FEDERICO PILATI

SOCIOLOGIA DIGITALE 2.0

INTELLIGENZA ARTIFICIALE GENERATIVA COME OGGETTO E STRUMENTO DELLA RICERCA SOCIALE

Ledizioni
The Innovative LEDIpublishing Company

Federico Pilati

SOCIOLOGIA DIGITALE 2.0

Intelligenza Artificiale Generativa come oggetto e strumento della ricerca sociale

Ledizioni

Via Antonio Boselli, 10 – 20136 Milano – Italy
www.ledizioni.it
info@ledizioni.it

Federico Pilati, *Sociologia digitale 2.0. L'Intelligenza Artificiale Generativa come oggetto e strumento della ricerca sociale*

Seconda edizione: novembre 2025

ISBN cartaceo: 9791256003952
ISBN eBook: 9791256003969
ISBN PDF in Open Access: 9791256005819

Progetto grafico: ufficio grafico Ledizioni
In copertina: immagine di Mathias Reding su unsplash.com.

Informazioni sul catalogo e sulle ristampe dell'editore: www.ledizioni.it

Indice

Prefazione

di Elena Esposito

Come usare la macchia cieca

Questo libro, che dichiara esplicitamente di non essere un manuale e di non volerlo essere, è di fatto la cosa più vicina possibile a una introduzione all'Intelligenza Artificiale Generativa (IAG) – cioè, come si legge giustamente nel titolo, a una nuova sociologia digitale. Il lettore ottiene un'idea chiara e esaustiva di quello che si sa oggi sulla IAG – cioè di fatto che una serie di questioni sono ancora aperte, e mettono in discussione non solo i fondamenti della vita sociale, ma anche quelli della disciplina con cui li osserviamo, la sociologia. È vero quindi che non è un manuale che fornisce istruzioni chiare e non controverse, perché un plausibile manuale dell'IAG non ci può essere.

Sociologia Digitale 2.0 è un testo introduttivo che non presuppone conoscenze tecniche e ci accompagna passo per passo in un percorso che può trattare a volte anche di questioni estremamente tecniche, ma le rende sempre accessibili. Tutti conosciamo ChatGPT e simili strumenti dopo il lancio fragoroso nel Novembre 2022, ma spesso in modo lacunoso. Qui si comincia dall'inizio e si segue lo sviluppo della IAG: da dove viene, come lavora, che cosa fa e come può essere usata. In tutti questi passaggi,

Federico Pilati ci porta a confrontarci con questioni nuove e con modi nuovi di gestirle, mostrando nello stesso tempo il loro impatto destabilizzante e il fermento di soluzioni inventive che stanno emergendo. Tipica del testo in molti passaggi è la contrapposizione tra i "profondi ripensamenti epistemologici e pratici" che ci sono richiesti e le proposte innovative con cui li si sta affrontando. Se ne esce un po' frastornati ma tutto sommato fiduciosi.

Ci si confronta con problemi dilaganti e irrisolti: il bias dei dati e dei programmi, l'intrasparenza degli algoritmi, gli enigmi della gestione di una privacy irrinunciabile e impossibile, l'esigenza contraddittoria di regolamentazione e di creatività, il devastante impatto ambientale e sociale di processi che sembrano irreversibili. Il libro non offre soluzioni (non potrebbe), ma mostra come la pressione di queste sfide stia portando a un'estesa e dinamica revisione degli approcci per trattarle. Siamo costretti a ripensare la posizione centrale degli esseri umani, il ruolo degli artefatti e dei media, la distribuzione geografica e politica delle risorse e delle prospettive, la neutralità degli osservatori e dei discorsi, e molto altro.

Non che tutto questo sia nuovo. Ciascuno di questi elementi è stato evidenziato e discusso da tempo con varie etichette: post-umanesimo, Actor-Network-Theory (ANT), Science and Technology Studies (STS), teoria critica, teoria dei sistemi, post-colonialismo, performatività, e altro. Pilati li richiama, ma mostra come il riferimento alla IAG li combini e focalizzi sotto la spinta di interrogativi inediti. E ci fa anche una sua proposta, la "riflessività sintetica": "la capacità di riconoscere e analizzare criticamente non solo la propria soggettività di ricercatori umani, ma anche quella degli strumenti algoritmici utilizzati, nonché le modalità di interazione e co-costruzione tra queste diverse forme di soggettività" (p. 8). Il testo poi estende e declina questa definizione in diverse categorie teoriche e

pratiche come la "soggettività algoritmica" (p. 132), l'"etnografia sintetica" (p. 87), gli "informanti sintetici" (p. 100) e altre.

Non le anticipo, il lettore le incontrerà nel testo con le relative spiegazioni. Vorrei solo menzionare quella che dalla mia prospettiva (ovviamente biased e soggettiva) emerge come la nozione centrale (per me, non necessariamente per l'autore): il concetto di macchia cieca. È stato introdotto da Heinz von Foerster mezzo secolo fa, senza nessun riferimento agli algoritmi di apprendimento autonomo o alla IAG. Faceva invece riferimento all'ottica, che ci dice che nella zona della retina dove il nervo ottico si innesta nell'occhio non ci sono fotorecettori, e quindi non percepiamo le immagini che ricadono in essa. In quella "macchia" siamo ciechi, ma questa circoscritta cecità è la condizione per poter vedere in generale, perché consente al nervo ottico di inviare gli stimoli al cervello. Per poter vedere qualsiasi cosa, quindi, serve una parziale cecità – ma di questo ci si può rendere conto. Von Foerster diceva che si può vedere di non vedere, e allora si vede meglio. Chi è consapevole della propria macchia cieca se di non poter vedere tutto, ma sa che questo non è un difetto o una lacuna nel suo campo visivo, bensì una condizione inevitabile. Quando osserva qualcosa, quindi, osserva nello stesso tempo se stesso e la sua cecità – in una "osservazione di secondo ordine" che gli consente di vedere meglio anche e proprio perché vede di non vedere tutto. In conseguenza della macchia cieca non si può osservare il mondo nella sua completezza collocandosi al di fuori di esso, perché si co-osserva anche l'osservatore che fa parte del mondo che osserva. Ma si vede molto e in modo più complesso.

Cosa c'entra con questo libro? La distinzione centrale intorno a cui (giustamente) Federico Pilati struttura il testo – tra IAG come oggetto di ricerca e come strumento di

ricerca – mi sembra riproporre la stessa condizione sul piano tecnico. Gli studiosi delle scienze sociali sono invitati a usare la IAG per studiare la IAG, riflettendo sul proprio coinvolgimento, il proprio ruolo e i propri limiti. Questa condizione può essere vista come una pratica della macchia cieca: la riflessione sulla IAG ci costringe a osservare gli strumenti che utilizziamo, i loro vincoli e i limiti delle nostre conoscenze, e nello stesso tempo ci offre la possibilità di farlo con una chiarezza e un'efficacia altrimenti irraggiungibili. Il risultato è una forma inedita di riflessività – che, con Pilati, si può chiamare sintetica.

Si vede di più? I lettori risponderanno da soli dopo aver seguito gli argomenti del libro. In ogni caso si saranno confrontati con i dilemmi e le opportunità che si presentano a una sociologia che abbandona una posizione esterna e osserva se stessa osservando i suoi strumenti. Una delle novità dirompenti della IAG è che ci porta a farlo, e questo libro ce lo mostra in modo straordinariamente informato e informativo.

Palo Alto, Novembre 2025

Introduzione

Nel panorama contemporaneo delle scienze sociali, l'emergere dell'intelligenza artificiale generativa rappresenta non solo una rivoluzione tecnologica, ma un vero e proprio punto di svolta epistemologico. Ci troviamo di fronte a strumenti che non si limitano a elaborare dati esistenti, ma che creano autonomamente contenuti testuali, visivi e sonori, sfumando i confini tra produzione umana e artificiale del sapere. Questa capacità generativa solleva interrogativi fondamentali sulle modalità di costruzione della conoscenza sociale e sugli strumenti metodologici attraverso cui comprendiamo e interpretiamo il mondo che ci circonda.

La sociologia si trova oggi a confrontarsi con queste tecnologie emergenti in una duplice prospettiva: da un lato, i sistemi di intelligenza artificiale generativa diventano essi stessi oggetto di studio, in quanto fenomeni sociotecnici che modificano le pratiche culturali, le relazioni sociali e i processi di significazione; dall'altro, questi stessi sistemi si propongono come strumenti innovativi a disposizione del ricercatore sociale, capaci di amplificare le potenzialità analitiche e interpretative dell'immaginazione sociologica. Questa integrazione non è priva di tensioni e contraddizioni, poiché mette in discussione nozioni fondamentali come l'autenticità dell'esperienza di ricerca, la neutralità dell'osservazione e il ruolo dell'interpretazione soggettiva nella costruzione del sapere scientifico.

La rivoluzione digitale ha progressivamente trasformato le scienze sociali, introducendo nuove metodologie, strumenti analitici e campi di indagine: le tecnologie

dell'informazione hanno ridefinito il modo in cui osserviamo, analizziamo e interpretiamo la realtà sociale. Questa trasformazione ha attraversato diverse fasi, ciascuna caratterizzata dall'introduzione di specifiche innovazioni tecnologiche e concettuali.

La prima fase ha visto l'emergere degli strumenti digitali come supporto alla ricerca tradizionale: spostamento dei questionari online, software per l'analisi qualitativa, database per la gestione dei dati, strumenti di visualizzazione per rappresentare reti sociali (Corbetta 2003). In questa fase, il digitale si configura essenzialmente come un'estensione dei metodi esistenti, senza mettere in discussione i fondamenti epistemologici della ricerca sociale.

La seconda fase è stata caratterizzata dalla nascita di nuovi terreni di indagine specificamente digitali: comunità virtuali, interazioni mediate dalla tecnologia, identità online (Rheingold 1993). L'etnografia virtuale o netnografia si è affermata come metodologia specifica per lo studio di questi contesti, adattando principi e tecniche dell'osservazione partecipante alle peculiarità degli spazi digitali (Kozinets 2006).

La terza fase ha visto l'introduzione dei big data e delle tecniche computazionali come nuovo paradigma per le scienze sociali, con l'emergere di approcci quantitativi su larga scala (Salganik 2019) e metodologie nativamente digitali per l'analisi delle tracce online (Rogers 2009). Questa svolta ha sollevato interrogativi sul rapporto tra dati e teoria, tra descrizione e interpretazione, tra correlazioni statistiche e comprensione dei significati sociali (Marres 2017).

Oggi ci troviamo di fronte a una quarta fase, caratterizzata dall'emergere dell'intelligenza artificiale generativa, che non si limita ad analizzare dati esistenti ma produce autonomamente contenuti, simulazioni, interpretazioni. Questa svolta generativa rappresenta un salto qualitativo rispetto alle fasi precedenti, poiché introduce nella ricer-

ca sociale non solo nuovi strumenti analitici, ma vere e proprie entità capaci di produrre conoscenza in modo semi-autonomo.

Negli ultimi anni, l'intelligenza artificiale generativa ha compiuto progressi straordinari, passando da semplici applicazioni sperimentali a sistemi complessi in grado di produrre testi, immagini, video e audio di qualità sorprendentemente elevata. Modelli come GPT (Generative Pre-trained Transformer), Claude, Midjourney, Stable Diffusion e altri hanno dimostrato capacità generative che superano le aspettative anche degli esperti del settore, avvicinandosi in molti casi a produzioni indistinguibili da quelle umane.

Questi sviluppi sono il risultato di avanzamenti in diversi ambiti dell'intelligenza artificiale, in particolare nell'apprendimento profondo (deep learning) e nell'elaborazione del linguaggio naturale (natural language processing). I modelli generativi contemporanei si basano su architetture neurali complesse, addestrate su enormi quantità di dati testuali e multimediali, che consentono loro di apprendere strutture linguistiche, schemi comunicativi e convenzioni culturali.

Ma cosa distingue l'intelligenza artificiale generativa dalle precedenti forme di automazione e computazione? La differenza fondamentale risiede nella capacità di questi sistemi di produrre contenuti originali, non limitandosi a classificare, analizzare o trasformare dati esistenti. Un modello generativo non si limita a rispondere a domande predefinite o a eseguire compiti specifici, ma può creare testi argomentativi, narrazioni, descrizioni, dialoghi, immagini e altri artefatti culturali con un grado di autonomia e creatività prima impensabile per una macchina.

Questa capacità generativa ha rapide e profonde implicazioni per numerosi ambiti sociali e culturali: dalla produzione artistica e letteraria al giornalismo, dall'educazione

alla ricerca scientifica, dalla comunicazione pubblicitaria alla produzione di contenuti per il web. In ciascuno di questi settori, l'introduzione di strumenti generativi sta ridefinendo ruoli professionali, pratiche creative, modalità di produzione e distribuzione dei contenuti, sollevando interrogativi sul futuro del lavoro intellettuale e creativo nell'era dell'automazione cognitiva.

