplice via istituzionale e organizzativa.

Anche l'interrogazione degli LLM è un esempio di come il pluralismo lavori sulla natura dinamica e opaca delle TE, articolando rilevazione, sperimentazione e riflessione critica all'interno di un unico processo di ricerca. Il loro studio richiede oggi approcci metodologici robusti, eventualmente fondati su logiche sperimentali, supportati da una documentazione

rigorosa e orientati alla replicabilità dei processi di ricerca. Inserire questo tipo di interrogazione in un disegno pluralista significa riconoscere che i sistemi indagati possono partecipare al processo di ricerca anche come dispositivi, e non soltanto come oggetti di studio.

L'analisi delle TE in una prospettiva pluralista ridefinisce il ruolo del ricercatore, chiamato non solo a scegliere strumenti adeguati ma a negoziare continuamente tra cornici teoriche, vincoli tecnici e implicazioni etiche. Qui la grande sfida, anticipata in apertura di queste conclusioni, è quella di stabilire un dialogo interdisciplinare efficace tra diverse comunità di ricerca, al di là della sociologia, che combini competenze e punti di vista integrando abilmente metodi e matrici disciplinari. Il pluralismo richiede ricercatrici e ricercatori capaci di sostenere posizionamenti differenti, più o meno intrusivi, più o meno coinvolti, e di tradurre tra linguaggi specialistici diversi senza ridurli a un comune denominatore semplificato. È in questa capacità di intermediazione che si gioca la possibilità per le scienze sociali di entrare nei processi di sviluppo delle TE fin dalle prime fasi, partecipando alla loro costruzione attraverso l'impegno della co-creazione, anziché limitarsi alla critica successiva.

Il punto di cambiamento epistemologico decisivo del pluralismo riguarda, dunque, il rapporto tra metodi e tecnologie. Non sosteniamo l'integrazione di metodi tecnologicamente intensivi nella ricerca sociale perché crediamo che il cambiamento tecnologico abbia un impatto sul metodo, ma piuttosto, crediamo che i metodi si evolvano parallelamente al cambiamento tecnologico, in una relazione co-costitutiva in cui le tecnologie ridefiniscono il campo del conoscibile e i metodi contribuiscono, a loro volta, a stabilizzare o destabilizzare i modi in cui le tecnologie stesse si presentano come oggetti di ricerca. Riconoscere questa co-evoluzione ha conseguenze metodologiche immediate, perché impone di interrogare le condizioni materiali – disponibilità di dati, accesso alle infrastrutture della ricerca, capacità interdisciplinari – che rendono possibile o ostacolano determinati disegni di ricerca, e di considerare queste condizioni come parte integrante della riflessione metodologica anziché come premesse pratiche neutre.

Proprio per questa ragione il pluralismo metodologico non è privo di rischi.

Il primo è quello di scivolare in una giustapposizione non sorvegliata di tecniche, in cui la combinazione diventa un fine in sé, slegato da una domanda di ricerca chiara. Il secondo riguarda l'economia politica delle tecnologie emergenti dove i benefici derivanti dal loro uso per la ricerca sociale rischiano di essere distribuiti in modo diseguale tra le comunità accademiche in base alla loro ubicazione, agli obiettivi di ricerca e alle risorse economiche e organizzative di cui dispongono, e disegni pluralisti

ambiziosi possono accentuare le disuguaglianze crescenti che attraversano l'ecosistema della conoscenza digitale.

Per evitare questi esiti, il pluralismo deve restare ancorato all'imperativo di rendere esplicite, in ogni disegno di ricerca, le scelte ontologiche e politiche che lo orientano.

Vi è infine una dimensione del pluralismo legata al riconoscimento dell'ingiustizia epistemica che attraversa il modo in cui alcuni attori vengono riconosciuti come soggetti credibili mentre altri restano ai margini. La pluralità dei dispositivi della ricerca è anche pluralità dei soggetti riconosciuti come legittimi produttori di conoscenza. Pratiche come la consultazione strutturata, la co-progettazione e l'inclusione degli stakeholder nei disegni di ricerca non sono semplici modi per ampliare il campo, ma strumenti attraverso cui il pluralismo si fa carico delle asimmetrie che attraversano l'ecosistema delle TE. Il legame con le "tecnologie dell'umiltà" richiamate nella prima parte del volume è qui particolarmente stretto in quanto il pluralismo metodologico può essere inteso come la traduzione operativa di un ethos di responsabilità epistemica. Concepito in questi termini, il pluralismo metodologico tiene insieme la dimensione epistemologica del riconoscimento delle TE come mediatrici epistemiche, la dimensione metodologica della costruzione di disegni capaci di integrare in modo flessibile strumenti computazionali, approcci qualitativi e pratiche sperimentali e di fonderne premesse ontologiche ed epistemologiche e la dimensione politica della responsabilità del ricercatore nei confronti delle traiettorie sociodigitali che la sua ricerca contribuisce a stabilizzare o a contestare. È dunque in questa convergenza che il pluralismo metodologico assume il proprio statuto: una modalità di ricerca che non si riduce alla selezione tra approcci alternativi né alla loro composizione conciliatoria, ma si configura come un'onto-epistemologia (Barad, 2007) per una scienza sociale capace di contribuire alla progettazione e implementazione di tecnologie emergenti allineate ai valori umani e ai principi etici.

La strada per comprendere le implicazioni sociali delle TE è ancora lunga, e questo libro dovrebbe essere inteso solo come un punto di partenza. Un aspetto non marginale riguarda l'economia politica delle tecnologie emergenti. Il loro sviluppo nelle economie avanzate è stato guidato da agende politiche sostenute dall'urgenza, dalla competitività internazionale e dalla logica di mercato. Ecco che diventa centrale interrogarsi su come immaginare e costruire forme di governo delle TE a partire dalle evidenze prodotte dalla ricerca sociale. Ciò implica non solo la capacità di analizzare criticamente tali tecnologie, ma anche di tradurre i risultati della ricerca in dispositivi di policy in grado di orientarne concretamente lo sviluppo e l'uso. Affinché ciò sia possibile, la produzione di conoscenza non può

restare confinata nello spazio accademico, ma deve essere messa nelle condizioni di incidere sui processi decisionali, contribuendo alla definizione di regole, standard e pratiche di intervento.

Molte delle forme di governance oggi prevalenti si caratterizzano per un approccio prevalentemente reattivo e settoriale, che interviene a valle dei processi di innovazione, quando le tecnologie sono già diffuse e le loro implicazioni difficilmente modificabili. Questa dinamica è evidente, ad esempio, nei processi di regolamentazione delle piattaforme digitali e dei sistemi di intelligenza artificiale, in cui gli interventi normativi tendono a seguire (più che ad anticipare) le direzioni di sviluppo tecnologico. Ne deriva una difficoltà strutturale nel governare fenomeni già stabilizzati, per i quali le possibilità di intervento risultano spesso limitate; a ciò si aggiunge una frammentazione delle responsabilità tra diversi livelli istituzionali e ambiti regolatori, che rende più complessa l'elaborazione di strategie coordinate in grado di cogliere le interdipendenze tra dimensioni tecniche e sociali. In molti casi, inoltre, la traduzione delle evidenze prodotte dalla ricerca in strumenti operativi di intervento resta parziale o discontinua, contribuendo a un disallineamento tra sviluppo tecnologico e capacità regolativa.

Emerge, quindi, la necessità di ripensare la governance come un processo che non può certo limitarsi a interventi *ex post* (volti a correggere gli effetti indesiderati) ma deve essere concepita come un processo anticipatorio e continuo, in grado di intervenire lungo l'intero ciclo di vita delle tecnologie, dalla progettazione alla diffusione, tenendo conto delle possibili implicazioni.

Ecco perché sviluppare una governance adeguata significa anche dotarsi di strumenti capaci, da un lato, di identificare e monitorare i rischi, dall'altro, di intervenire in modo tempestivo e proporzionato, affinché non si rinunci al potenziale innovativo delle tecnologie, ma lo si adotti in modo critico, rendendo esplicite le condizioni tecniche ed epistemiche che ne orientano lo sviluppo e il funzionamento. In questo senso, pratiche come la valutazione continua, l'anticipazione e la co-creazione, già discusse nel volume, possono essere lette come componenti fondamentali di un approccio alla governance che non separa conoscenza e azione, ma le integra in un processo riflessivo e orientato all'intervento.