Nel contesto specifico delle scienze sociali, l'emergere dell'intelligenza artificiale generativa pone questioni ancora più complesse e affascinanti, poiché questi strumenti non sono semplicemente oggetti di studio o ausili tecnici, ma possono configurarsi come veri e propri attori sociali, capaci di interagire, influenzare percezioni, produrre significati e partecipare attivamente alla costruzione della realtà sociale che pretendono di analizzare.

La ricerca sociale si trova dunque oggi a confrontarsi con l'emergere dell'intelligenza artificiale generativa in modi che richiedono un profondo ripensamento dei suoi presupposti epistemologici e delle sue pratiche di ricerca (Pilati, Munk e Venturini 2024).

Tradizionalmente, la sociologia classica si basa su alcuni principi fondamentali: l'immersione del ricercatore nel contesto studiato, l'osservazione diretta delle pratiche sociali, l'interazione faccia a faccia con i soggetti della ricerca, la raccolta di dati attraverso note di campo, interviste e documentazione multimediale, l'interpretazione riflessiva dei significati culturali. Questi principi presuppongono una netta distinzione tra il soggetto conoscente (il ricercatore) e l'oggetto della conoscenza (il fenomeno sociale studiato), nonché una chiara separazione tra strumenti metodologici e dati empirici.

L'introduzione dell'intelligenza artificiale generativa complica notevolmente questo quadro. In primo luogo, i sistemi di IA generativa diventano essi stessi oggetti di studio, in quanto fenomeni sociotecnici che influenzano

pratiche, relazioni e significati culturali. Il ricercatore che studia l'impatto sociale dell'IA generativa si trova ad osservare non solo come gli esseri umani interagiscono con questi sistemi, ma anche come questi sistemi partecipano attivamente alla costruzione della realtà sociale, generando contenuti, influenzando percezioni, modificando pratiche comunicative e relazionali.

In secondo luogo, questi stessi sistemi possono essere utilizzati come strumenti di ricerca, supportando lo studioso nelle diverse fasi del processo di indagine: dalla raccolta dei dati alla loro analisi, dall'interpretazione alla comunicazione dei risultati. Un modello generativo può trascrivere automaticamente interviste, analizzare tesi identificando temi ricorrenti, suggerire connessioni teoriche generando interpretazioni alternative, e via dicendo.

In terzo luogo, l'interazione tra utente e intelligenza artificiale può dar vita a forme ibride di conoscenza, in cui l'interpretazione umana e quella algoritmica si intrecciano in modi complessi e non sempre trasparenti. Questa ibridazione solleva interrogativi fondamentali sulla natura della ricerca tramite intelligenza artificiale, sull'autenticità dell'esperienza, sulla soggettività dell'interpretazione, fino alla costruzione del significato.

Uno degli aspetti più innovativi e al contempo problematici dell'introduzione delle tecnologie generative riguarda proprio la riflessività del ricercatore, ovvero la capacità di riconoscere e analizzare criticamente il proprio ruolo nel processo di ricerca, i propri presupposti teorici, i propri bias cognitivi e culturali, la propria posizione sociale (Bourdieu e Wacquant 1992).

La riflessività è considerata un elemento fondamentale della ricerca sociale contemporanea, che riconosce l'impossibilità di un'osservazione neutrale e oggettiva e valorizza invece la consapevolezza critica del proprio posizionamento come parte integrante del processo conoscitivo.

Il ricercatore riflessivo non pretende di eliminare la propria soggettività, ma di renderla esplicita e di integrarla consapevolmente nell'analisi.

L'introduzione dell'intelligenza artificiale generativa come strumento di ricerca complica notevolmente questa dimensione riflessiva, introducendo nuove forme di mediazione tra il ricercatore e il suo oggetto di studio. I sistemi di IA non sono strumenti neutri, ma incorporano visioni del mondo, presupposti culturali, bias di vario tipo derivanti dai dati su cui sono stati addestrati e dalle scelte progettuali che li hanno informati (Munk 2023).

Utilizzare questi sistemi nel processo di ricerca significa quindi introdurre una soggettività algoritmica che si intreccia con quella umana in modi non sempre trasparenti e controllabili. Come distinguere, in un'interpretazione generata da un sistema di IA, ciò che deriva dai dati empirici, ciò che riflette bias algoritmici, ciò che emerge dalla interazione con il ricercatore umano?

Anziché ignorare questa complessità o considerarla un limite invalicabile qui si propone di trasformarla in una risorsa riflessiva. La riflessività sintetica consiste nella capacità di riconoscere e analizzare criticamente non solo la propria soggettività di ricercatori umani, ma anche quella degli strumenti algoritmici utilizzati, nonché le modalità di interazione e co-costruzione tra queste diverse forme di soggettività.

Questo approccio riflessivo può concretizzarsi in diverse pratiche di ricerca:

1. Esplicitazione algoritmica: rendere visibili e comprensibili gli strumenti di IA utilizzati, le loro caratteristiche tecniche, i dati su cui sono stati addestrati, i limiti conosciuti.
2. Comparazione interpretativa: confrontare sistematicamente interpretazioni umane e algoritmiche de-

gli stessi dati, analizzando convergenze, divergenze, complementarità.
3. Archeologia dei bias: esplorare criticamente i bias incorporati sia nella soggettività del ricercatore, sia negli algoritmi utilizzati, sia nella loro interazione.
4. Triangolazione riflessiva: utilizzare diversi strumenti di IA, con diverse architetture e training set, per generare interpretazioni multiple e metterle in relazione con quelle del ricercatore.
5. Metanalisi generativa: chiedere agli stessi sistemi di IA di riflettere criticamente sulle proprie interpretazioni, esplicitando limiti, incertezze, possibili bias.

La riflessività sintetica non pretende di risolvere completamente le tensioni e le contraddizioni dell'uso dell'IA, ma prova a renderle esplicite e integrarle consapevolmente nel processo di ricerca, trasformando potenziali limiti in risorse analitiche e interpretative.

L'IA generativa presenta infatti una peculiare duplicità epistemologica nel contesto della ricerca sociale: si configura contemporaneamente come oggetto di indagine e come potenziale soggetto conoscente, come fenomeno da studiare e come strumento per studiare altri fenomeni (Esposito 2022). Questa duplicità solleva questioni fondamentali relative ai presupposti epistemologici della ricerca sociale, alle relazioni tra soggetto e oggetto della conoscenza, ai criteri di validazione del sapere scientifico. In questi contesti ibridi, l'IA non è semplicemente un oggetto da osservare, ma una controparte che partecipa attivamente alla co-creazione del campo. I modelli linguistici di grandi dimensioni, in particolare, sfidano le nostre categorizzazioni abituali. Essi sono simultaneamente:

- Artefatti tecnici prodotti da sistemi socio-economici specifici.
- Attori che esercitano una forma di agency nella pro-

duzione di testi e significati.
- Assemblaggi di molteplici voci umane incorporate attraverso i dati di addestramento.
- Interfacce attraverso cui emergono nuove forme di relazionalità.

Questa natura ibrida richiede un ripensamento delle metodologie tradizionali. Se nella ricerca sociale classica la distinzione tra lo studioso (soggetto) e la cultura studiata (oggetto) era già problematizzata, tramite l'IA questa distinzione diventa ancora più fluida e complessa (Pilati, Munk e Venturini 2024).

Un ulteriore spostamento concettuale fondamentale riguarda il passaggio da una preoccupazione esclusiva per le rappresentazioni a un'attenzione per le performatività. I sistemi di IA generativa non si limitano a rappresentare il mondo sociale, ma partecipano attivamente alla sua configurazione attraverso atti performativi che hanno conseguenze materiali.

Questo slittamento richiama il concetto di performatività in una cornice post-umanista. I fenomeni non preesistono semplicemente alle loro misurazioni, ma emergono attraverso processi relazionali di intra-azione. Ciò implica prestare attenzione a:

1. La performatività algoritmica: Come i sistemi di IA generativa non si limitano a descrivere il mondo, ma contribuiscono a configurarlo attraverso previsioni che diventano prescrizioni, classificazioni che producono effetti reali, e generazioni che plasmano immaginari collettivi.
2. La materialità dei processi semiotici: Come i significati prodotti attraverso l'interazione con sistemi di IA sono inseparabili dalle infrastrutture materiali che li rendono possibili (data center, reti di fibra ottica, dispositivi, consumi energetici).

3. L'emergenza di nuove soggettività: Come l'interazione quotidiana con sistemi di IA contribuisce alla formazione di nuove forme di soggettività, alterando i processi di costruzione identitaria e le pratiche di relazionalità.

La sfida metodologica che ne consegue è quella di sviluppare strumenti capaci di cogliere questi processi performativi nei loro aspetti sia discorsivi che materiali, evitando sia il determinismo tecnologico che il costruzionismo sociale ingenuo.

Concludendo questo primo capitolo emerge con chiarezza la necessità di sviluppare metodologie ibride che attraversino confini disciplinari e ontologici tradizionalmente separati. Queste metodologie possono attingere a diverse tradizioni:

- Le riflessioni comunicative sulla mediazione tecnologica.
- Gli studi sociali della scienza e della tecnologia (STS) e la loro attenzione alle pratiche socio-materiali.
- La teoria critica dei media e la sua analisi delle logiche infrastrutturali.
- L'antropologia post-umanista e la sua esplorazione di forme di agency non-umane.
- Gli approcci decoloniali e la loro critica delle gerarchie epistemiche globali.

L'ibridazione metodologica non è semplicemente una questione di combinare tecniche diverse, ma implica un ripensamento più profondo delle premesse ontologiche ed epistemologiche che guidano la ricerca nell'era dell'IA generativa.

La sfida che ci attende è quella di sviluppare una ricerca sociale capace di navigare la complessità dei mondi ibridi in cui viviamo, rendendo visibili le strutture di potere che li attraversano e contribuendo alla costruzione di relazioni più eque e generative tra umani e intelligenze artificiali.

Infine, l'uso dell'IA nella ricerca sociale, come ogni innovazione, solleva una serie di sfide etiche che richiedono attenta considerazione:

1. Privacy e consenso: l'utilizzo di IA generativa per al'analisi dati solleva questioni sulla privacy dei partecipanti e sulla natura del consenso informato. I partecipanti alla ricerca consentono l'analisi dei loro dati da parte di algoritmi? Comprendono le implicazioni di questa analisi automatizzata?
2. Proprietà intellettuale: chi detiene i diritti sui contenuti generati dall'IA? Come attribuire correttamente la paternità di interpretazioni co-costruite tra ricercatore e algoritmo?
3. Trasparenza algoritmica: come garantire che i processi algoritmici utilizzati nella ricerca siano sufficientemente trasparenti e comprensibili, sia per i ricercatori stessi, sia per i partecipanti, sia per la comunità scientifica?
4. Bias e discriminazione: come prevenire che bias algoritmici si traducano in interpretazioni discriminatorie o rappresentazioni stereotipate di gruppi sociali già marginali o vulnerabili?
5. Responsabilità e accountability: chi è responsabile delle conseguenze di interpretazioni o rappresentazioni generate algoritmicamente? Come articolare forme appropriate di accountability in contesti di ricerca ibridi uomo-macchina?

Queste sfide non rappresentano ostacoli insormontabili, ma questioni aperte che richiedono riflessione critica, dibattito scientifico, sviluppo di linee guida, e sperimentazione responsabile. Solo in questo modo, tramite una riflessione interna alla comunità dei ricercatori sociali, sarà possibile gestire l'emergere di nuovi paradigmi.

A partire da questa esigenza, il presente volume esplora i molteplici ruoli ed usi dell'Intelligenza Artificiale Generativa attraverso un percorso che intreccia riflessio-

ne teorica, analisi metodologica e casi di studio empirici. La struttura del libro riflette la duplice prospettiva che guida la nostra indagine: l'intelligenza artificiale generativa come oggetto di studio e come strumento innovativo per la ricerca sociale.

Il primo capitolo offre una panoramica storica e concettuale dello sviluppo dei modelli generativi, dalle prime applicazioni fino agli sviluppi più recenti.

Nel secondo capitolo, "L'intelligenza artificiale generativa come oggetto di studio", vengono esplorate le implicazioni sociali e culturali di questi sistemi, con particolare attenzione al loro ruolo nella produzione di significato e rappresentazioni culturali.

Il terzo capitolo, "L'intelligenza artificiale generativa come strumento di ricerca", esamina le potenzialità dell'IA come ausilio per la ricerca sociale. Vengono analizzati strumenti specifici per la trascrizione, il tagging automatico e l'indicizzazione, con attenzione sull'integrazione tra computazione e riflessività.

Il quarto capitolo, "Pratiche sperimentali ed esperimenti pratici", presenta casi concreti di applicazione dell'IA in diversi contesti, con un'attenzione specifica sulle strategie per sviluppare una riflessività sintetica che utilizzi l'IA come strumento per riflettere su bias e posizionamenti sociali.

Lungi dal proporsi come un manuale definitivo o un compendio esaustivo, questo libro si configura come un contributo al dibattito in corso. L'auspicio è che possa stimolare ulteriori riflessioni, sperimentazioni e innovazioni, contribuendo a definire i contorni di una rinvigorita sociologia digitale capace di confrontarsi criticamente e creativamente con le sfide e le opportunità dell'intelligenza artificiale.

Capitolo 1. Origini e sviluppo dell'Intelligenza Artificiale Generativa

L'intelligenza artificiale generativa rappresenta uno degli sviluppi più significativi e trasformativi nel panorama tecnologico contemporaneo. Per comprenderne appieno le implicazioni sociali e culturali, è necessario ripercorrere il percorso evolutivo che ha portato da concetti teorici e primi esperimenti a sistemi complessi capaci di generare contenuti indistinguibili da quelli prodotti dall'intelletto umano. Questa genealogia non è soltanto tecnica, ma profondamente intrecciata con evoluzione di idee, visioni culturali, contesti sociali e rapporti di potere che hanno influenzato lo sviluppo di questi sistemi.

Le radici concettuali dell'intelligenza artificiale generativa possono essere rintracciate nelle teorie probabilistiche dell'informazione sviluppate a metà del XX secolo. I lavori seminali degli anni '50 di Claude Shannon sulla teoria dell'informazione e le ricerche di Alan Turing sull'intelligenza artificiale hanno posto le basi teoriche per lo sviluppo di sistemi capaci non solo di elaborare informazioni, ma anche di generare nuovi contenuti. Tuttavia, è solo con l'avvento dei modelli probabilistici più sofisticati, come le catene di Markov, che si iniziano a vedere i primi tentativi concreti di generazione automatica di contenuti, principalmente nel campo della linguistica computazionale.

Negli anni '80 e '90, l'emergere delle reti neurali artificiali ha aperto nuove possibilità per l'IA generativa. In particolare, i modelli connessionisti, ispirati al funzionamento del cervello umano, hanno mostrato capacità di apprendi-

mento e generalizzazione superiori rispetto ai precedenti approcci simbolici. Le reti neurali ricorrenti (RNN) hanno consentito di modellare sequenze temporali, come testi o musica, mentre le reti neurali convoluzionali (CNN) hanno rivoluzionato l'elaborazione di immagini.

Un salto qualitativo fondamentale si è verificato nel 2014 con l'introduzione delle Generative Adversarial Networks (GAN) da parte di Ian Goodfellow e colleghi. Il concetto innovativo alla base delle GAN è la competizione tra due reti neurali: un generatore, che crea contenuti cercando di imitare esempi reali, e un discriminatore, che tenta di distinguere i contenuti generati da quelli autentici. Questa competizione "avversaria" porta a un miglioramento progressivo della qualità dei contenuti generati, in un processo che è stato paragonato a quello di un artista falsario che affina continuamente la propria tecnica per ingannare un esperto d'arte.

Le GAN hanno segnato un punto di svolta nella generazione di immagini realistiche, consentendo la creazione di volti umani, paesaggi, opere d'arte e altri contenuti visivi di qualità sorprendente. Modelli come StyleGAN, BigGAN e Progressive GAN hanno progressivamente migliorato la risoluzione, la diversità e il realismo delle immagini generate, aprendo la strada a nuove applicazioni in ambito artistico, del design, dell'intrattenimento e della comunicazione visiva.

Parallelamente all'evoluzione dei modelli generativi per le immagini, si è assistito a un progresso significativo nei modelli linguistici, culminato nell'era dei Large Language Models (LLM). La svolta decisiva è avvenuta con l'introduzione dell'architettura Transformer nel 2017 da parte di ricercatori di Google. A differenza delle precedenti reti neurali ricorrenti, l'architettura Transformer utilizza meccanismi di attenzione che consentono di elaborare in parallelo intere sequenze di testo, superando i limiti delle

RNN nell'elaborazione di contesti ampi e relazioni a lunga distanza nel testo.

Basandosi sull'architettura Transformer, nel 2018 è stato introdotto BERT (Bidirectional Encoder Representations from Transformers), che ha rivoluzionato la comprensione del linguaggio naturale grazie alla capacità di considerare il contesto bidirezionale di una parola. Nello stesso periodo, OpenAI ha sviluppato la serie GPT (Generative Pre-trained Transformer), focalizzata specificamente sulla generazione di testo.

Con il rilascio di GPT-3 nel 2020, si è assistito a un salto di scala senza precedenti: con 175 miliardi di parametri, questo modello ha dimostrato capacità generative stupefacenti, in grado di produrre testi coerenti, articolati e stilisticamente diversificati su praticamente qualsiasi argomento. GPT-3 e i successivi modelli della serie non sono stati addestrati per compiti specifici, ma hanno acquisito una comprensione generale del linguaggio attraverso l'apprendimento non supervisionato su vasti corpus testuali disponibili su internet.

Questo approccio di "pre-addestramento generale seguito da fine-tuning specifico" è diventato il paradigma dominante nello sviluppo di modelli linguistici generativi. Modelli come PaLM di Google, Claude di Anthropic, LLaMA di Meta e i successivi sviluppi della serie GPT hanno progressivamente migliorato in termini di dimensioni, capacità di ragionamento, comprensione contestuale e generazione di contenuti.

Un aspetto particolarmente rilevante di questi sviluppi è stata l'introduzione di tecniche di allineamento con valori e preferenze umane, come il Reinforcement Learning from Human Feedback (RLHF). Queste tecniche mirano a rendere i modelli non solo più capaci, ma anche più sicuri, etici e utili, cercando di allineare il loro comportamento con le aspettative e i valori umani.

L'evoluzione più recente dell'IA generativa è caratterizzata dalla convergenza di diverse modalità (testo, immagine, audio, video) in sistemi multimodali integrati. Modelli come DALL-E, Midjourney e Stable Diffusion hanno innovato la generazione di immagini a partire da descrizioni testuali, creando un ponte tra la comprensione linguistica e la sintesi visiva. Questi sistemi di "text-to-image" utilizzano varianti dell'architettura Transformer o architetture di diffusione per tradurre concetti linguistici in rappresentazioni visive dettagliate e creative.

Nel campo audio, sistemi come WaveNet, Jukebox e MusicLM hanno consentito la generazione di voci umane sintetiche, musica e suoni ambientali di qualità sorprendentemente realistica. Parallelamente, modelli come Whisper hanno rivoluzionato il riconoscimento vocale e la trascrizione automatica, creando un ulteriore ponte tra modalità audio e testuale.

La frontiera più recente è rappresentata dalla generazione video, con sistemi come Sora, Runway ML e numerosi altri modelli che stanno dimostrando capacità sempre più sofisticate nel creare sequenze video realistiche a partire da prompt testuali o immagini statiche.

A questa frontiera si affianca la stessa integrazione di queste diverse capacità in sistemi unificati, capaci di comprendere e generare contenuti multimodali in modo coerente e contestualizzato. GPT-4 di OpenAI, Claude di Anthropic, Gemini di Google e Llama di Meta rappresentano esempi di questa tendenza verso modelli generali multimodali, che possono elaborare input in diverse modalità e generare output altrettanto diversificati. Inoltre, la questione di come gli stessi modelli di IA generativa possano compiere azioni sul web per conto dell'utente, ovvero come agenti, è in continua evoluzione.

Un fenomeno particolarmente rilevante degli ultimi anni è inoltre la crescente democratizzazione computa-

zionale dell'IA generativa, con l'emergere di modelli open source e l'accessibilità di strumenti generativi a un pubblico sempre più ampio. Progetti come Llama 2, Mistral, Falcon e Stable Diffusion hanno reso disponibili modelli potenti che possono essere eseguiti anche su hardware consumer, consentendo a sviluppatori indipendenti, ricercatori, artisti e utenti comuni di sperimentare con l'IA generativa.

Questo "rinascimento open source" ha stimolato un'ondata di innovazione distribuita, con comunità di sviluppatori che adattano, migliorano e specializzano i modelli base per applicazioni specifiche. Piattaforme come Hugging Face hanno facilitato la condivisione di modelli e il loro utilizzo attraverso interfacce semplificate, mentre framework come LangChain e LlamaIndex hanno reso più accessibile l'integrazione di LLM nelle applicazioni.

È importante sottolineare che questa narrativa evolutiva dell'IA generativa, pur essendo descrittivamente accurata, rischia di veicolare una visione deterministica e teleologica dello sviluppo tecnologico. La traiettoria che abbiamo delineato non era predeterminata né inevitabile, ma il risultato di scelte specifiche, priorità di ricerca, investimenti economici, contesti culturali e quadri regolatori (Bender et al. 2021).

Lo sviluppo dell'IA generativa è stato fortemente influenzato da fattori socioeconomici come la disponibilità di enormi quantità di dati (spesso raccolti senza esplicito consenso), l'accesso a potenza di calcolo crescente, gli investimenti massicci da parte di grandi corporation tecnologiche, specifiche visioni culturali predominanti nelle comunità di ricerca sull'IA, e la mancanza di regolamentazione in molti ambiti (Cardon 2015).

Una prospettiva sociologica critica deve riconoscere che questi sistemi non sono emersi in un vuoto sociale, ma sono profondamente radicati in relazioni di potere, visio-

ni culturali e strutture economiche esistenti (Crawford 2021). La loro evoluzione non è guidata solo da miglioramenti tecnici, ma anche da obiettivi commerciali, considerazioni geopolitiche, valori culturali dominanti e concezioni specifiche dell'intelligenza e della creatività.

Capitolo 2. L'Intelligenza Artificiale Generativa come oggetto di studio

Considerare l'intelligenza artificiale generativa come oggetto di studio significa osservare e interpretare come questi sistemi vengono sviluppati, implementati, utilizzati e integrati nei contesti sociali e culturali. L'indagine dei sistemi di IA generativa può concentrarsi su diversi livelli di analisi:

1. Il livello della produzione: studiare i contesti sociali, culturali ed economici in cui i sistemi di IA generativa vengono sviluppati, le comunità di pratica dei programmatori e ricercatori, le visioni e i valori che guidano la progettazione algoritmica, le scelte tecniche e le loro implicazioni sociali.
2. Il livello dell'implementazione: analizzare come i sistemi generativi vengono introdotti e adattati in specifici contesti organizzativi e istituzionali, le trasformazioni delle pratiche lavorative, le resistenze e le negoziazioni, i processi di appropriazione e risignificazione.
3. Il livello dell'uso: osservare come diversi attori sociali interagiscono con i sistemi generativi, le modalità di utilizzo previste e impreviste, le strategie di adattamento e sovversione, le nuove pratiche sociali che emergono dall'interazione uomo-macchina.
4. Il livello del significato: interpretare i simboli associati all'intelligenza artificiale generativa, le rappresentazioni sociali, i discorsi pubblici, le narrazioni mediatiche, i frame interpretativi attraverso cui questi sistemi vengono compresi e significati.

5. Il livello dell'impatto: analizzare le conseguenze sociali, culturali, economiche e politiche dell'introduzione di sistemi generativi, i cambiamenti nelle relazioni di potere, nelle dinamiche di inclusione ed esclusione, nei processi di produzione e distribuzione del sapere.

Questo capitolo esplora l'IA generativa come oggetto di indagine sociologica in sé, esaminando le sue implicazioni sociali, culturali ed epistemologiche. La rapida evoluzione di queste tecnologie sta producendo trasformazioni profonde nelle dinamiche sociali, nelle pratiche culturali e nei processi di costruzione della conoscenza, offrendo un terreno fertile per l'analisi sociologica (Joyce e Cruz 2024). Nel fare ciò cercheremo di superare visioni deterministiche e riduzionistiche dell'intelligenza artificiale, riconoscendone la natura socialmente costruita e culturalmente situata. Non si tratta di studiare gli algoritmi come entità astratte e autonome, ma di osservare come essi si inseriscono in contesti sociali specifici, interagiscono con attori umani, incorporano e riproducono visioni del mondo, valori, pregiudizi e relazioni di potere.

Allo stesso tempo, la sociologia dei sistemi di intelligenza artificiale deve confrontarsi con sfide peculiari: come osservare processi algoritmici spesso opachi e inaccessibili? Come documentare interazioni mediate dalla tecnologia? Come interpretare output generati automaticamente? Come distinguere tra intenzionalità umana e comportamento algoritmico? Queste sfide richiedono lo sviluppo di nuovi strumenti concettuali, capaci di cogliere la complessità sociotecnica dei fenomeni studiati.