Affinché la ricerca sociale possa non solo seguire le traiettorie delle tecnologie emergenti, ma anche contribuire attivamente alla loro anticipazione, e intervenire sul loro reindirizzamento, è necessario sviluppare una capacità di intervento che si collochi nelle fasi iniziali dei processi di innovazione, quando le configurazioni tecniche, organizzative e normative sono ancora in via di definizione. Anticipare non significa prevedere

in modo deterministico gli esiti dello sviluppo tecnologico, ma costruire scenari plausibili, rendere visibili le implicazioni potenziali e aprire spazi di discussione e di scelta. In questo modo, la ricerca sociale può assumere una funzione non soltanto analitica, ma anche orientativa, contribuendo a delineare possibili alternative e a problematizzare le traiettorie dominanti. Ciò richiede di rafforzare l'integrazione tra la produzione di conoscenza, le pratiche di valutazione e i dispositivi di governance, così da rendere la ricerca stessa parte attiva dei processi di configurazione delle tecnologie, anziché semplice osservatrice dei loro effetti. Si sta realizzando una trasformazione profonda del modo in cui la conoscenza viene prodotta e utilizzata, sempre più situata e co-prodotta all'interno di configurazioni sociodigitali date, ma sempre più connessa a pratiche economiche, organizzative e regolamentari di ampia scala. In questo quadro, il nostro compito come ricercatori sociali sarà di costruire pratiche riflessive, responsabili e realmente pluraliste, in grado di connettere la produzione di conoscenza, l'azione pubblica e il governo delle tecnologie.

Bibliografia

- Abdul A., Vermeulen J., Wang D., Lim B.Y., Kankanhalli M. (2018), “Trends and trajectories for explainable, accountable and intelligible systems: An HCI research agenda”, in Aa.Vv., *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, pp. 1-18.
- Agarwal S., Alok S., Ghosh P., Gupta S. (2020), “Financial inclusion and alternate credit scoring for the millennials: role of big data and machine learning in fintech”, *SSRN Working Paper*, 3507827.
- Ahmad A., Desouza K.C., Maynard S.B., Naseer H., Baskerville R.L. (2020), How integration of cyber security management and incident response enables organizational learning, «*Journal of the Association for Information Science and Technology*», 71, 8, pp. 939-953.
- Airoidi M. (2022), *Machine habitus: Toward a sociology of algorithms*, Polity Press.
- Airoidi M. (2023), “Computational authority in platform society: dimensions of power in machine learning”, in C. Borch, J.P. Pardo-Guerra (a cura di), *The Oxford Handbook of the Sociology of Machine Learning*, Oxford University Press, Oxford.
- Airoidi M., Rokka J. (2022), Algorithmic consumer culture, «*Consumption Markets & Culture*», 25, 5, pp. 411-428.
- Akrich M. (1992), “The de-scription of technical objects”, in W.E. Bijker, J. Law (a cura di), *Shaping Technology/Building Society: Studies in Sociotechnical Change*, MIT Press, pp. 205-224.
- Almeida F., Okon E. (2025), Assessing the impact of high-performance computing on digital transformation: benefits, challenges, and size-dependent differences, «*The Journal of Supercomputing*», 81, 6, p. 795.
- Amato F., Aragona B., De Angelis M. (2023), Factors and possible application scenarios of Explainable AI, «*Rivista di Digital Politics*», 3, 3, pp. 543-564.
- Amato F. (2025), Participant observation and evaluative thinking: The case of an Italian software house, «*Sociologia Italiana*», pp. 49-72.

- Amaturo E., Aragona B. (2019), Per un'epistemologia del digitale: note sull'uso di big data e computazione nella ricerca sociale, «Quaderni di sociologia», 81, LXIII, pp. 71-90.
- Amaturo E., Aragona B. (2021), Critical optimism: A methodological posture to shape the future of digital social research, «Italian Sociological Review», 11, 4S, pp. 167-182.
- Amoore L. (2011), Data derivatives: On the emergence of a security risk calculus for our times, «Theory, Culture & Society», 28, 6, pp. 24-43.
- Amoore L. (2020), *Cloud ethics: Algorithms and the attributes of ourselves and others*, Duke University Press, Durham.
- Ananny M., Crawford K. (2018), Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability, «New Media & Society», 20, 3, pp. 973-989.
- Aragona B. (2022), *Algorithm audit: Why, What and How?*, Routledge.
- Aragona B., Acampa S. (2025), Definitions of artificial intelligence and quantum technologies: A comparative thematic analysis of the strategic documents of 29 European countries, «Futures», 103722.
- Aragona B., Amato F. (2022), Rischi algoritmici e strumenti di mitigazione, «RISKELABORATION», 3, 1, pp. 41-46.
- Aragona B., Felaco C. (2018), La costruzione socio-tecnica degli algoritmi. Una ricerca nelle infrastrutture di dati, «The Lab's Quarterly», 20, 4, pp. 97-116.
- Aragona B., Arvidsson A., Felaco C. (2020), Introduction. Ethnography of algorithms. The cultural analysis of a sociotechnical construct, «Etnografia e Ricerca Qualitativa», 3.
- Aragona B., Felaco C. (2020), Understanding algorithms. Spaces, expert communities, and cultural artifacts, «Etnografia e Ricerca Qualitativa», 3.
- Argota Sánchez-Vaquerizo J.A. (2025), Urban Digital Twins and metaverses towards city multiplicities: uniting or dividing urban experiences?, «Ethics and Information Technology», 27, 1, p. 4.
- Bannon L.J., Ehn P. (2012), "Design: Design matters in participatory design", in J. Simonsen, T. Robertson (a cura di), *Routledge international handbook of participatory design*, Routledge, pp. 37-63.
- Baracca A., Ferrari G., Renzetti R. (2017), The "go-stop-go" of Italian civil nuclear programs, beset by lack of strategic planning, exploitation for personal gain and unscrupulous political conspiracies: 1946-1987, arXiv preprint arXiv:1709.05195.
- Barad K. (2007), *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*, Duke University Press, Durham.
- Bareis J., Katzenbach C. (2022), Talking AI into being: The narratives and imaginaries of national AI strategies, «Science, Technology, & Human Values», 47, 2, pp. 268-293.
- Barocas S., Hood S., Ziewitz M. (2013), Governing Algorithms: A Provocation Piece, «SSRN Electronic Journal».
- Batty M. (2018), Digital twins, «Environment and Planning B: Urban Analytics and City Science», 45, pp. 817-820.
- Batty M. (2018), *Inventing future cities*, MIT Press.
- Bazzani L. (2023), *Sociologia dei futuri*, Mimesis, Milano.

- Beach D. (2020), "Process tracing methods", in Aa.Vv., *Handbuch Methoden der Politikwissenschaft*, Springer Fachmedien Wiesbaden, pp. 699-719.
- Beck U. (1992), *Risk society: Towards a new modernity*, Sage (Orig. pubb. 1986, Risikogesellschaft).
- Beckert J. (2016), *Imagined futures: Fictional expectations and capitalist dynamics*, Harvard University Press, Cambridge.
- Beer D. (2018), *The data gaze: Capitalism, power and perception*, SAGE Publications.
- Bender E.M., Gebru T., McMillan-Major A., Shmitchell S. (2021), "On the dangers of stochastic parrots: Can language models be too big?", in Aa.Vv., *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pp. 610-623.
- Benenti G., Casati G., Montangero S. (2025), *Il computer impossibile. Come il calcolatore quantistico cambierà il mondo*, Raffaello Cortina Editore, Milano.
- Bergemann D., Bonatti A. (2024), Data, competition, and digital platforms, «American Economic Review», 114, 8, pp. 2553-2595.
- Bijker W.E., Hughes T.P., Pinch T.J. (a cura di) (1987), *The social construction of technological systems. New directions in the sociology and history of technology*, MIT Press, Cambridge.
- Binns R., Van Kleek M., Veale M., Lyngs U., Zhao J., Shadbolt N. (2018), "'It's reducing a human being to a percentage': Perceptions of justice in algorithmic decisions", in Aa.Vv., *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, pp. 1-14.
- Birhane A., Prabhu V.U. (2021), "Large image datasets: A pyrrhic win for computer vision?", in Aa.Vv., *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, pp. 1536-1546.
- Bjögvinsson E., Ehn P., Hillgren P.-A. (2012), Design things and design thinking: Contemporary participatory design challenges, «Design Issues», 28, 3, pp. 101-116.
- Bonini T., Treré E. (2024), *Algorithms of resistance: The everyday fight against platform power*, MIT Press.
- Borgogno O., Savini Zangrandi M. (2024), Chinese data governance and trade policy: from cyber sovereignty to the quest for digital hegemony?, «Journal of Contemporary China», 33, 148, pp. 578-602.
- Borrás S., Edler J. (2014), "The governance of change in socio-technical and innovation systems: three pillars for a conceptual framework", in Aa.Vv., *The Governance of Socio-Technical Systems*, Edward Elgar Publishing.
- Borup M., Brown N., Konrad K., Van Lente H. (2006), The sociology of expectations in science and technology, «Technology Analysis, Strategic Management», 18, 3-4, pp. 285-298.
- Boschert S., Rosen R. (2016), "Digital Twin-The Simulation Aspect", in Aa.Vv., *Mechatronic Futures*, Springer International Publishing, pp. 59-74.
- Bowker G.C., Star S.L. (1999), *Sorting things out: Classification and its consequences*, MIT Press.