La comprensione dell'IA generativa richiede un approccio che vada oltre gli aspetti puramente tecnici, per interrogare le complesse intersezioni tra algoritmi, dati, potere e società. Queste tecnologie non emergono in un vacuum, ma sono radicate in specifici contesti socioculturali, in-

corporano visioni del mondo particolari e riflettono le strutture di potere esistenti. Allo stesso tempo, esse contribuiscono attivamente a configurare tali contesti, visioni e strutture in modi che richiedono un'attenta analisi critica (Airoldi 2021).

In questo capitolo, esamineremo le principali correnti teoriche che hanno affrontato l'IA generativa come fenomeno sociale, le strategie più adatte per studiarne gli effetti, e i dibattiti contemporanei sulle sue implicazioni etiche e politiche. Questa prospettiva ci consente di superare una visione strumentale e apparentemente neutrale dell'IA, rivelando invece il complesso ecosistema di relazioni, pratiche e strutture di potere in cui è immersa. La ricerca sociale, infatti, come pratica di indagine, non può prescindere da un'analisi delle condizioni materiali e dei rapporti sociali che rendono possibile l'esistenza stessa dei sistemi di IA che stiamo studiando. Analizzeremo inoltre come queste tecnologie stiano trasformando ambiti diversi come il lavoro, l'identità, la creatività e la produzione culturale, evidenziando tensioni e contraddizioni che emergono dalla loro diffusione. L'obiettivo è fornire una base concettuale solida per comprendere l'IA generativa non solo come innovazione tecnologica, ma come fenomeno socialmente costruito e socialmente trasformativo che richiede nuovi strumenti analitici e interpretativi.

Il lavoro invisibile: l'economia del micro-lavoro

Dietro l'apparente autonomia dei sistemi di intelligenza artificiale si nasconde una vasta rete di lavoratori umani che ne rendono possibile il funzionamento. L'antropologa Mary Gray, insieme a Siddharth Suri, ha coniato il termine "ghost workers" per descrivere questa forza lavoro globale spesso invisibile e sottopagata che costituisce l'infrastruttura umana dell'IA. Il processo di addestramento

dei modelli richiede enormi quantità di dati etichettati, e dietro questa etichettatura si celano milioni di lavoratori, principalmente localizzati nel Sud Globale, che svolgono micro-compiti ripetitivi e talvolta psicologicamente gravosi (Crawford 2021).

Questi lavoratori digitali etichettano immagini per il riconoscimento visivo, moderano contenuti potenzialmente traumatici, trascrivono audio e verificano i risultati generati dai modelli. Le loro condizioni lavorative sono caratterizzate da precarietà, compensi minimi e assenza di protezioni sociali adeguate. Nel corridoio di un'azienda tecnologica di San Francisco, un ingegnere può parlare con entusiasmo dell'ultimo modello sviluppato, mentre a migliaia di chilometri di distanza, in un centro di etichettatura dati di Nairobi o Manila, decine di lavoratori visualizzano immagini violente per insegnare ai sistemi di moderazione a riconoscere contenuti problematici.

Questa geografia del lavoro digitale riflette e perpetua disuguaglianze globali preesistenti. Come osserva Casilli (2025), l'innovazione tecnologica occidentale dipende da forme di lavoro invisibilizzate che vengono strategicamente nascoste per mantenere intatta la narrazione dell'autonomia tecnologica. È fondamentale riconoscere che i modelli che utilizziamo sono il prodotto di queste relazioni di lavoro asimmetriche. Il sociologo che utilizza o studia sistemi di IA deve quindi interrogarsi su come le condizioni di questo lavoro invisibile influenzino la qualità dei dati e, di conseguenza, i comportamenti dei modelli che producono i testi che analizziamo.

Bias nei training set: la politica dei dati e le epistemologie incorporate

I modelli di IA non sono entità neutre che elaborano informazioni in modo oggettivo, ma incorporano le visioni del mondo, i pregiudizi e le strutture di potere presenti nei dati con cui vengono addestrati. Questi bias non rappresentano semplici "errori" tecnici da correggere, ma rivelano le fondamenta epistemologiche e politiche su cui l'IA è costruita. Come sostiene Kate Crawford (2021), i sistemi di IA sono registri politici che codificano relazioni di potere attraverso la loro architettura e i dati che utilizzano.

Il bias di rappresentazione si manifesta nella tendenza dei dataset a sovrarappresentare determinate popolazioni – tipicamente uomini bianchi delle nazioni occidentali – e a sottorappresentare o rappresentare in modo stereotipato altri gruppi (Esposito 2022). Quando un ricercatore utilizza un modello linguistico per analizzare narrazioni su temi come il genere, la razza o la disabilità, deve essere consapevole che le risposte che ottiene sono filtrate attraverso queste lenti distorte. I bias storici si riflettono nei dati raccolti che inevitabilmente incorporano le disuguaglianze e le ingiustizie sociali del passato e del presente. Un corpus testuale che include decenni di letteratura medica, ad esempio, porterà con sé le tracce di paradigmi medici storicamente androcentrici che hanno considerato il corpo maschile come norma e quello femminile come deviazione.

Le scelte su cosa misurare, come misurarlo e quali dati considerare rilevanti rappresentano decisioni politiche che riflettono specifiche priorità e visioni del mondo. Quando decidiamo di utilizzare determinate metriche per valutare la "performance" di un modello linguistico, stiamo implicitamente definendo cosa conta come successo e cosa viene invece considerato irrilevante (O'Neil 2017). I processi di selezione e aggregazione dei dati tendo-

no inoltre a rendere invisibili le esperienze marginali e a privilegiare le narrazioni dominanti, creando sistemi che funzionano meglio per alcuni utenti rispetto ad altri.

Ciò implica la necessità di un'analisi critica dei dataset utilizzati per addestrare i modelli con cui interagiamo. Bisogna interrogarsi costantemente su quali voci e prospettive siano incluse o escluse, chi decida quali dati sono considerati validi o rilevanti, e come le epistemologie incorporate nei dati influenzino le interazioni con i sistemi di IA. Solo attraverso questa consapevolezza critica è possibile contestualizzare adeguatamente le osservazioni e le interpretazioni che ne derivano.

Sorveglianza e oligopolio delle piattaforme

I sistemi di IA sono sempre più integrati nelle infrastrutture di sorveglianza e controllo sociale, creando nuove forme di potere disciplinare e biopolitico che Michel Foucault avrebbe riconosciuto, seppur in forme tecnologicamente evolute (Zuboff 2019). Questa dimensione è in realtà fondamentale per comprendere i contesti sociali in cui l'IA opera e viene percepita.

La sorveglianza predittiva, che utilizza algoritmi per anticipare comportamenti considerati "criminali" o "devianti", tende a replicare e rafforzare pregiudizi esistenti contro comunità già marginalizzate (Angwin et al 2022). In diverse città americane, i sistemi di polizia predittiva hanno concentrato la sorveglianza in quartieri a prevalenza afroamericana o latina, creando loop di feedback che giustificano ulteriore sorveglianza. Questi sistemi non operano in un vuoto sociale, ma si inseriscono in contesti storicamente segnati da discriminazione sistematica (Zuboff 2019).

Le tecnologie di riconoscimento biometrico – facciale, del portamento o della voce – sollevano questioni crucia-

li sul consenso, sulla privacy e sul diritto all'anonimato. L'espansione di queste tecnologie crea quello che potremmo definire un panopticon digitale che disciplina i comportamenti attraverso la consapevolezza di essere costantemente osservati. Un individuo che cammina in una città densamente monitorata da telecamere equipaggiate con sistemi di riconoscimento facciale modifica inconsciamente il proprio comportamento, sapendo di essere identificabile e tracciabile (Lyon 2024).

I sistemi di valutazione algoritmica, dal credito sociale cinese alle valutazioni sul posto di lavoro utilizzate in molte aziende occidentali, diventano strumenti di classificazione e stratificazione degli individui, spesso operando attraverso logiche opache e difficilmente contestabili. Un lavoratore della gig economy, costantemente valutato attraverso metriche automatizzate, sviluppa forme di auto-disciplina e adattamento strategico a questi sistemi che influenzano profondamente la sua esperienza lavorativa quotidiana (Bonini e Treré 2024).

Per chi studia questi sistemi, emergono domande cruciali su come queste tecnologie modifichino le relazioni di potere tra stato, corporazioni e cittadini; in che modo la consapevolezza della sorveglianza alteri i comportamenti e le pratiche quotidiane; e come si possano sviluppare metodologie che rendano visibili questi meccanismi di controllo. In questo contesto il ricercatore deve essere consapevole di come i sistemi di IA con cui interagisce siano parte di più ampi regimi di sorveglianza e controllo.

Lo sviluppo dell'intelligenza artificiale è inoltre caratterizzato da una crescente concentrazione di potere nelle mani di poche corporazioni tecnologiche globali (Srnicek 2017). Questa concentrazione ha profonde implicazioni economiche, politiche ed epistemologiche. I costi computazionali per l'addestramento di modelli avanzati sono talmente elevati da essere accessibili solo alle grandi

aziende tecnologiche o ai laboratori da esse finanziati, creando barriere all'ingresso quasi insormontabili per attori più piccoli o con minori risorse.

La monopolizzazione delle risorse computazionali si accompagna a una privatizzazione dei dati sempre più pervasiva. Le piattaforme digitali accumulano enormi quantità di dati generati dagli utenti, trasformandoli in proprietà privata utilizzata per addestrare modelli proprietari. Il paradosso è che i contenuti culturali prodotti collettivamente dall'umanità – libri, articoli, conversazioni, immagini – diventano la materia prima per sistemi di IA che vengono poi commercializzati e monetizzati privatamente (Couldry e Mejias 2019).

Si assiste inoltre a una centralizzazione del sapere, con la conoscenza tecnica necessaria per lo sviluppo dell'IA che si concentra in pochi hub globali, principalmente in Nord America e Cina, creando nuove geografie della conoscenza e dell'innovazione. Questa centralizzazione non è solo geografica ma anche epistemica: determinate visioni del mondo, priorità di ricerca e paradigmi tecnologici diventano dominanti, mentre approcci alternativi vengono marginalizzati.

La dipendenza infrastrutturale che ne deriva vede governi, università, organizzazioni non-profit e piccole imprese diventare sempre più dipendenti dalle infrastrutture di IA fornite dai giganti tecnologici. Un ricercatore che utilizza modelli linguistici per analizzare testi culturali dipende inevitabilmente dalle infrastrutture proprietarie sviluppate da queste aziende, con tutte le limitazioni e i condizionamenti che questo comporta (Burkhardt e Rieder 2024).

Questi processi di oligopolizzazione sollevano questioni fondamentali: come la concentrazione di potere influenza le direzioni di sviluppo dell'IA e i valori incorporati nei sistemi; quali forme di resistenza, appropriazione o uso

alternativo dei sistemi di IA emergono in risposta a questa centralizzazione; e come si possa sviluppare un'analisi che non dipenda essa stessa dalle infrastrutture oligopolistiche che intende analizzare.

Impatto ambientale e colonialità

I sistemi di IA hanno un impatto materiale significativo sull'ambiente, spesso invisibile nelle discussioni tecnocentriche sull'intelligenza artificiale. La materialità dell'IA, apparentemente eterea e virtuale, si manifesta in un consumo energetico massiccio necessario per l'addestramento di modelli di grandi dimensioni, contribuendo all'emissione di gas serra e al cambiamento climatico. L'addestramento di un singolo modello linguistico di grandi dimensioni può produrre emissioni di CO2 equivalenti a quelle di centinaia di voli transatlantici.

Questa impronta ecologica si estende all'estrazione mineraria necessaria per la produzione dell'hardware che sostiene l'IA. Chip, server e dispositivi dipendono dall'estrazione di minerali rari, spesso in condizioni di sfruttamento del lavoro e con gravi danni ambientali nelle comunità locali. Nelle miniere di coltan della Repubblica Democratica del Congo, ad esempio, si estraggono materiali essenziali per l'elettronica in condizioni di grave sfruttamento umano e devastazione ambientale (Miconi e Marrone 2021).

Il rapido sviluppo di nuovi modelli e hardware crea cicli di obsolescenza accelerata, generando rifiuti elettronici difficili da smaltire. Un server dismesso da un data center negli Stati Uniti può finire in una discarica elettronica in Ghana, dove i componenti tossici contaminano il suolo e l'acqua delle comunità circostanti. Gli impatti ambientali negativi dell'IA tendono a colpire in modo sproporzionato le comunità già vulnerabili, mentre i benefici si

concentrano principalmente nel Nord Globale, creando e perpetuando disuguaglianze climatiche globali.