- Breazu P., Katsos N. (2024), ChatGPT-4 as a journalist: Whose perspectives is it reproducing?, «Discourse & Society», 35, 6, pp. 687-707.
- Brown N., Rappert B. (2017), *Contested futures*, Routledge.
- Brown N., Rappert B., Webster A. (2000), "Introducing contested futures: From looking into the future to looking at the future", in N. Brown, B. Rappert, A. Webster (a cura di), *Contested futures*, Routledge, pp. 3-20.
- Brown T.B., Mann B., Ryder N., Subbiah M., Kaplan J., Dhariwal P., Neelakantan A., Shyam P., Sastry G., Askell A., Agarwal S., Herbert-Voss A., Krueger G., Henighan T., Child R., Ramesh A., Ziegler D.M., Wu J., Winter C., Amodei D. (2020), Language Models are Few-Shot Learners, arXiv preprint arXiv:2005.14165.
- Bubeck S., Chandrasekaran V., Eldan R., Gehrke J., Horvitz E., Kamar E., Lee P., Lee Y.T., Li Y., Lundberg S., Nori H., Palangi H., Ribeiro M.T., Zhang Y. (2023), Sparks of Artificial General Intelligence: Early experiments with GPT-4, arXiv preprint arXiv:2303.12712.
- Bucher T. (2019), "The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms", in Aa.Vv., *The social power of algorithms*, Routledge, pp. 30-44.
- Burrell J. (2016), How the machine "thinks": Understanding opacity in machine learning algorithms, «Big Data & Society», 3, 1.
- Butler J. (1999), *Gender trouble: Feminism and the subversion of identity*, Routledge.
- Cai Z.G., Duan X., Haslett D.A., Wang S., Pickering M.J. (2024), Do large language models resemble humans in language use?, arXiv preprint arXiv:2303.08014.
- Callon M. (1984), Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay, «The Sociological Review», 32, 1_suppl, pp. 196-233.
- Callon M. (2001), "Actor-network theory", in N.J. Smelser, P.B. Baltes (a cura di), *International Encyclopedia of the Social and Behavioral Sciences*, Elsevier, Amsterdam, pp. 62-66.
- Callon M. (2007), "What does it mean to say that economics is performative", in Aa.Vv., *Do Economists Make Markets*.
- Callon M. (2016), Revisiting marketization: from interface-markets to market-agencements, «Consumption Markets & Culture», 19, 1, pp. 17-37.
- Caloffi A., Colovic A., Rizzoli V., Rossi F. (2023), Innovation intermediaries' types and functions: A computational analysis of the literature, «Technological Forecasting and Social Change», 189, 122351.
- Calzati S., De Kerckhove D. (2024), *Quantum ecology: Why and how new information technologies will reshape societies*, MIT Press, Cambridge.
- Campbell-Kelly M. (2003), *From airline reservations to sonic the hedgehog. A History of the Software Industry*, MIT Press, Cambridge-London.
- Campolo A., Crawford K. (2020), Enchanted determinism: Power without responsibility in artificial intelligence, «Engaging Science, Technology, and Society», 6, pp. 1-19.
- Cantó-Milà N., Seebach S. (a cura di) (2024), *Imaginaries of the future: An introduction*, Routledge.

- Carayannis E.G., Campbell D.F.J. (2009), “Mode 3” and “Quadruple Helix”: Toward a 21st century fractal innovation ecosystem, «International Journal of Technology Management», 46, 3-4, pp. 201-234.
- Carayannis E.G., Barth T.D., Campbell D.F.J. (2012), The Quintuple Helix innovation model: global warming as a challenge and driver for innovation, «Journal of Innovation and Entrepreneurship», 1, 1, p. 2.
- Carter M., Egliston B. (2021), What are the risks of Virtual Reality data? Learning analytics, algorithmic bias and a fantasy of perfect data, «New Media & Society», 25, 3, pp. 485-504.
- Castelo N., Bos M.W., Lehmann D.R. (2019), Task-dependent algorithm aversion, «Journal of Marketing Research», 56, 5, pp. 809-825.
- Cave S., Dihal K. (2019), Hopes and fears for intelligent machines in fiction and reality, «Nature Machine Intelligence», 1, 2, pp. 74-78.
- Ceron T., Falk N., Barić A., Nikolaev D., Padó S. (2024), Beyond Prompt Brittleness: Evaluating the Reliability and Consistency of Political Worldviews in LLMs, «Transactions of the Association for Computational Linguistics», 12, pp. 1378-1400.
- Chianese D. (in corso di stampa), “Learned Futures: using Future Methods on Generative AI systems”, in B. Aragona, R. Coleman, S. Halford (a cura di), *Methods for Sociodigital Futures Research: Approaches, Challenges, Openings*, Bristol University Press.
- Christiano P.F., Leike J., Brown T.B., Martic M., Legg S., Amodei D. (2017), “Deep reinforcement learning from human preferences”, in Aa.Vv., *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS’17)*, pp. 4302-4310.
- Christin A. (2020), The ethnographer and the algorithm: Beyond the black box, «Theory and Society», 49, 5-6, pp. 897-918.
- Chuang C.M. (2023), The conceptualization of smart tourism service platforms on tourist value co-creation behaviours: An integrative perspective of smart tourism services, «Humanities and Social Sciences Communications», 10, 1, p. 367.
- Cinelli M., De Francisci Morales G., Galeazzi A., Quattrocioni W., Starnini M. (2021), The echo chamber effect on social media, «Proceedings of the National Academy of Sciences», 118, 9, e2023301118.
- Collins H., Evans R. (2019), *Rethinking Expertise*, University of Chicago Press.
- Conte R., Gilbert N., Bonelli G., Cioffi-Revilla C., Deffuant G., Kertesz J., Helbing D. (2012), Manifesto of computational social science, «The European Physical Journal Special Topics», 214, 1, pp. 325-346.
- Corriere della Sera (2022, 22 aprile), Disastro di Chernobyl: quel 26 aprile del 1986..., disponibile al link: https://www.corriere.it/esteri/22_aprile_26/disastro-chernobyl-26-aprile-oggi-3bd6e44e-c417-11ec-b866-99e8752a9d0f.shtml
- Corriere della Sera (2023, 30 marzo), Il vento di Chernobyl soffia sul referendum per il nucleare, disponibile al link: https://www.corriere.it/cronache/23_marzo_30/vento-chernobyl-soffia-referendum-il-nucleare-7cd20222-ce79-11ed-86b7-ece540d3f123.shtml
- Cosgrove D. (2003), *Apollo’s eye: A cartographic genealogy of the earth in the western imagination*, Johns Hopkins University Press.

- Couldry N., Mejias U.A. (2020), *The costs of connection: How data are colonizing human life and appropriating it for capitalism*, Stanford University Press.
- Crampton J.W. (2011), *Mapping: A critical introduction to cartography and GIS*, John Wiley & Sons.
- Crang M., Graham S. (2007), Sentient cities: Ambient intelligence and the politics of urban space, «Information, Communication & Society», 10.
- Crawford K. (2021), *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*, Yale University Press.
- Crawford K., Joler V. (2018), *Anatomy of an AI System*, AI Now Institute.
- Creemers R. (2023), Cybersecurity law and regulation in China: Securing the smart state, «China Law and Society Review», 6, 2, pp. 111-145.
- Creswell J.W., Clark V.P. (2007), *Mixed methods research*, Sage, Thousand Oaks.
- Culot G., Orzes G., Sartor M., Nassimbeni G. (2020), The future of manufacturing: A Delphi-based scenario analysis on Industry 4.0, «Technological Forecasting and Social Change», 157, 120092.
- Curry M.R. (1998), *Digital places: Living with geographic information technologies*, Routledge.
- Davenport T.H., Harris J.G. (2005), Automated decision making comes of age, «MIT Sloan Management Review», 46, 4, p. 83.
- Dawkins O., Kitchin R. (2025), Urban digital twins: Digital twins for participatory steering, «New Media & Society», 27, 8, pp. 4402-4419.
- Del Casino V.J., Hanna S.P. (2005), Beyond the “binaries”: A methodological intervention for interrogating maps as representational practices, «ACME: An International Journal for Critical Geographies», 4, 1, pp. 34-56.
- Deleuze G., Guattari F. (1980), *Capitalisme et schizophrénie*, Les Editions de Minuit, Paris.
- Deleuze G., Guattari F. (2017), *Mille piani. Capitalismo e schizofrenia*, Orthotes, Napoli.
- Denis J., Mongili A., Pontille D. (2015), Maintenance & repair in Science and Technology Studies, «Tecnoscienza: Italian Journal of Science & Technology Studies», 6, 2, pp. 5-15.
- Devlin J., Chang M.-W., Lee K., Toutanova K. (2019), BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, arXiv preprint arXiv:1810.04805.
- Di Zio S., Calleo Y., Bolzan M. (2023), Delphi-based visual scenarios: an innovative use of generative adversarial networks, «Futures», 154, 103280.
- Dominguez-Olmedo R., Hardt M., Mender-Dünner C. (2024), Questioning the Survey Responses of Large Language Models, arXiv preprint arXiv:2306.07951.
- Douglas M., Wildavsky A. (1982), *Risk and culture: An essay on the selection of technical and environmental dangers*, University of California Press.
- Dourish P. (2017), *The Stuff of Bits: An Essay on the Materialities of Information*, MIT Press.
- Dourish P., Bell G. (2014), *Divining a digital future: Mess and mythology in ubiquitous computing*, MIT Press.
- Dowling J.P., Milburn G.J. (2003), Quantum technology: The second quantum revolution, «Philosophical Transactions of the Royal Society of London.