Questi aspetti materiali dell'IA come infrastruttura sociotecnica pongono domande essenziali su come incorporare nella ricerca la dimensione ecologica e materiale dei sistemi di IA; quali relazioni esistano tra le geografie dell'estrazione materiale e le geografie del lavoro digitale invisibile; e come le comunità locali resistano, si adattino o negozino con questi impatti ambientali. Una sociologia che ignori questa dimensione materiale rischia di perpetuare una visione disincarnata e dematerializzata dell'IA che nasconde i suoi reali costi ecologici e sociali.

Da questo punto di vista l'intelligenza artificiale si inserisce in relazioni geopolitiche globali segnate da storiche disuguaglianze coloniali e post-coloniali. I modelli linguistici su larga scala, ad esempio, tendono a privilegiare lingue dominanti come l'inglese, mentre la diversità linguistica globale viene marginalizzata. Un modello addestrato principalmente su testi in inglese mostra prestazioni significativamente inferiori quando interagisce con lingue con minori risorse digitali, riproducendo e amplificando gerarchie linguistiche globali. Inoltre, anche all'interno dello stesso contesto linguistico le gerarchie tradizionali di discriminazione strutturale sono spesso presenti, come nel caso degli afroamericani e la loro rappresentazione in Google (Noble 2018).

Assistiamo a forme di estrattivismo dei dati in cui questi vengono prelevati dalle comunità del Sud Globale senza consenso informato o benefici adeguati, in un processo che richiama le dinamiche estrattive del colonialismo storico (Gray and Suri 2019). Le fotografie personali caricate sui social media da utenti in paesi africani, asiatici o latinoamericani possono essere raccolte senza autorizzazione per addestrare sistemi di riconoscimento facciale, creando una nuova forma di appropriazione.

Le norme, i valori e gli standard incorporati nei sistemi di IA vengono sviluppati principalmente in contesti occidentali e poi imposti globalmente come universali, in un'imposizione di standard tecnologici che ricorda dinamiche coloniali. Concetti come "privacy", "consenso" o "equità" vengono definiti secondo paradigmi occidentali e poi codificati in sistemi tecnologici che circolano globalmente, ignorando concezioni alternative di questi valori in diverse tradizioni culturali (Crawford 2021).

I paesi del Sud Globale diventano consumatori e non produttori di tecnologie di IA, creando nuove forme di dipendenza tecnologica. Un paese che utilizza sistemi di IA sviluppati altrove per la gestione di servizi pubblici essenziali diventa dipendente da aggiornamenti, modifiche e decisioni prese da aziende straniere, limitando la propria sovranità tecnologica. I modelli addestrati principalmente su dati provenienti da culture dominanti tendono inoltre a marginalizzare epistemologie, ontologie e forme di conoscenza alternative, contribuendo a processi di omologazione culturale (Noble 2018).

Queste dinamiche richiedono di sviluppare prospettive decoloniali che riconoscano la pluralità di epistemologie e ontologie che i sistemi di IA tendono a oscurare; analizzino criticamente le relazioni di potere globali incorporate nelle infrastrutture di IA; ed esplorino forme di appropriazione e "indigenizzazione" delle tecnologie di IA in contesti diversi per essere consapevoli di come i sistemi con cui interagisce siano prodotti di specifiche storie coloniali e geopolitiche, e di come questa consapevolezza debba informare sia la loro analisi che l'interpretazione dei loro risultati.

Governance e regolamentazione

L'IA come infrastruttura socio-tecnica è oggetto di crescenti tentativi di regolamentazione e governance, che riflettono diverse visioni politiche e interessi in competizione. Questa dimensione regolativa crea un campo di studio fondamentale per comprendere come valori, norme e visioni del futuro vengano negoziati e incorporati nei sistemi tecnologici.

Diversi paradigmi regolativi emergono globalmente, dai modelli basati sul rischio dell'Unione Europea, esemplificati dall'AI Act, agli approcci più market-friendly degli Stati Uniti, fino ai sistemi di controllo statale centralizzato della Cina. Queste differenze non sono puramente tecniche ma riflettono profonde divergenze nelle concezioni del rapporto tra tecnologia, mercato, stato e cittadini. Un sistema di IA sviluppato per conformarsi alla regolamentazione europea incorpora necessariamente valori di trasparenza, privacy e valutazione del rischio che possono essere assenti in sistemi sviluppati in contesti regolativi diversi (Floridi 2023).

Iniziative come i "model cards" o i "system impact assessments" cercano di rendere più trasparenti i processi di sviluppo e implementazione dell'IA, ma sollevano questioni su cosa costituisca una trasparenza significativa. La documentazione tecnica di un modello linguistico può contenere informazioni dettagliate sui parametri di addestramento, ma rimanere opaca rispetto alle decisioni valoriali che hanno guidato la selezione dei dati o la definizione di cosa costituisca un "buon" risultato. Da ciò emergono tensioni tra approcci tecnocratici alla governance dell'IA e richieste di maggiore partecipazione democratica nelle decisioni che riguardano queste tecnologie. Mentre panel di esperti tecnici definiscono linee guida e standard, crescono le richieste da parte della società civile di processi decisionali più inclusivi che tengano conto delle preoccu-

pazioni e delle esperienze delle comunità potenzialmente impattate dai sistemi di IA (de Neufville e Baum 2021).

Queste dinamiche di governance creano un campo di indagine fondamentale: come i diversi attori – corporazioni, stati, società civile, comunità tecniche – negoziano e contestano le regole che governano l'IA; quali visioni del futuro, valori e interessi siano incorporati nei diversi regimi regolativi; e come le pratiche di governance influenzino lo sviluppo tecnico e l'implementazione dell'IA. Chi studia i sistemi di IA deve essere consapevole di come questi sistemi siano prodotti di specifici regimi regolativi che ne condizionano il funzionamento e la percezione sociale.

Parallelamente l'IA generativa sta profondamente riconfigurando anche l'economia dell'informazione, modificando sia le modalità di produzione e distribuzione dei contenuti, sia le pratiche di consumo e fruizione mediale. La capacità di produrre contenuti personalizzati su scala massiva, adattati a specifici segmenti di pubblico o persino a singoli individui, introduce nuove dinamiche nella competizione per l'attenzione degli utenti (Pasquale 2015).

Piattaforme digitali e social media, già caratterizzati da logiche algoritmiche di personalizzazione, possono ora offrire contenuti non solo selezionati, ma anche generati specificamente per massimizzare l'engagement di ciascun utente. Questo solleva questioni fondamentali relative alla formazione dell'opinione pubblica, alla costruzione di spazi informativi comuni, alla crescente frammentazione e personalizzazione delle esperienze mediali (Jungherr e Schroeder 2023)

La sovrabbondanza informativa, già problematica nell'era digitale, viene amplificata dalla facilità con cui i sistemi generativi possono produrre contenuti in quantità virtualmente illimitate. Questa inflazione contenutistica modifica il valore percepito dell'informazione, le

pratiche di verifica e validazione, le dinamiche di fiducia e credibilità nei confronti delle fonti (Vaccari e Chadwick 2020).

Un fenomeno particolarmente rilevante riguarda la circolazione di contenuti "sintetici" indistinguibili da quelli "autentici", con il conseguente offuscamento dei confini tra reale e simulato, tra documentazione e generazione, tra testimonianza e invenzione. Questa ibridazione complica notevolmente le pratiche sociali di attribuzione del valore e della credibilità, che tradizionalmente si basano sulla distinzione tra categorie come "vero" e "falso", "originale" e "copia", "umano" e "automatico" (Esposito 2022).

L'IA generativa sta provocando una profonda ristrutturazione dei regimi di verità, ovvero dei sistemi sociali attraverso cui vengono stabilite e validate determinate conoscenze come vere o false. La capacità di generare contenuti testuali, visivi e audiovisivi indistinguibili da quelli prodotti da esseri umani mette in crisi meccanismi tradizionali di validazione basati su nozioni come l'autenticità, la testimonianza diretta, la tracciabilità delle fonti (Vaccari e Chadwick 2020).

Questa trasformazione non riguarda solo il problema delle "deepfake" o della disinformazione, ma tocca questioni epistemologiche più profonde relative ai fondamenti sociali della conoscenza. Come vengono ristabiliti i confini tra reale e generato in un contesto in cui questa distinzione diventa tecnicamente indeterminabile? Quali nuove pratiche sociali, tecniche verificative, istituzioni di validazione emergono per gestire questa indeterminatezza?

Riconfigurazione del lavoro cognitivo e creativo

L'intelligenza artificiale generativa non è semplicemente uno strumento neutro che amplifica capacità umane preesistenti, ma un attore attivo nei processi di produzio-

ne, circolazione e contestazione del significato sociale. I contenuti generati algoritmicamente non si limitano a riflettere significati esistenti, ma partecipano alla loro costruzione, trasformazione e negoziazione attraverso meccanismi complessi che meritano un'analisi sociologica approfondita. Queste trasformazioni meritano un'analisi sociologica approfondita, che possa evidenziare non solo gli effetti superficiali, ma anche le riconfigurazioni più profonde nelle relazioni sociali, nei processi di significazione e nelle strutture istituzionali.

Un aspetto frequentemente evidenziato nel dibattito sull'IA generativa è il suo potenziale democratizzante. Strumenti come DALL-E, Midjourney, ChatGPT o Runway ML sembrano effettivamente ridurre le barriere tecniche ed economiche all'espressione creativa, consentendo a persone senza formazione specifica di produrre testi, immagini, musica o video di qualità professionale (Elgammal 2018).

Questa democratizzazione dell'accesso a capacità creative ha effetti complessi. Da un lato, osserviamo l'emergere di nuove comunità di pratica, nuove forme di espressione vernacolare, nuove possibilità di partecipazione culturale per gruppi precedentemente esclusi dai circuiti della produzione creativa istituzionalizzata (Doshi e Hauser 2024).

Dall'altro, riscontriamo il formarsi di nuove gerarchie e disuguaglianze: nell'accesso ai modelli più avanzati e potenti, nella disponibilità di risorse computazionali, nella capacità di utilizzare efficacemente questi strumenti, nella possibilità di integrare output generati in circuiti economici e culturali di valorizzazione (Epstein et al. 2023).

Uno degli aspetti più immediati e visibili dell'IA generativa riguarda la trasformazione del lavoro creativo e cognitivo. Professioni tradizionalmente considerate intrinsecamente umane e resistenti all'automazione – come la scrittura creativa, il design grafico, la composizione mu-

sicale, la creazione di contenuti audiovisivi – si trovano ora a confrontarsi con sistemi capaci di generare contenuti di qualità comparabile in tempi estremamente ridotti. Questa evoluzione sta ridefinendo il valore economico e sociale di diverse forme di lavoro creativo. Da un lato, possiamo osservare dinamiche di dequalificazione e svalorizzazione di specifiche competenze tecniche, quando queste possono essere efficientemente replicate da sistemi automatizzati. Dall'altro, emergono nuove forme di competenza legate alla capacità di utilizzare efficacemente strumenti generativi, di dirigere e curare contenuti generati algoritmicamente, di integrare input umani e artificiali in processi creativi ibridi. Le ricerche condotte in contesti professionali come agenzie di comunicazione, studi di design, redazioni giornalistiche e case editrici mostrano pattern complessi di adattamento, resistenza, appropriazione e risignificazione di questi strumenti. Piuttosto che una semplice sostituzione del lavoro umano, assistiamo alla formazione di nuove ecologie creative in cui esseri umani e sistemi generativi interagiscono in modi diversificati, dando vita a nuove pratiche professionali, nuove identità lavorative e nuove forme di collaborazione (Lee 2022).

Particolarmente interessante è l'emergere di quella che potremmo definire una "estetica dell'IA", caratterizzata da specifici tratti stilistici, convenzioni rappresentative e marche generative che diventano riconoscibili e vengono attivamente ricercate o evitate in diversi contesti creativi (Manovich 2024). Si osservano anche fenomeni di specializzazione professionale attorno a competenze di "prompt engineering", ovvero l'arte di formulare istruzioni efficaci per sistemi generativi al fine di ottenere risultati specifici. Infine, un tema nascente è quello dei diritti d'autore e di come questo si interseca con il lavoro creativo umano, laddove l'intersezione tra uomo e macchina nella produzione di contenuti si fa sempre più sfocata e

rende necessario ripensare le norme attuali (Aufderheide e Jaszi 2018).

Contemporaneamente l'intelligenza artificiale generativa sta contribuendo all'emergere di nuove estetiche, caratterizzate da specifici tratti stilistici, convenzioni rappresentative, modalità espressive. Questi sviluppi non sono semplicemente innovazioni formali, ma hanno profonde implicazioni per i regimi di valore estetico, ovvero per i sistemi sociali attraverso cui determinate espressioni artistiche vengono riconosciute, valorizzate e legittimate come esteticamente significative (Manovich 2018).

Le istituzioni tradizionali di validazione estetica – musei, gallerie, case editrici, critici specializzati – si trovano a dover negoziare con forme espressive che sfidano categorie consolidate come autorialità, originalità, maestria tecnica, intenzionalità artistica. Quali criteri utilizzare per valutare un'opera generata algoritmicamente? Come attribuire valore a creazioni ibride umano-macchina? Che tipo di competenze critiche sono necessarie per apprezzare e interpretare adeguatamente queste nuove forme espressive?

Un aspetto meno evidente ma profondamente significativo dell'IA generativa riguarda la sua influenza sulle temporalità della produzione culturale. La capacità di generare contenuti complessi in tempi estremamente ridotti non modifica solo la quantità di artefatti culturali disponibili, ma anche i ritmi di produzione, circolazione e obsolescenza dei significati sociali (Manovich 2024).

Questa "accelerazione generativa" ha implicazioni profonde per i processi di formazione del valore culturale, che tradizionalmente richiedevano tempi lunghi di sedimentazione, riconoscimento e canonizzazione. Come si sviluppano processi di valorizzazione culturale in contesti di sovrabbondanza e rapida obsolescenza? Quali nuo-

ve pratiche di cura, preservazione e memoria culturale emergono in risposta a questa accelerazione?

Questa tensione tra accelerazione e rallentamento, tra l'immediatezza della generazione algoritmica e la temporalità estesa della sedimentazione culturale, rappresenta uno dei terreni più interessanti per l'analisi sociologica dell'IA generativa come agente di produzione di significato.

Costruzione algoritmica di immaginari sociali

I modelli generativi, addestrati su enormi corpus di dati culturali, assimilano e riproducono specifiche visioni del mondo, rappresentazioni sociali, cornici interpretative e strutture narrative. Quando generano nuovi contenuti, questi modelli non creano dal nulla, ma ricombinano, estendono e trasformano pattern appresi, contribuendo alla riproduzione o alla contestazione di particolari immaginari sociali (Munk 2023).

Un esempio illuminante è rappresentato dalla generazione di immagini di specifiche categorie sociali o professionali. Se chiediamo a DALL-E o Midjourney di visualizzare "un medico", "un insegnante", "un criminale" o "una famiglia", le immagini prodotte tendono a riflettere e rinforzare stereotipi dominanti relativi a genere, razza, classe sociale, abilità fisica. Queste rappresentazioni non sono semplicemente il risultato di bias tecnici, ma il prodotto di complesse stratificazioni di significati sociali incorporati nei dati di addestramento (Luccioni et al. 2024).

Allo stesso tempo, i modelli generativi possono anche contribuire all'emergere di nuovi immaginari, combinando elementi in modi inaspettati, esplorando possibilità rappresentative marginali, amplificando visioni alternative. Questa tensione tra riproduzione e innovazione, tra conferma dello status quo e apertura verso l'ignoto, carat-

terizza profondamente il funzionamento sociale dell'IA generativa (Esposito 2022).

Un aspetto particolarmente rilevante riguarda il ruolo dell'IA generativa nelle politiche della rappresentazione, ovvero nelle lotte sociali attorno a chi ha il diritto di rappresentare determinate identità, esperienze, culture. La capacità di questi sistemi di generare immagini o testi che rappresentano praticamente qualsiasi individuo, gruppo o cultura solleva questioni di appropriazione culturale, di autorità rappresentativa, di autenticità espressiva. Chi ha il diritto di generare immagini che rappresentano specifiche comunità marginali? Chi può utilizzare algoritmi per simulare voci o stili di gruppi minoritari? Come vengono stabiliti i confini tra rappresentazione rispettosa e appropriazione indebita in un contesto di generazione algoritmica?

L'IA generativa, specialmente nei suoi sviluppi linguistici, solleva questioni fondamentali relative al rapporto tra linguaggio e potere. I Large Language Models non sono semplici strumenti di elaborazione linguistica, ma potenti agenti di riproduzione, trasformazione e contestazione di discorsi sociali dominanti. Questi modelli, addestrati su vasti corpus testuali che riflettono relazioni di potere esistenti, tendono naturalmente a riprodurre e amplificare discorsi egemonici, marginalizzando prospettive minoritarie o dissenzienti (Ferrara 2024). Al tempo stesso, la loro capacità di generare testi su qualsiasi argomento, in qualsiasi stile, assumendo qualsiasi voce o posizionamento, introduce elementi di instabilità e potenziale sovversione nelle economie discorsive consolidate.

Particolarmente significative sono le pratiche di "jailbreaking" o di elusione dei limiti imposti ai modelli dai loro sviluppatori. Queste pratiche possono essere interpretate non solo come tentativi di aggirare misure di sicurezza, ma come forme di contestazione dell'autorità epistemica

e discorsiva incorporata nei sistemi generativi. Chi decide quali discorsi sono accettabili e quali problematici? Chi stabilisce i confini tra contenuti appropriati e inappropriati? Queste non sono questioni meramente tecniche, ma profondamente politiche, che implicano specifiche visioni del mondo e relazioni di potere.

Capitolo 3. L'Intelligenza Artificiale Generativa come strumento di ricerca

Oltre ad essere oggetto di studio, l'intelligenza artificiale generativa può configurarsi come strumento innovativo per la ricerca sociale, supportando il ricercatore nelle diverse fasi del processo di indagine:

1. Nella fase di preparazione: i sistemi generativi possono aiutare a esplorare la letteratura esistente, identificare gap conoscitivi, formulare domande di ricerca, sviluppare protocolli di osservazione, generare ipotesi preliminari.
2. Nella fase di raccolta dati: strumenti come Whisper AI o altri software possono trascrivere automaticamente interviste e conversazioni, sistemi di riconoscimento visivo possono classificare immagini e video, chatbot specializzati possono condurre interviste strutturate o semistrutturate.
3. Nella fase di analisi: applicativi come NotebookLM possono indicizzare, recuperare, riassumere e confrontare grandi quantità di dati testuali, identificare temi ricorrenti, suggerire connessioni tra concetti, generare codifiche preliminari.
4. Nella fase di interpretazione: differenti modelli generativi possono proporre letture alternative dei dati, suggerire fondamenti teorici, sviluppare quadri interpretativi, stimolare la riflessività del ricercatore attraverso il confronto con le interpretazioni algoritmiche.
5. Nella fase di comunicazione: l'IA generativa può supportare la scrittura di resoconti, la visualizzazione dei

dati, la traduzione in linguaggi accessibili a diversi pubblici, la diffusione multimediale dei risultati.

L'utilizzo dell'intelligenza artificiale come strumento di ricerca non implica una delega automatica del processo di ricerca agli algoritmi, ma piuttosto una collaborazione uomo-macchina in cui le capacità computazionali e generative dell'IA si integrano con la sensibilità contestuale, l'intuizione teorica e la riflessività critica del ricercatore umano.

Questa collaborazione solleva questioni fondamentali sul ruolo del ricercatore, sulla natura dell'osservazione e dell'interpretazione, sulla costruzione della conoscenza sociale. Se l'approccio tradizionale si basa sull'idea che il ricercatore sia lo strumento primario della ricerca, con la sua soggettività consapevole e riflessiva, l'introduzione di strumenti generativi complica notevolmente questo quadro, introducendo forme di soggettività algoritmica che interagiscono in modi complessi con quella umana. Ciò non sostituisce l'interpretazione del ricercatore, ma offre un ulteriore livello di analisi che può arricchire la comprensione del fenomeno studiato (Morgan 2023).

Al tempo stesso, l'utilizzo dell'IA generativa può amplificare le potenzialità dell'immaginazione sociologica, consentendo di analizzare quantità di dati altrimenti ingestibili, di identificare pattern non immediatamente evidenti, di generare interpretazioni alternative, di sperimentare nuove forme di rappresentazione e comunicazione dei risultati. Questa integrazione non è semplicemente una sostituzione di metodi tradizionali, ma piuttosto un ampliamento del repertorio di strumenti a disposizione del ricercatore (Davidson 2024).

La sfida consiste nel trovare un equilibrio tra amplificazione e delega, tra potenziamento delle capacità analitiche e mantenimento del controllo interpretativo,

tra sfruttamento delle potenzialità generative e preservazione della riflessività critica. Un equilibrio che richiede consapevolezza metodologica, sensibilità epistemologica e responsabilità etica.

La ricerca sociale è stata tradizionalmente caratterizzata da un'attenzione meticolosa al dettaglio, da un impegno profondo con i dati e da un processo interpretativo riflessivo che richiede tempo e immersione. Gli strumenti di intelligenza artificiale generativa non sostituiscono questi aspetti fondamentali, ma offrono modalità complementari per affrontare alcune delle sfide più onerose della ricerca qualitativa: la gestione di grandi volumi di dati testuali, l'identificazione di pattern nascosti, la trascrizione di interviste, e la navigazione attraverso complessi corpus (Zhang et al. 2025). Questo processo non solo riduce il carico di lavoro manuale, ma ne facilita anche l'esplorazione (Hitch 2024).

È importante sottolineare che l'integrazione dell'IA nel lavoro non implica una semplificazione del processo di ricerca. Al contrario, richiede un livello aggiuntivo di competenza e riflessività da parte del ricercatore, che deve essere in grado di valutare criticamente gli output generati dall'IA, riconoscendone i limiti e i bias potenziali. Come abbiamo ampiamente descritto nel precedente capitolo, il ricercatore deve essere consapevole che l'IA non è uno strumento neutrale, ma porta con sé specifiche logiche algoritmiche e bias potenziali.

In questo capitolo, esamineremo come strumenti specifici di IA generativa possano essere impiegati nelle diverse fasi della ricerca qualitativa, dalle trascrizioni automatiche all'analisi tematica, dalla codifica all'interpretazione. Discuteremo anche le sfide metodologiche, etiche ed epistemologiche che emergono quando integriamo questi strumenti nelle nostre pratiche di ricerca. L'obiettivo non è proporre l'IA come sostituto del pensiero critico e

dell'immaginazione sociologica, ma piuttosto esplorare come essa possa amplificare e arricchire le capacità analitiche e interpretative del ricercatore sociale.

Trascrizione e tagging automatico

La trascrizione di interviste, registrazioni di campo e altri materiali audio rappresenta una delle attività più dispendiose in termini di tempo nella ricerca qualitativa. Quello che un tempo chiedeva giorni o addirittura settimane di lavoro meticoloso può ora essere completato in una frazione del tempo grazie a strumenti come Whisper AI di OpenAI, un sistema di riconoscimento vocale open-source progettato per la trascrizione automatica di audio in vari formati e lingue.

Whisper AI si basa su un'architettura di rete neurale addestrata su un vasto corpus di dati audio multilingue, che le conferisce una notevole robustezza rispetto alle variazioni di accento, al rumore di fondo e ad altri fattori che tradizionalmente complicano la trascrizione automatica. A differenza dei precedenti sistemi di riconoscimento vocale, che spesso richiedevano un addestramento specifico per ogni lingua o contesto, Whisper AI opera come un modello generalista capace di adattarsi a diverse situazioni comunicative.

La sua architettura encoder-decoder, simile a quella utilizzata da altri modelli di linguaggio, le consente di elaborare l'input audio come una sequenza di caratteristiche spettrali e di generare output testuali attraverso un processo sequenziale. Questo approccio permette non solo di trascrivere accuratamente il parlato, ma anche di identificare automaticamente la lingua utilizzata e di inserire la punteggiatura appropriata, creando trascrizioni leggibili e ben strutturate.

Un aspetto particolarmente rilevante per la ricerca qualitativa è la capacità di Whisper AI di gestire registrazioni in contesti non ideali, come discussioni di gruppo dove più persone parlano, interviste in ambienti rumorosi, o conversazioni che mescolano più lingue. Questa flessibilità la rende adatta per diverse situazioni di ricerca sul campo, dalle interviste formali alle osservazioni partecipanti in contesti caotici.

L'integrazione di Whisper AI nel processo di ricerca qualitativa inizia tipicamente con la conversione delle registrazioni audio in trascrizioni grezze. Questo processo può essere eseguito localmente su un computer del ricercatore o attraverso servizi basati su cloud che utilizzano l'API di Whisper, a seconda delle esigenze di sicurezza e privacy dei dati.

Una volta ottenute le trascrizioni iniziali, il ricercatore può procedere con una revisione e correzione manuale. Questo passaggio rimane essenziale: nonostante l'elevata accuratezza di WhisperAI (specialmente in condizioni audio ottimali), alcuni errori o ambiguità possono persistere, particolarmente in presenza di terminologia specialistica, riferimenti culturali specifici o espressioni idiomatiche. La revisione manuale offre anche l'opportunità di aggiungere note contestuali, osservazioni non verbali e altri elementi che il sistema automatico non può catturare.

Il vero valore aggiunto di Whisper AI emerge quando si passa dalla semplice trascrizione al tagging automatico. Il sistema può essere configurato per identificare automaticamente cambiamenti di interlocutore, segmentare il testo in unità tematiche, o evidenziare passaggi significativi in base a parole chiave o pattern linguistici predefiniti. Questa capacità di pre-strutturare i dati facilita enormemente le successive fasi di codifica e analisi.