- Series A: Mathematical, Physical and Engineering Sciences», 361, 1809, pp. 1655-1674.
- Dunlap L., Mandal K., Darrell T., Steinhardt J., Gonzalez J.E. (2025), VibeCheck: Discover and Quantify Qualitative Differences in Large Language Models, arXiv preprint arXiv:2410.12851.
- Durmus E., Nguyen K., Liao T.I., Schiefer N., Askill A., Bakhtin A., Chen C., Hatfield-Dodds Z., Hernandez D., Joseph N., Lovitt L., McCandlish S., Sikder O., Tamkin A., Thamkul J., Kaplan J., Clark J., Ganguli D. (2024), Towards Measuring the Representation of Subjective Global Opinions in Language Models, arXiv preprint arXiv:2306.16388.
- Edquist C. (2009), "Systems of innovation: Perspectives and challenges", in J. Fagerberg, D.C. Mowery, R.R. Nelson (a cura di), *The Oxford handbook of innovation*, Oxford University Press, pp. 181-208.
- Edwards P.N. (2013), *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*, MIT Press.
- Edwards P.N., Mayernik M.S., Batcheller A.L., Bowker G.C., Borgman C.L. (2011), Science friction: Data, metadata, and collaboration, «Social Studies of Science», 41, 5, pp. 667-690.
- Emirbayer M., Goodwin J. (1994), Network analysis, culture, and the problem of agency, «American Journal of Sociology», 99, 6, pp. 1411-1454.
- Esposito E., Maturo A., Meloni C. (2021), The imaginaries of smart cities: Between global scripts and local adaptations, «Urban Studies», 58, 9, pp. 1773-1789.
- Etzkowitz H., Leydesdorff L. (2000), The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university-industry-government relations, «Research Policy», 29, 2, pp. 109-123.
- Etzkowitz H. (2003), Research groups as 'quasi-firms': the invention of the entrepreneurial university, «Research Policy», 32, 1, pp. 109-121.
- Eubanks V. (2018), *Automating inequality: How high-tech tools profile, police, and punish the poor*, St. Martin's Press.
- Fan A., Jernite Y., Perez E., Grangier D., Weston J., Auli M. (2019), ELI5: Long Form Question Answering, arXiv preprint arXiv:1907.09190.
- Farías I., Bender T. (a cura di) (2010), *Urban assemblages: How actor-network theory changes urban studies*, Routledge.
- Farrand B., Carrapiço H. (2015), "The governance of online expression in a networked world: free-speech, surveillance and censorship", in H. Carrapiço, B. Farrand (a cura di), *The Governance of Online Expression in a Networked World*, Routledge.
- Felt U. (2009), *Knowing and living in academic research: Convergences and heterogeneity in research cultures in the European context*, University of Vienna.
- Felt U., Fouché R., Miller C.A., Smith-Doerr L. (a cura di) (2016), *The handbook of science and technology studies*, MIT Press.
- Felt U., Wynne B., Callon M., Gonçalves M.E., Jasanoff S., Jepsen M., Tallacchini M. (2007), *Taking European knowledge society seriously*, Publications Office of the European Union, Luxembourg.

- Ferré-Bigorra J., Casals M., Gangolells M. (2022), The adoption of urban digital twins, «Cities», 131, 103905.
- Floridi L. (2022), *Etica dell'intelligenza artificiale: Sviluppi, opportunità, sfide*, a cura di M. Durante, Raffaello Cortina Editore, Milano.
- Foucault M. (1980), *Power/knowledge: Selected Interviews and Other Writings, 1972-1977*, Pantheon.
- Freeman L.C. (1978), Centrality in social networks conceptual clarification, «Social Networks», 1, 3, pp. 215-239.
- Freire P. (1996), *Pedagogy of hope: Reliving Pedagogy of the Oppressed*, trad. di R.R. Barr, Continuum, New York.
- Fricker M. (2007), *Epistemic Injustice: Power and the Ethics of Knowing*, Oxford University Press, Oxford.
- Friis K., Lysne O. (2021), Huawei, 5G and security: Technological limitations and political responses, «Development and Change», 52, 5, pp. 1174-1195.
- Frow P., Nenonen S., Payne A., Storbacka K. (2015), Managing co-creation design: A strategic approach to innovation, «British Journal of Management», 26, 3, pp. 463-483.
- Fuenfschilling L., Truffer B. (2014), The structuration of socio-technical regimes: Conceptual foundations from institutional theory, «Research Policy», 43, 4, pp. 772-791.
- Gabrys J. (2014), Programming environments: Environmentality and citizen sensing in the smart city, «Environment and Planning D: Society & Space», 32, 1, pp. 30-48.
- Gal M.S., Aviv O. (2020), The competitive effects of the GDPR, «Journal of Competition Law & Economics», 16, 3, pp. 349-391.
- Garud R., Tuertscher P., Van de Ven A.H. (2013), Perspectives on innovation processes, «Academy of Management Annals», 7, 1, pp. 775-819.
- Gebru T., Morgenstern J., Vecchione B., Vaughan J.W., Wallach H., Iii H.D., Crawford K. (2021), Datasheets for datasets, «Communications of the ACM», 64, 12, pp. 86-92.
- Geels F.W. (2004), "Understanding system innovations: a critical literature review and a conceptual synthesis", in Aa.Vv., *System innovation and the transition to sustainability: Theory, evidence and policy*, pp. 19-47.
- Geiger R.S. (2017), Beyond opening up the black box: Investigating the role of algorithmic systems in Wikipedian organizational culture, «Big Data & Society», 4, 2, pp. 1-14.
- Ghenai C., Husein L.A., Al Nahlawi M., Hamid A.K., Bettayeb M. (2022), Recent trends of digital twin technologies in the energy sector: A comprehensive review, «Sustainable Energy Technologies and Assessments», 54, 102837.
- Gherardi S. (2012), *How to conduct a practice-based study*, Edward Elgar.
- Gherardi S. (2016), "Sociomateriality in posthuman practice theory", in Aa.Vv., *The nexus of practices*, Routledge, pp. 50-63.
- Gherardi S., Murgia A., Bellè E., Miele F., Carreri A. (2019), Tracking the sociomaterial traces of affect at the crossroads of affect and practice theories, «Qualitative Research in Organizations and Management: An International Journal», 14, 3, pp. 295-316.
- Giddens A. (1990), *The Consequences of Modernity*, Polity Press.