Nonostante i suoi evidenti vantaggi, l'utilizzo di Whisper AI nella ricerca qualitativa presenta anche alcu-

ne limitazioni che meritano attenzione critica. In primo luogo, come tutti i sistemi di IA, Whisper riflette i bias presenti nei dati di addestramento. Questi possono manifestarsi in una minore accuratezza nella trascrizione di accenti non standard, dialetti regionali o varietà linguistiche minoritarie.

In secondo luogo, la dipendenza da trascrizioni automatiche può potenzialmente distanziare il ricercatore dal materiale originale. Il processo tradizionale di trascrizione manuale, sebbene laborioso, costituisce spesso un primo momento di immersione nei dati e di riflessione analitica. Delegando questo processo a un sistema automatico, il ricercatore potrebbe perdere alcune di queste intuizioni iniziali.

Infine, è importante ricordare che Whisper AI, come qualsiasi strumento tecnologico, non è neutrale ma incorpora specifiche visioni del linguaggio e della comunicazione. Ad esempio, la sua tendenza a normalizzare il parlato inserendo punteggiatura convenzionale potrebbe oscurare caratteristiche significative del discorso orale come interruzioni, sovrapposizioni o esitazioni che potrebbero essere rilevanti per l'analisi.

Per queste ragioni, è essenziale approcciarsi all'utilizzo di Whisper AI nella ricerca qualitativa con una consapevolezza critica, considerandolo come uno strumento complementare piuttosto che sostitutivo dell'engagement attivo del ricercatore con i dati empirici.

Indicizzazione, recupero informazioni, sommari e comparazione di contenuti testuali

Una volta che le trascrizioni e altri materiali testuali sono stati raccolti, il ricercatore qualitativo si trova spesso di fronte alla sfida di gestire, organizzare e navigare attraverso un vasto corpus di dati. In questo contesto, strumenti

come NotebookLM rappresentano un significativo passo avanti, offrendo funzionalità avanzate di indicizzazione, recupero, sintesi e comparazione dei dati qualitativi.

NotebookLM è un ambiente di lavoro che combina la flessibilità dei notebook interattivi con le capacità dei modelli linguistici di grandi dimensioni (LLM). Basato su Google Gemini, NotebookLM consente di integrare analisi qualitative tradizionali con funzionalità avanzate di elaborazione del linguaggio naturale.

L'ambiente permette ai ricercatori di caricare, organizzare e annotare documenti testuali mantenendo una struttura flessibile che facilita tanto l'analisi sistematica quanto l'esplorazione intuitiva. A differenza dei software tradizionali di analisi qualitativa dei dati (CAQDAS), NotebookLM integra nativamente capacità di elaborazione semantica profonda che vanno oltre la semplice ricerca per parole chiave.

Un aspetto distintivo di NotebookLM è la sua interfaccia, che permette ai ricercatori di definire workflow analitici personalizzati e di integrarli con altre analisi quantitative o visualizzazioni, creando un ambiente misto tra diversi metodi di ricerca.

L'indicizzazione semantica rappresenta uno dei contributi più significativi che l'IA generativa offre alla ricerca qualitativa. A differenza dei sistemi tradizionali di indicizzazione basati su parole chiave o operatori booleani, l'indicizzazione semantica in NotebookLM si basa sulla rappresentazione vettoriale del testo attraverso embedding linguistici.

Questo approccio consente di mappare il significato dei testi in uno spazio multidimensionale dove la prossimità riflette la similarità semantica. In pratica, questo significa che il ricercatore può individuare passaggi concettualmente correlati anche quando non condividono lo stesso lessico specifico. Ad esempio, una ricerca sulla "resilienza

comunitaria" potrebbe recuperare passaggi che parlano di "capacità di adattamento collettivo" o "strategie di sopravvivenza sociale", anche se queste espressioni non condividono parole in comune.

La funzione di retrieval potenziata da IA permette interrogazioni in linguaggio naturale come "Trova tutti i passaggi dove gli intervistati esprimono sentimenti ambivalenti verso le istituzioni locali" o "Identifica le descrizioni di pratiche culturali tradizionali adattate al contesto urbano contemporaneo". Questo tipo di ricerca concettuale rappresenta un salto qualitativo rispetto ai tradizionali metodi di ricerca testuale, avvicinandosi maggiormente al modo in cui i ricercatori interagiscono naturalmente con i loro dati.

Una delle applicazioni più promettenti dell'IA generativa è la capacità di sintetizzare ampie sezioni di dati mantenendo al contempo la ricchezza e la complessità del materiale originale. NotebookLM offre funzionalità di summarizing che vanno oltre il semplice riassunto estrattivo (che seleziona frasi esistenti) per produrre sintesi astrattive che riformulano le informazioni chiave in nuove formulazioni.

Queste sintesi possono essere generate a diversi livelli di granularità: dalla sintesi di una singola intervista, alla comparazione di più interviste su un tema specifico, fino alla generalizzazione di pattern emergenti attraverso l'intero corpus di dati. Il ricercatore mantiene il controllo sul livello di dettaglio e sulla focalizzazione tematica della sintesi, permettendo analisi sia ampie che profonde.

Un'applicazione particolarmente utile è la generazione assistita di memo analitici, un elemento fondamentale nella Grounded Theory e in altri approcci qualitativi. NotebookLM può suggerire memo basati sull'analisi di specifici segmenti di dati, evidenziando connessioni potenzialmente significative e proponendo interpretazioni

preliminari che il ricercatore può poi valutare, modificare o estendere.

Questo tipo di memo, generato automaticamente ma sempre sottoposto alla valutazione critica del ricercatore, può stimolare nuove direzioni analitiche o evidenziare connessioni che potrebbero altrimenti restare implicite.

La comparazione sistematica tra diversi segmenti di dati è un processo fondamentale nell'analisi qualitativa, alla base di tecniche come la comparazione costante nella Grounded Theory o l'analisi del discorso comparativa. NotebookLM offre diverse modalità di comparazione assistita da IA che possono arricchire questo processo:

1. Comparazione tematica: Il sistema può identificare automaticamente temi comuni e divergenti tra diverse interviste o gruppi di interviste, evidenziando pattern di similarità e contrasto a livello di contenuto.
2. Comparazione retorica: Può analizzare le strutture discorsive e le strategie retoriche utilizzate dai diversi partecipanti, evidenziando differenze nell'uso di metafore, narrativi, argomentazioni o posizionamenti.
3. Comparazione contestuale: Permette di esaminare come temi simili siano articolati in contesti diversi, evidenziando l'influenza di fattori situazionali o strutturali sulle narrative individuali.
4. Comparazione longitudinale: Consente di tracciare l'evoluzione di specifici temi o discorsi nel tempo, particolarmente utile per studi che coinvolgono interviste ripetute o dati raccolti in fasi diverse.

L'integrazione di NotebookLM e strumenti simili nel processo di ricerca qualitativa solleva questioni metodologiche ed epistemologiche che meritano attenzione critica:

La dipendenza tecnologica introduce nuove vulnerabilità nel processo di ricerca. L'affidamento a piattaforme proprietarie o a modelli linguistici sviluppati da grandi

aziende tecnologiche solleva questioni di accesso, controllo dei dati e sostenibilità a lungo termine della ricerca.

La black-box epistemologica dei sistemi di IA generativa può minare la trasparenza metodologica che è fondamentale nella ricerca qualitativa. Il ricercatore deve interrogarsi criticamente su come vengono generate le sintesi e le interpretazioni automatiche, e su quali presupposti teorici e computazionali si basano.

Il rischio di distacco analitico emerge quando l'automazione di parti del processo analitico potrebbe ridurre l'immersione profonda del ricercatore nei dati, un elemento considerato cruciale in molte tradizioni qualitative.

Infine, la standardizzazione metodologica implicita in questi strumenti potrebbe privilegiare certi tipi di analisi e interpretazioni rispetto ad altri, potenzialmente riducendo la diversità metodologica e teorica nel campo della ricerca qualitativa.

Per mitigare questi rischi, è essenziale che l'utilizzo di NotebookLM e strumenti simili sia accompagnato da una costante riflessività metodologica e da un'integrazione consapevole con pratiche di ricerca qualitativa consolidate.

Brainstorming mediato dalla collaborazione con modelli linguistici multipli

L'identificazione di temi significativi e la codifica sistematica dei dati qualitativi sono processi fondamentali che richiedono tradizionalmente un intenso lavoro interpretativo da parte di un team di ricercatori. L'intelligenza artificiale generativa offre nuove possibilità per supportare e arricchire questi processi, non tanto sostituendo il giudizio umano quanto ampliando le capacità analitiche attraverso l'uso di modelli multipli in un approccio collaborativo.

I modelli di linguaggio di grande scala come GPT, Claude, o Llama possono essere utilizzati per identificare pattern tematici nei dati qualitativi attraverso diverse logiche computazionali:

1. Analisi semantica latente: I modelli possono identificare cluster di significato all'interno del testo, raggruppando passaggi concettualmente correlati anche quando utilizzano terminologie diverse.
2. Rilevamento di anomalie discorsive: Possono evidenziare passaggi che si discostano significativamente dai pattern dominanti nel corpus, potenzialmente indicando temi emergenti o prospettive divergenti.
3. Analisi di sentiment e stance: Oltre ai contenuti espliciti, i modelli possono rilevare atteggiamenti, emozioni e posizionamenti impliciti nel discorso, arricchendo l'analisi tematica con dimensioni affettive e valutative.
4. Mappatura di reti concettuali: Possono tracciare relazioni tra concetti all'interno del corpus, visualizzando come diversi temi si intersecano e si influenzano tra loro.

A differenza dei tradizionali approcci computazionali all'analisi testuale, che spesso si basano su conteggi di frequenza o co-occorrenze lessicali, i modelli di linguaggio attuali possono cogliere sfumature semantiche e contestuali, avvicinandosi maggiormente alla comprensione interpretativa che caratterizza l'analisi qualitativa umana.

Un approccio particolarmente promettente consiste nell'utilizzo di modelli diversi in parallelo per la codifica dei dati qualitativi, sfruttando le loro differenti architetture, basi di addestramento e capacità. Questo metodo, che potremmo definire "triangolazione algoritmica", si ispira al principio metodologico della triangolazione nelle scienze sociali, dove l'uso di molteplici prospettive analitiche arricchisce la comprensione del fenomeno studiato.

In pratica, lo stesso corpus di dati viene processato da modelli diversi, ciascuno configurato con parametri specifici per enfatizzare determinate dimensioni analitiche. Ad esempio:

- Un modello potrebbe essere istruito per adottare un approccio bottom-up, identificando temi emergenti senza categorie predefinite, in linea con la logica della Grounded Theory.
- Un secondo modello potrebbe operare in modalità top-down, applicando un framework teorico esplicito specificato dal ricercatore.
- Un terzo modello potrebbe focalizzarsi specificamente su elementi discorsivi e retorici, analizzando come i temi vengono articolati linguisticamente.

Le codifiche generate dai diversi modelli vengono poi confrontate e integrate. Le convergenze tra modelli possono suggerire temi robusti, mentre le divergenze possono evidenziare aree di ambiguità o complessità interpretativa che meritano un'attenzione particolare da parte del ricercatore.

La codifica collaborativa rappresenta un cambio di paradigma rispetto sia alla codifica puramente manuale sia all'automazione completa. In questo approccio, ricercatore e sistema di IA interagiscono iterativamente, generando un processo dialettico che valorizza tanto la sensibilità interpretativa umana quanto le capacità computazionali della macchina.

Il processo tipicamente si articola in diverse fasi:

1. Codifica esplorativa: Il ricercatore inizia con una lettura approfondita di un sottoinsieme di dati, sviluppando un insieme preliminare di codici e annotazioni. Contemporaneamente, il sistema di IA genera una propria codifica indipendente dello stesso materiale.
2. Comparazione e negoziazione: Le due codifiche vengono confrontate, identificando convergenze, divergenze

e complementarità. Il ricercatore valuta criticamente i codici proposti dall'IA, potenzialmente incorporando intuizioni che potrebbero essere sfuggite all'analisi umana iniziale.
3. Co-costruzione del libro codici: Emerge un libro codici affinato che riflette questa negoziazione interpretativa. Il sistema di IA viene ricalibrato in base a questo framework condiviso.
4. Applicazione estesa: Il libro codici co-costruito viene applicato al resto del corpus, con il sistema di IA che propone codifiche che il ricercatore può accettare, modificare o respingere.
5. Raffinamento iterativo: Il processo continua ciclicamente, con il libro codici che evolve in risposta a nuovi dati e intuizioni.