- Gillespie T. (2014), "The relevance of algorithms", in Aa.Vv., *Media technologies: Essays on communication, materiality, and society*, MIT Press, pp. 167-194.
- Giorgi S., Liu T., Aich A., Isman K.J., Sherman G., Fried Z., Sedoc J., Ungar L., Curtis B. (2024), "Modeling Human Subjectivity in LLMs Using Explicit and Implicit Human Factors in Personas", in Y. Al-Onaizan, M. Bansal, Y.-N. Chen (a cura di), *Findings of the Association for Computational Linguistics: EMNLP 2024*, Association for Computational Linguistics, pp. 7174-7188.
- Gitelman L. (a cura di) (2013), *"Raw data" is an oxymoron*, MIT Press.
- Goodall N.J. (2014), "Machine ethics and automated vehicles", in Aa.Vv., *Road vehicle automation*, pp. 93-102.
- Grievies M. (2014), *Digital twin: manufacturing excellence through virtual factory replication*, White Paper, 1, pp. 1-7.
- Grievies M., Vickers J. (2017), "Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems", in Aa.Vv., *Transdisciplinary Perspectives on Complex Systems*, Springer International Publishing, pp. 85-113.
- Grote T., Berens P. (2020), On the ethics of algorithmic decision-making in healthcare, «Journal of Medical Ethics».
- Grunwald A. (2009), "Technology assessment: Concepts and methods", in A.W.M. Meijers (a cura di), *Philosophy of technology and engineering sciences*, Elsevier, pp. 1103-1146.
- Guenduez A.A., Mettler T. (2023), The role of narratives in the digital transformation of public organizations, «Government Information Quarterly», 40, 2, pp. 101-118.
- Guo F. (2023), GPT in Game Theory Experiments, arXiv preprint arXiv:2305.05516.
- Guston D.H. (2014), Understanding "anticipatory governance", «Social Studies of Science», 44, 2, pp. 218-242.
- Gutierrez Lopez M., Halford S. (2024), Explaining machine learning practice: Findings from an engaged science and technology studies project, «Information, Communication & Society», pp. 1-17.
- Hahn D., Munir A., Behzadan V. (2019), Security and privacy issues in intelligent transportation systems: Classification and challenges, «IEEE Intelligent Transportation Systems Magazine», 13, 1, pp. 181-196.
- Haider J., Sundin O. (2023), *Paradoxes of digital disconnection*, MIT Press.
- Halford S., Southerton D. (2023), *Futures in the making: Practices of anticipatory governance*, Polity Press.
- Halford S., Southerton D. (2024), Sociodigital Futures? An agenda for sociological research and practice, «Sociologia Italiana», 26.
- Haraway D. (1988), Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective, «Feminist Studies», 14, 3, pp. 575-599.
- Harley J.B. (1989), Deconstructing the map, «Cartographica: The International Journal for Geographic Information and Geovisualization», 26, 2, pp. 1-20.
- Hartmann J., Schwenzow J., Witte M. (2023), The political ideology of conversational AI: Converging evidence on ChatGPT's pro-environmental, left-libertarian orientation, arXiv preprint arXiv:2301.01768.
- Hesse-Biber S., Johnson R.B. (2013), Coming at things differently: Future directions of possible engagement with mixed methods research, «Journal of Mixed Methods Research», 7, 2, pp. 103-109.

- Hodapp D., Hawlitschek F., Kramer D. (2019), "Value co-creation in nascent platform ecosystems: A Delphi study in the context of the Internet of Things", in Aa.Vv., *Proceedings of the 40th International Conference on Information Systems (ICIS 2019)*, Munich.
- Holstein K., Vaughan J.W., Daumé H. III, Dudík M., Wallach H. (2018), Improving fairness in machine learning systems: What do industry practitioners need?, arXiv preprint arXiv:1812.05239.
- Howells J. (2006), Intermediation and the role of intermediaries in innovation, «Research Policy», 35, 5, pp. 715-728.
- Hughes T.P. (1993), *Networks of power: electrification in Western society, 1880-1930*, Johns Hopkins University Press.
- Hull M.S. (2012), *Government of paper: The materiality of bureaucracy in urban Pakistan*, University of California Press.
- Introna L.D. (2013), "Epilogue: Performativity and the becoming of sociomaterial assemblages", in Aa.Vv., *Materiality and space: Organizations, artefacts and practices*, Palgrave Macmillan, London, pp. 330-342.
- Jasanoff S. (2003), Technologies of Humility: Citizen Participation in Governing Science, «Minerva», 41, 3, pp. 223-244.
- Jasanoff S. (2004), *States of Knowledge: The Co-production of Science and Social Order*, Routledge.
- Jasanoff S. (2004), "The idiom of co-production", in S. Jasanoff (a cura di), *States of Knowledge*, Routledge, pp. 1-12.
- Jasanoff S. (2005), *Designs on nature: Science and democracy in Europe and the United States*, Princeton University Press.
- Jasanoff S. (2015), *Dreamscapes of modernity: Sociotechnical imaginaries and the fabrication of power*, University of Chicago Press.
- Jasanoff S. (2015), "Future imperfect: Science, technology, and the imaginations of modernity", in S. Jasanoff, S.-H. Kim (a cura di), *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*, University of Chicago Press, Chicago, pp. 1-33.
- Jasanoff S., Kim S.H. (2009), Containing the atom: Sociotechnical imaginaries and nuclear power in the United States and South Korea, «Minerva», 47, 2, pp. 119-146.
- Jasanoff S., Kim S.-H. (a cura di) (2015), *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*, University of Chicago Press.
- Jensen C.B., Morita A. (2017), "Anthropology as a study of the multiverse: Ontology and the work of speculation", in C.B. Jensen, A. Morita (a cura di), *Infrastructures and social complexity: A companion*, pp. 121-137.
- Jiang H., Zhang X., Cao X., Breazeal C., Roy D., Kabbara J. (2024), PersonaLLM: Investigating the Ability of Large Language Models to Express Personality Traits, arXiv preprint arXiv:2305.02547.
- Joyce K., Smith-Doerr L., Alegria S., Bell S., Cruz T., Hoffman S.G., Noble S.U., Shestakofsky B. (2021), Toward a Sociology of Artificial Intelligence: A Call for Research on Inequalities and Structural Change, «Socius», 7, 2378023121999581.

- Kaku M. (2023), *Quantum supremacy: How the quantum computer revolution will change everything*, Knopf Doubleday Publishing Group, New York.
- Kanger L., Schot J. (2019), Deep transitions: Theorizing the long-term patterns of socio-technical change, «Environmental Innovation and Societal Transitions», 32, pp. 7-21.
- Katzenbach C. (2021), Imagining digital futures: Shaping the role of platform governance, «Internet Policy Review», 10, 2.
- Kitchin R. (2014), *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*, SAGE.
- Kitchin R. (2016), The ethics of smart cities and urban science, «Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences», 374, 2083, 20160115.
- Kitchin R. (2017), Thinking critically about and researching algorithms, «Information, Communication & Society», 20, 1, pp. 14-29.
- Kitchin R. (2021), *Data Lives: How Data Are Made and Shape Our World*, Bristol University Press.
- Kitchin R., Lauriault T. (2014), *Towards critical data studies: Charting and unpacking data assemblages and their work*.
- Kitchin R., Lauriault T.P., McArdle G. (2015), Knowing and governing cities through urban indicators, city benchmarking and real-time dashboards, «Regional Studies, Regional Science», 2, 1, pp. 6-28.
- Knorr Cetina K. (2001), "Objectual Practice", in T.R. Schatzki, K. Knorr Cetina, E. von Savigny (a cura di), *The Practice Turn in Contemporary Theory*, Routledge, pp. 175-188.
- Koch G. (2017), "The ethnography of infrastructures: digital humanities and cultural anthropology", in Aa.Vv., *Cultural Heritage Infrastructures in Digital Humanities*, pp. 63-81.
- Konrad K., Böhle K. (2019), Socio-technical futures and the governance of innovation processes: An introduction to the special issue, «Futures», 109, pp. 101-107.
- Kozachek D. (2026), Entering the age of hybrid futures: A comparative study of human and GPT-generated scenarios, «Futures», 176, 103705.
- Kritzinger W., Karner M., Traar G., Henjes J., Sihn W. (2018), Digital Twin in manufacturing: A categorical literature review and classification, «IFAC-PapersOnLine», 51, 11, pp. 1016-1022.
- Kuhlmann S., Rip A. (2018), Next-generation innovation policy and grand challenges, «Science & Public Policy», 45, 4, pp. 448-454.
- La Repubblica (2022, 9 novembre), 9 novembre 1987: il referendum antinucleare in Italia, disponibile al link: https://www.repubblica.it/rubriche/urs/2022/11/09/video/stefano_massini_ufficio_racconti_smarriti_-_9_novembre_1987_il_referendum_antinucleare_in_italia-422548168/
- Lakoff G., Johnson M. (2008), *Metaphors we live by*, University of Chicago Press. (Orig. pubb. 1980).
- Lash S. (2007), Power after hegemony: Cultural studies in mutation?, «Theory, Culture & Society», 24, 3, pp. 55-78.
- Latour B. (1987), *Science in action: How to follow scientists and engineers through society*, Harvard University Press, Cambridge.

- Latour B. (1992), "Where are the missing masses? The sociology of a few mundane artifacts", in W. Bijker, J. Law (a cura di), *Shaping Technology-Building Society. Studies in Sociotechnical Change*, MIT Press, Cambridge, pp. 225-259.
- Latour B. (1999), *Pandora's hope: Essays on the reality of science studies*, Harvard University Press, Cambridge.
- Latour B. (2005), *Reassembling the social: An introduction to actor-network-theory*, Oxford University Press.
- Law J. (1992), Notes on the theory of the actor-network: Ordering, strategy, and heterogeneity, «Systems Practice», 5, 4, pp. 379-393.
- Law J. (2004), *After Method: Mess in Social Science Research*, Routledge.
- Law J. (2008), "Actor-network theory and material semiotics", in B.S. Turner (a cura di), *The new Blackwell companion to social theory*, Blackwell, pp. 141-158.
- Law J., Singleton V. (2000), Performing technology's stories: On social constructivism, performance, and performativity, «Technology and Culture», 41, 4, pp. 765-775.
- Leonelli S. (2019), *Data-centric biology: A philosophical study*, University of Chicago Press.
- Levidow L., Papaioannou T. (2013), *Confronting uncertainties in the bioeconomy: Governance, sustainability, and innovation*, Routledge.
- Li X., Li Y., Qiu L., Joty S., Bing L. (2024), Evaluating Psychological Safety of Large Language Models, arXiv preprint arXiv:2212.10529.
- Light B., Burgess J., Duguay S. (2018), The walkthrough method: An approach to the study of apps, «New Media & Society», 20, 3, pp. 881-900.
- Liu Y., Panwang Y., Gu C. (2025), "Turning right"? An experimental study on the political value shift in large language models, «Humanities and Social Sciences Communications», 12, 1, pp. 1-10.
- Luciani M. (2018), La decisione giudiziaria robotica, «Rivista AIC», 3.
- Luhmann N. (1995), *Social systems*, Stanford University Press.
- Mackenzie A. (2017), *Machine learners: Archaeology of a data practice*, MIT Press.
- Macnaghten P., Givant J. (2020), Converging narratives: Artificial intelligence, risk and governance, «Public Understanding of Science», 29, 1, pp. 3-14.
- Madaio M.A., Stark L., Vaughan J.W., Wallach H. (2020), "Co-designing checklists to understand organizational challenges and opportunities around fairness in AI", in Aa.Vv., *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, pp. 1-14.
- Mager A., Katzenbach C. (2021), Future imaginaries in the making: Media, infrastructures and the politics of anticipation, «Communication Theory», 31, 4, pp. 674-695.
- Manon H.S., Temkin D. (2011), Notes on glitch, «World Picture», 6, p. 118.
- Mariani K., Lozada F.V. (2023), The use of AI and algorithms for decision making in workplace recruitment practices, «Journal of Student Research», 12, 1.
- Marres N. (2017), *Digital sociology: The reinvention of social research*, John Wiley & Sons.

- Marx L. (2000), *The machine in the garden: Technology and the pastoral ideal in America*, Oxford University Press.
- McAfee A., Brynjolfsson E. (2018), *Machine, Platform, Crowd: Harnessing Our Digital Future*, Norton.
- McCarthy J., Minsky M.L., Rochester N., Shannon C.E. (2006), A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955, «AI Magazine», 27, 4, pp. 12-12.
- McIntosh T.R., Liu T., Susnjak T., Watters P., Ng A., Halgamuge M.N. (2024), A Culturally Sensitive Test to Evaluate Nuanced GPT Hallucination, «IEEE Transactions on Artificial Intelligence», 5, 6, pp. 2739-2751.
- Meskó B., Drobni Z., Bényei É., Gergely B., Györfly Z. (2017), Digital health is a cultural transformation of traditional healthcare, «Mhealth», 3.
- Meuser M., Nagel U. (2009), “The expert interview and changes in knowledge production”, in A. Bogner, B. Littig, W. Menz (a cura di), *Interviewing Experts*, Palgrave Macmillan, London, pp. 17-42.
- Mikolov T., Chen K., Corrado G., Dean J. (2013), Efficient Estimation of Word Representations in Vector Space, arXiv preprint arXiv:1301.3781.
- Miller C.A. (2005), *Science and democracy: Making knowledge and making power in the biosciences and beyond*, Routledge.
- Miller T. (2019), Explanation in artificial intelligence: Insights from the social sciences, «Artificial Intelligence», 267, pp. 1-38.
- Ministero delle Imprese e del Made in Italy (2025), *Strategia Italiana per le Tecnologie Quantistiche*, Ministero delle Imprese e del Made in Italy, Roma.
- Mische A. (2009), Projects and possibilities: Researching futures in action, «Sociological Forum», 24, 3, pp. 694-704.
- Mitchell M., Wu S., Zaldivar A., Barnes P., Vasserman L., Hutchinson B., Spitzer E., Raji I.D., Gebru T. (2019), “Model cards for model reporting”, in Aa.Vv., *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 220-229.
- Mittelstadt B.D., Allo P., Taddeo M., Wachter S., Floridi L. (2016), The ethics of algorithms: Mapping the debate, «Big Data & Society», 3, 2, 2053951716679679.
- Mökander J., Axente M. (2023), Ethics-based auditing of automated decision making systems: intervention points and policy implications, «AI & Society», 38, 1, pp. 153-171.
- Mökander J., Morley J., Taddeo M., Floridi L. (2021), Ethics-based auditing of automated decision-making systems: Nature, scope, and limitations, «Science and Engineering Ethics», 27, 4, p. 44.
- Mol A. (1999), Ontological politics. A word and some questions, «The Sociological Review», 47, S1, pp. 74-89.
- Moore P.V. (2020), *The Quantified Self in Precarity: Work, Technology and Capabilities*, Routledge.
- Morozov E. (2013), *To save everything, click here. The folly of technological solutionism*, PublicAffairs, New York.
- Mostafa N., Hamdy W., Alawady H. (2019), Impacts of Internet of Things on supply chains: A framework for warehousing, «Social Sciences», 8, 3, p. 84.

- Motoki F.Y.S., Pinho Neto V., Rangel V. (2025), Assessing political bias and value misalignment in generative artificial intelligence, «Journal of Economic Behavior & Organization», 106904.
- Munk A.K., Meunier A., Venturini T. (2019), “Data sprints: A collaborative format in digital controversy mapping”, in Aa.Vv., *Digital STS: A Field Guide for Science & Technology Studies*, pp. 472-496.
- Natale S. (2016), *Inverni ed estati dell'intelligenza artificiale: Metafore, miti e cicli di entusiasmo tecnologico*, Il Mulino, Bologna.
- Natale S., Ballatore A. (2018), Imagining the thinking machine: Technological myths and the rise of artificial intelligence, «Convergence», 24, 6, pp. 627-641.
- Natale S., Ballatore A. (2020), Imagining the future of AI: Anthropomorphic and machine metaphors in public discourse, «New Media & Society», 22, 11, pp. 1-19.
- Negri E., Fumagalli L., Macchi M. (2017), A review of the roles of digital twin in CPS-based production systems, «Procedia Manufacturing», 11, pp. 939-948.
- Nelson L.K. (2020), Computational grounded theory: A methodological framework, «Sociological Methods & Research», 49, 1, pp. 3-42.
- Nelson R.R. (1993), *National Innovation Systems: A Comparative Analysis*, Oxford University Press.
- Nicolini D. (2012), *Practice theory, work, and organization: An introduction*, Oxford University Press.
- Nielsen M.A., Chuang I.L. (2010), *Quantum computation and quantum information*, Cambridge University Press, Cambridge.
- Noble S.U. (2018), *Algorithms of oppression: How search engines reinforce racism*, New York University Press.
- Nye D.E. (1996), *American technological sublime*, MIT Press.
- O'Neil C. (2016), *Weapons of math destruction: How big data increases inequality and threatens democracy*, Crown Random House, New York.
- OECD (2019), *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449), OECD.
- OECD (2021), *State of implementation of the OECD AI Principles: Insights from national AI policies* (OECD Digital Economy Papers No. 311), OECD Publishing.
- OECD (2023), *The state of implementation of the OECD AI Principles four years on* (OECD Artificial Intelligence Papers No. 3), OECD Publishing.
- OECD (2024), *Policy framework for quantum technologies*, Organisation for Economic Co-operation and Development, Paris.
- Orlikowski W.J. (1992), The duality of technology: Rethinking the concept of technology in organizations, «Organization Science», 3, 3, pp. 398-427.
- Orlikowski W.J., Iacono C.S. (2001), Research commentary: Desperately seeking the “IT” in IT research—A call to theorizing the IT artifact, «Information Systems Research», 12, 2, pp. 121-134.
- Osterwalder A., Pigneur Y. (2010), *Business model generation: a handbook for visionaries, game changers, and challengers*, John Wiley & Sons.
- Osterwalder A., Pigneur Y., Bernarda G., Smith A. (2015), *Value proposition design: How to create products and services customers want*, John Wiley & Sons.

- Owen R., Stilgoe J., Macnaghten P., Gorman M., Fisher E., Guston D. (2013), "A framework for responsible innovation", in R. Owen, J. Stilgoe, P. Macnaghten, M. Gorman, E. Fisher (a cura di), *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*, Wiley, Chichester, pp. 27-50.
- Paliwoda S.J. (1983), Predicting the future using Delphi, «Management Decision», 21, 1, pp. 31-38.
- Parsons T. (1951), *The Social System*, Routledge & Kegan Paul.
- Pasquale F. (2015), *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard University Press, Cambridge.
- Pickles J. (1995), *Ground truth: The social implications of geographic information systems*, Guilford Press.
- Pinch T.J., Bijker W.E. (1984), The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other, «Social Studies of Science», 14, 3, pp. 399-441.
- Pink S. (2019), Digital social futures research, «Journal of Digital Social Research», 1, 1, pp. 41-48.
- Pink S. (2025), Artificial intelligence and the futures turn: an anticipatory infrastructure for qualitative methods, «Qualitative Research in Psychology», 0, 0, pp. 1-20.
- Plantin J.-C., Lagoze C., Edwards P.N., Sandvig C. (2018), Infrastructure studies meet platform studies in the age of Google and Facebook, «New Media & Society», 20, 1, pp. 293-310.
- Powell W.W., Grodal S. (2006), "Networks of innovators", in Aa.Vv., *The Oxford Handbook of Innovation*, Oxford University Press, pp. 56-85.
- Qiu Y., Yao H., Ren P., Tian X., You M. (2025), Regulatory sandbox expansion: Exploring the leap from fintech to medical artificial intelligence, «Intelligent Oncology», 1, 2, pp. 120-127.
- Qu Y., Wang J. (2024), Performance and biases of Large Language Models in public opinion simulation, «Humanities and Social Sciences Communications», 11, 1, pp. 1-13.
- Radford A., Kim J.W., Hallacy C., Ramesh A., Goh G., Agarwal S., Sastry G., Askell A., Mishkin P., Clark J., Krueger G., Sutskever I. (2021), Learning Transferable Visual Models From Natural Language Supervision, arXiv preprint arXiv:2103.00020.
- Ranga M., Etzkowitz H. (2013), Triple Helix systems: an analytical framework for innovation policy and practice in the Knowledge Society, «Industry and Higher Education», 27, pp. 237-262.
- Ratto M., Boler M. (a cura di) (2014), *DIY citizenship: Critical making and social media*, MIT Press.
- Renzi A.B., Freitas S. (2015), The Delphi method for future scenarios construction, «Procedia Manufacturing», 3, pp. 5785-5791.
- Ribes D., Bowker G.C. (2009), Between meaning and machine: Learning to represent the knowledge of communities, «Information and Organization», 19, 4, pp. 199-217.
- Ribes D., Jackson S.J. (2013), "Data bite man: The work of sustaining a long-term study", in L. Gitelman (a cura di), *"Raw data" is an oxymoron*, MIT Press, pp. 147-166.

- Richardson R. (2021), Defining and demystifying automated decision systems, «Maryland Law Review», 81, p. 785.
- Richterich A. (2018), *The big data agenda: Data ethics and critical data studies*, University of Westminster Press.
- Rip A. (2018), “De facto governance of nanotechnologies”, in Aa.Vv., *Futures of Science and Technology in Society*, Springer Fachmedien Wiesbaden, pp. 75-96.
- Rogers E.M. (2003), *Diffusion of Innovations*, Free Press.
- Rosner D.K. (2020), *Critical fabulations: Reworking the methods and margins of design*, MIT Press.
- Rosyidah A., Fajriyah N. (2022), AI narratives and algorithmic governance in Southeast Asia, «Journal of Asian Public Policy», 15, 3, pp. 351-367.
- Rouvroy A., Berns T. (2013), Algorithmic governmentality and prospects of emancipation: Disparateness as a precondition for individuation through relationships?, «Réseaux», 177, pp. 163-196.
- Royal Society (2020), *Quantum technologies: Policy considerations*, Royal Society, London.
- Ruppert E., Isin E., Bigo D. (2017), Data politics, «Big Data & Society», 4.
- Ruppert E., Law J., Savage M. (2013), Reassembling social science methods: The challenge of digital devices, «Theory, Culture & Society», 30, 4, pp. 22-46.
- Sablitzky T. (2022), The Delphi method, «Hypothesis: Research Journal for Health Information Professionals», 34, 1.
- Saltelli A., Di Fiore M. (a cura di) (2023), *The politics of modelling: Numbers between science and policy*, Oxford University Press.
- Salvini P., Reinmund T., Hardin B., Grieman K., Ten Holter C., Johnson A., Jirotko M. (2023), Human involvement in autonomous decision making systems. Lessons learned from three case studies in aviation, social care and road vehicles, «Frontiers in Political Science», 5, 1238461.
- Scherrer N., Shi C., Feder A., Blei D.M. (2023), *Evaluating the Moral Beliefs Encoded in LLMs*.
- Schmidt V.A. (2008), Discursive institutionalism: The explanatory power of ideas and discourse, «Annual Review of Political Science», 11, pp. 303-326.
- Schrotter G., Hürzeler C. (2020), The digital twin of the city of Zurich for urban planning, «PFG – Journal of Photogrammetry Remote Sensing and Geoinformation Science», 88, 1, pp. 99-112.
- Scott J.C. (1999), *Seeing like a state: How certain schemes to improve the human condition have failed*, Yale University Press.
- Seaver N. (2017), Algorithms as culture: Some tactics for the ethnography of algorithmic systems, «Big Data & Society», 4, 2, pp. 1-12.
- Senato della Repubblica (2011), Dossier n. 273, L'energia nucleare in Italia, disponibile al link: https://www.senato.it/documenti/repository/dossier/studi/2011/Dossier_273.pdf
- Sharkey A. (2014), Robots and human dignity: a consideration of the effects of robot care on the dignity of older people, «Ethics and Information Technology», 16, pp. 63-75.
- Shelton T., Zook M., Wiig A. (2015), The ‘actually existing smart city’, «Cambridge Journal of Regions Economy and Society», 8, 1, pp. 13-25.

- Shkedova O., Sester M. (2022), Visualization of Space Occupancy Uncertainty in a 3D Voxel-based Urban Model, «ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences», V-4-2022, pp. 357-364.
- Silverstone R., Hirsch E. (1992), *Consuming technologies*, Taylor & Francis.
- Simonsen J., Robertson T. (a cura di) (2013), *Routledge international handbook of participatory design*, Routledge.
- Star S.L. (1999), The ethnography of infrastructure, «The American Behavioral Scientist», 43, 3, pp. 377-391.
- Star S.L., Bowker G.C. (2010), “How to Infrastructure”, in Aa.Vv., *Handbook of New Media: Social Shaping and Social Consequences of ICTs*, SAGE Publications, pp. 230-245.
- Star S.L., Ruhleder K. (1996), Steps toward an ecology of infrastructure: Design and access for large information spaces, «Information Systems Research», 7, 1, pp. 111-134.
- Steen M. (2013), Co-design as a process of joint inquiry and imagination, «Design Issues», 29, 2, pp. 16-28.
- Steenbergen M.R., Marks G. (2007), Evaluating expert judgments, «European Journal of Political Research», 46, 3, pp. 347-366.
- Stilgoe J., Owen R., Macnaghten P. (2013), Developing a framework for responsible innovation, «Research Policy», 42, 9, pp. 1568-1580.
- Suchman L. (2007), *Human-machine reconfigurations: Plans and situated actions*, Cambridge University Press.
- Sundin O., Haider J. (2023), “Sociomateriality”, in Aa.Vv., *Information Literacy through Theory*, Facet Publishing.
- Tabassi E. (2023), *Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1)*, National Institute of Standards and Technology.
- Tan W. (2022), Narratives of AI governance in policy discourse, «Government Information Quarterly», 39, 4.
- Tanenbaum A. (2009), *Modern operating systems*, Pearson Education.
- Taubenfeld A., Dover Y., Reichart R., Goldstein A. (2024), “Systematic Biases in LLM Simulations of Debates”, in Y. Al-Onaizan, M. Bansal, Y.-N. Chen (a cura di), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, pp. 251-267.
- Taylor L. (2017), What is data justice? The case for connecting digital rights and freedoms globally, «Big Data & Society», 4, 2.
- Timmermans S., Epstein S. (2010), A world of standards but not a standard world: Toward a sociology of standards and standardization, «Annual Review of Sociology», 36, 1, pp. 69-89.
- Törnberg P., Uitermark J. (2021), For a heterodox computational social science, «Big Data & Society», 8, 2, 20539517211047725.
- Triandafyllidou A. (1995), The Chernobyl Accident in the Italian Press: A Media Story-Line, «Discourse & Society», 6, 4, pp. 517-536.
- Truby J., Brown R.D., Ibrahim I.A., Caudevilla Parellada O. (2022), A sandbox approach to regulating high-risk artificial intelligence applications, «European Journal of Risk Regulation», 13, 2, pp. 270-294.
- Turpin M., Michael J., Perez E., Bowman S.R. (2023), Language Models Don't Always Say What They Think: Unfaithful Explanations in Chain-of-Thought Prompting, arXiv preprint arXiv:2305.04388.

- UNESCO (2021), *Recommendation on the ethics of artificial intelligence*, UNESCO.
- Urman A., Makhortykh M. (2025), The silence of the LLMs: Cross-lingual analysis of guardrail-related political bias and false information prevalence in ChatGPT, Google Bard (Gemini), and Bing Chat, «Telematics and Informatics», 96, 102211.
- Van Lente H. (2016), *Technological expectations and the emergence of new technologies*, Edward Elgar.
- van Zoonen L. (2016), Privacy concerns in smart cities, «Government Information Quarterly», 33, 3, pp. 472-480.
- Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser Ł., Polosukhin I. (2017), “Attention is all you need”, in Aa.Vv., *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS’17)*, pp. 6000-6010.
- Venturini T. (2010), Diving in magma: how to explore controversies with actor-network theory, «Public Understanding of Science», 19, 3, pp. 258-273.
- Vermaas P. (2025), A Vision on Societal Research on Quantum Technologies in Seven Directions: Introducing the Topical Collection Quantum Society, «Digital Society», 4, 1, p. 13.
- von Schomberg R. (2013), “A vision of responsible research and innovation”, in R. Owen, J. Bessant, M. Heintz (a cura di), *Responsible innovation: Managing the responsible emergence of science and innovation in society*, Wiley, pp. 51-74.
- Voorwinden A. (2021), The privatised city: Technology and public-private partnerships in the smart city, «Law, Innovation and Technology», 13, 2, pp. 439-463.
- Watson C., Kharrufa A. (2021), “HCI-H is also for hazard: Using HAZOP to identify undesirable consequences in socio-technical systems”, in Aa.Vv., *Proceedings of the 4th ACM SIGCAS Conference on Computing and Sustainable Societies*, pp. 230-242.
- Weil C., Bibri S.E., Longchamp R., Golay F., Alahi A. (2023), Urban digital twin challenges: A systematic review and perspectives for sustainable smart cities, «Sustainable Cities and Society», 99, 104862.
- Woolgar S. (1985), Why not a Sociology of Machines? The Case of Sociology and Artificial Intelligence, «Sociology», 19, 4, pp. 557-572.
- World Nuclear Association (2024), Italy – Country Profile, disponibile al link: <https://world-nuclear.org/information-library/country-profiles/countries-g-n/italy>
- Wyatt S. (2017), “Metaphors and the Internet: A contextual approach”, in J. Burgess et al. (a cura di), *The Routledge handbook of language and digital communication*, Routledge, pp. 285-298.
- Zhou D., Zhang Y. (2024), Political biases and inconsistencies in bilingual GPT models—The cases of the U.S. and China, «Scientific Reports», 14, 1, 25048.
- Zuboff S. (2019), *The age of surveillance capitalism: the fight for a human future at the new frontier of power*, PublicAffairs, New York.
- Zuboff S. (2023), “The age of surveillance capitalism”, in W. Longhofer, D. Winchester (a cura di), *Social theory re-wired: New connections to classical and contemporary perspectives*, Routledge.

Gli autori

Suania Acampa - Professore aggiunto di Sociologia della Pubblica Amministrazione, Scuola Interdipartimentale di Economia e Giurisprudenza, Università degli Studi di Napoli Parthenope.

Francesco Amato - Borsista di ricerca in Scienze sociali e statistiche, Dipartimento di Scienze Sociali, Università degli Studi di Napoli Federico II.

Enrica Amaturo - Professoressa ordinaria di Sociologia generale, Dipartimento di Scienze Sociali, Università degli Studi di Napoli Federico II.

Biagio Aragona - Professore ordinario di Sociologia generale, Dipartimento di Scienze Sociali, Università degli Studi di Napoli Federico II.

Dario Chianese - Dottorando in Scienze sociali e statistiche, Dipartimento di Scienze Sociali, Università degli Studi di Napoli Federico II.

Mattia De Angelis - Dottore di ricerca in Scienze sociali e statistiche, Dipartimento di Scienze Sociali, Università degli Studi di Napoli Federico II.

Vincenzo Esposito - Borsista di ricerca, Dipartimento di Studi Letterari, Linguistici e Comparati, Università degli Studi di Napoli L'Orientale.

Cristiano Felaco - Professore associato di Sociologia generale, Dipartimento di Scienze Sociali, Università degli Studi di Napoli Federico II.

Questo LIBRO



ti è piaciuto?

Comunicaci il tuo giudizio su:

www.francoangeli.it/opinione



**VUOI RICEVERE GLI AGGIORNAMENTI
SULLE NOSTRE NOVITÀ
NELLE AREE CHE TI INTERESSANO?**



ISCRIVITI ALLE NOSTRE NEWSLETTER

SEGUICI SU:



FrancoAngeli

La passione per le conoscenze

Vi aspettiamo su:

www.francoangeli.it

per scaricare (gratuitamente) i cataloghi delle nostre pubblicazioni

DIVISI PER ARGOMENTI E CENTINAIA DI VOCI: PER FACILITARE
LE VOSTRE RICERCHE.



Management, finanza,
marketing, operations, HR

Psicologia e psicoterapia:
teorie e tecniche

Didattica, scienze
della formazione

Economia,
economia aziendale

Sociologia

Antropologia

Comunicazione e media

Medicina, sanità



Architettura, design,
arte, territorio

Informatica, ingegneria
Scienze

Filosofia, letteratura,
linguistica, storia

Politica, diritto

Psicologia, benessere,
autoaiuto

Efficacia personale

Politiche
e servizi sociali



FrancoAngeli

La passione per le conoscenze

Il volume esplora il rapporto tra l'evoluzione tecnologica e la produzione di conoscenza in un'epoca segnata dalla diffusione di sistemi decisionali automatizzati, LLM, gemelli digitali e tecnologie quantistiche. Attraverso un approccio interdisciplinare e critico, il testo analizza le tecnologie emergenti non solo come oggetti di studio, ma anche come strumenti di ricerca che si stanno progressivamente inserendo nelle pratiche della ricerca sociale. Il libro propone inoltre una riflessione metodologica sulle nuove forme di ricerca sociotecnica, approfondendo temi quali immaginari sociotecnici, valutazione, simulazione, prompting e co-creazione dell'innovazione.

Frutto delle attività dell'Osservatorio FAST - Futuro, Algoritmi, Società e Tecnologie dell'Università degli Studi di Napoli Federico II, il volume si configura come una guida critica per studiosi, studenti e ricercatori interessati a comprendere le implicazioni sociali delle tecnologie emergenti e le trasformazioni contemporanee della metodologia della ricerca sociale nell'era digitale.

Enrica Amato è professore ordinario di Sociologia generale presso il Dipartimento di Scienze Sociali dell'Università degli Studi di Napoli Federico II. Nel corso della sua carriera ha ricoperto incarichi apicali di governo accademico, tra cui quelli di preside della Facoltà di Sociologia e direttrice del Dipartimento di Scienze Sociali. Nel 2020 è stata insignita dal Presidente della Repubblica del titolo di Ufficiale dell'Ordine al Merito della Repubblica Italiana.

Biagio Aragona è professore ordinario di Sociologia generale presso il Dipartimento di Scienze Sociali dell'Università degli Studi di Napoli Federico II. Coordinatore del Corso di laurea magistrale in Sociologia Digitale e Analisi del Web, è direttore dell'Osservatorio FAST e vicedirettore del Southern Centre for Digital Transformations.

Suana Acampa, dottoressa di ricerca in Scienze Sociali e Statistiche, è professoressa aggiunta presso l'Università degli Studi di Napoli Parthenope. È stata ricercatrice post-doc presso il Southern Centre for Digital Transformation dell'Università degli Studi di Napoli Federico II.

Francesco Amato, dottore di ricerca in Scienze Sociali e Statistiche, è borsista di ricerca presso l'Università degli Studi di Napoli Federico II. I suoi interessi di ricerca si concentrano sugli aspetti metodologici ed epistemologici della ricerca sociale digitale.