L'integrazione dell'IA nel processo di codifica solleva inevitabilmente questioni di affidabilità, validità e rigore metodologico. Per affrontare queste preoccupazioni, sono state sviluppate diverse tecniche specifiche:
1. Auditing algoritmico: Il sistema registra e rende esplicito il suo processo decisionale per ogni codifica, permettendo al ricercatore di esaminare come e perché certi passaggi sono stati associati a specifici codici.
2. Benchmark di sensibilità: Test sistematici su come variazioni minori nel testo influenzano la codifica, rivelando potenziali inconsistenze o bias nel sistema.
3. Controlli di coerenza inter-codificatore: Confronto statistico tra codifiche umane e automatiche per quantificare il livello di concordanza e identificare aree di sistematica divergenza.
4. Validazione incrociata contestuale: Verifica che il sistema mantenga coerenza interpretativa attraverso contesti diversi, ad esempio controllando che passaggi semanticamente simili in interviste diverse ricevano codifiche compatibili.

5. Test di robustezza temporale: Controlli periodici per assicurare che la codifica rimanga stabile nel tempo, evitando "derive interpretative" durante il processo di analisi.

Queste tecniche non solo migliorano la qualità della codifica assistita, ma contribuiscono anche a costruire una documentazione metodologica dettagliata che aumenta la trasparenza e la verificabilità della ricerca qualitativa.

L'approccio collaborativo alla codifica qualitativa ha profonde implicazioni epistemologiche che meritano una riflessione critica. Tradizionalmente, nella ricerca qualitativa il ricercatore è considerato lo strumento primario dell'indagine, e la sua soggettività interpretativa è vista non come un limite da minimizzare ma come una risorsa analitica da rendere esplicita attraverso la riflessività.

L'introduzione dell'IA in questo processo solleva interrogativi su chi (o cosa) produce conoscenza e su come la mediazione algoritmica influenzi la natura di questa conoscenza. La codifica collaborativa può essere vista attraverso diverse lenti epistemologiche.

Dalla prospettiva del realismo critico, l'approccio multi-modello può essere interpretato come un tentativo di triangolazione che cerca di avvicinarsi a una comprensione più accurata di meccanismi causali sottostanti, assumendo che esista una realtà sociale indipendente dalle nostre interpretazioni che può essere colta, sebbene imperfettamente, attraverso molteplici prospettive analitiche.

Per il costruttivismo sociale, la codifica collaborativa rappresenta un nuovo spazio di negoziazione di significato, dove le interpretazioni emergono dall'interazione tra diversi sistemi simbolici – quelli umani e quelli computazionali – ciascuno con i propri presupposti, limitazioni e potenzialità.

Dal punto di vista post-umanista, questo approccio sfida la distinzione binaria tra agentività umana e computazionale, suggerendo la possibilità di un'epistemologia distribuita dove la conoscenza emerge da assemblaggi sociotecnici ibridi piuttosto che da soggetti puramente umani.

Queste diverse letture epistemologiche non sono meramente teoriche, ma hanno implicazioni pratiche per come i ricercatori approcciano, documentano e legittimano il loro uso dell'IA nella ricerca qualitativa.

Integrare efficienza computazionale e interpretazione

L'integrazione dell'intelligenza artificiale generativa nella ricerca sociale comporta sfide significative che vanno oltre le questioni tecniche per toccare il cuore stesso della pratica riflessiva nelle scienze sociali. Questa sezione esplora le tensioni, le contraddizioni e le potenziali riconfigurazioni che emergono quando tecnologie computazionali avanzate incontrano tradizioni metodologiche profondamente radicate nella riflessività e nell'interpretazione situata.

Una delle tensioni più evidenti nell'uso dell'IA nella ricerca qualitativa riguarda il rapporto tra efficienza e immersione. La promessa di automazione e scalabilità offerta dagli strumenti di IA entra potenzialmente in conflitto con il valore dell'immersione profonda nei dati che caratterizza l'approccio qualitativo.

L'immersione prolungata e l'engagement riflessivo con i dati qualitativi non sono semplicemente passaggi metodologici, ma pratiche epistemiche che generano forme specifiche di conoscenza situata. L'automazione di parti del processo analitico rischia di alterare questa relazione intensiva con i dati, potenzialmente riducendo quella

che Clifford Geertz chiamava "descrizione densa" a una forma più superficiale di mappatura tematica.

La sfida metodologica, quindi, non è tanto scegliere tra efficienza e immersione, quanto riconfigurare il processo di ricerca in modo da sfruttare i vantaggi computazionali preservando al contempo gli spazi per una riflessività significativa.

La riflessività metodologica nelle scienze sociali richiede che il ricercatore renda espliciti i propri presupposti, posizionamenti e processi analitici. L'integrazione di sistemi di IA complica questo requisito a causa della loro relativa opacità e della difficoltà di articolare pienamente come generano le loro interpretazioni.

Il problema non è solo tecnico (la black-box dei modelli di deep learning) ma anche epistemologico: come rendere conto riflessivamente di processi analitici che sono parzialmente delegati a sistemi che operano attraverso logiche computazionali non completamente accessibili alla coscienza riflessiva del ricercatore?

Alcune strategie emergenti per affrontare questa sfida includono:

1. Documentazione algoritmica espansa: Elaborare documentazioni dettagliate non solo dei risultati dell'analisi assistita dall'IA, ma anche dei parametri, delle configurazioni e delle decisioni di progettazione che hanno influenzato il funzionamento del sistema.
2. Analisi riflessiva degli output divergenti: Esaminare sistematicamente i casi in cui l'interpretazione umana e quella algoritmica divergono significativamente, utilizzando queste divergenze come opportunità per interrogare criticamente tanto i presupposti del ricercatore quanto le tendenze del sistema.
3. Pratiche di co-interpretazione documentata: Sviluppare e documentare processi espliciti attraverso cui le interpretazioni algoritmiche vengono valutate,

modificate o integrate nella comprensione complessiva del ricercatore.
4. Audit riflessivi periodici: Condurre revisioni sistematiche di come l'uso di strumenti di IA influisce sulle direzioni analitiche, sulle conclusioni emergenti e sulle scelte metodologiche nel corso del progetto di ricerca.

Queste strategie non eliminano completamente il problema dell'opacità algoritmica, ma rappresentano tentativi pragmatici di estendere la pratica riflessiva per includere la mediazione tecnologica come oggetto esplicito di riflessione metodologica.

L'integrazione dell'IA nella ricerca qualitativa non solo modifica le pratiche analitiche, ma potenzialmente riconfigura la soggettività stessa del ricercatore e il suo rapporto con il processo conoscitivo. Il tradizionale modello del ricercatore come "strumento umano" primario dell'indagine qualitativa viene complicato dall'emergere di quella che potremmo definire una "soggettività estesa" o "distribuita".

In questo modello emergente, l'interpretazione non emerge esclusivamente dalla coscienza riflessiva individuale del ricercatore, ma dall'interazione dinamica tra capacità umane e computazionali. Il ricercatore si trova a navigare in uno spazio interpretativo parzialmente co-costituito da logiche algoritmiche, richiedendo nuove forme di riflessività che considerino questa ibridità epistemica.

Questa riconfigurazione della soggettività del ricercatore solleva questioni fondamentali:

- Come mantenere un'agentività riflessiva significativa quando parti del processo interpretativo sono delegate a sistemi computazionali?
- Come articolare e rendere conto delle diverse forme di mediazione che intervengono tra ricercatore e

dati empirici?
- Quali nuove pratiche di riflessività sono necessarie per navigare consapevolmente questa assemblaggio socio-tecnico?

Queste domande non hanno risposte definitive, ma indicano la necessità di una continua sperimentazione metodologica accompagnata da una profonda riflessione epistemologica.

Un paradosso fondamentale nell'uso dell'IA nella ricerca qualitativa riguarda la tensione tra automazione e interpretazione. La ricerca qualitativa, specialmente nelle sue varianti più interpretative e fenomenologiche, si basa su una comprensione del significato che è intrinsecamente contestuale, contingente e intersoggettiva. L'automazione, d'altra parte, funziona attraverso la standardizzazione e formalizzazione dei processi.

Questo paradosso si manifesta in vari modi:

1. Standardizzazione vs. sensibilità contestuale: I sistemi automatizzati tendono a favorire approcci standardizzati che possono entrare in tensione con l'attenzione alla specificità contestuale che caratterizza la ricerca qualitativa.
2. Scalabilità vs. profondità: L'automazione privilegia la scalabilità (analizzare più dati) potenzialmente a scapito della profondità interpretativa che richiede un engagement intensivo con porzioni limitate di dati.
3. Efficienza vs. apertura esplorativa: L'ottimizzazione per l'efficienza può ridurre gli spazi per l'esplorazione aperta e la serendipità analitica che spesso producono intuizioni significative nella ricerca qualitativa.
4. Generalizzazione vs. particolarità: I modelli computazionali tendono verso la generalizzazione, mentre molte tradizioni qualitative valorizzano la particolarità e l'unicità dei fenomeni sociali.

Riconoscere questi paradossi non significa rifiutare l'integrazione dell'IA nella ricerca qualitativa, ma piuttosto approcciarla con consapevolezza critica, sviluppando pratiche ibride che possano navigare produttivamente queste tensioni.

Di fronte alle sfide delineate, emerge la necessità di sviluppare approcci che non si limitino né ad abbracciare acriticamente né a rifiutare categoricamente l'integrazione dell'IA nella ricerca qualitativa. Piuttosto, si profila la possibilità di una "metodologia qualitativa criticamente aumentata" che consideri le tecnologie computazionali non come sostituti neutrali, ma come attori che partecipano attivamente alla configurazione delle pratiche conoscitive.

Questo approccio si caratterizza per:

1. Riflessività tecnologica espansa: Una riflessività che si estende esplicitamente alle mediazioni tecnologiche, considerando come gli strumenti computazionali modificano il processo di ricerca.
2. Ibridazione metodologica consapevole: L'integrazione deliberata di processi automatizzati e interpretativi, con attenzione esplicita a dove e come si intersecano.
3. Documentazione multi-livello: Pratiche di documentazione che rendono visibili tanto i processi interpretativi umani quanto le operazioni computazionali che hanno contribuito all'analisi.
4. Sperimentazione metodologica sistematica: Un approccio esplorativo che sperimenta diverse configurazioni di collaborazione uomo-macchina, documentando sistematicamente i loro effetti sui processi e risultati analitici.
5. Contestualizzazione socio-tecnica: L'attenzione alle condizioni sociali, economiche e politiche che plasmano lo sviluppo e l'implementazione delle tecnologie di IA utilizzate nella ricerca.

Questa metodologia criticamente aumentata non risolve definitivamente le tensioni discusse, ma offre un framework per navigarle produttivamente, mantenendo la profondità riflessiva della tradizione qualitativa mentre esplora le potenzialità analitiche offerte dall'intelligenza artificiale generativa.

L'integrazione dell'intelligenza artificiale nella ricerca qualitativa solleva questioni etiche significative che vanno oltre le considerazioni metodologiche ed epistemologiche discusse finora. Queste questioni riguardano non solo la protezione dei partecipanti alla ricerca, ma anche le implicazioni più ampie per la pratica della ricerca sociale e per le comunità studiate.

La natura dei modelli di linguaggio di grandi dimensioni solleva nuove sfide per i principi etici tradizionali della ricerca sociale. A differenza degli strumenti di analisi qualitativa tradizionali, che operano esclusivamente su dataset locali, molti strumenti di IA generativa funzionano attraverso API remote, sollevando questioni sulla custodia e circolazione dei dati.

Quando le trascrizioni delle interviste o le note di campo vengono inviate a questi sistemi, emergono numerose domande:

1. Consenso informato espanso: I partecipanti alla ricerca hanno veramente compreso e acconsentito all'elaborazione delle loro parole attraverso sistemi di IA? Come aggiornare i protocolli di consenso informato per includere esplicitamente queste nuove forme di analisi?
2. Rischi di re-identificazione: I modelli di linguaggio avanzati possono potenzialmente riconoscere pattern linguistici unici che potrebbero, in combinazione con altre informazioni, compromettere l'anonimato dei partecipanti anche quando gli identificatori diretti sono stati rimossi.
3. Custodia dei dati a lungo termine: Come garantire che i dati condivisi con sistemi di IA non vengano conserva-

ti o utilizzati in modi non previsti dal consenso originale, specialmente considerando la rapida evoluzione dei termini di servizio delle piattaforme commerciali?
4. Disparità di comprensione tecnologica: Come garantire un consenso veramente informato quando esiste spesso un divario significativo tra la comprensione tecnica del ricercatore e quella dei partecipanti riguardo al funzionamento dei sistemi di IA?

Un esempio potrebbe essere quello di un team di ricerca che studia esperienze di vulnerabilità sanitaria implementando un "protocollo di mediazione dei dati" in cui le trascrizioni vengono sistematicamente modificate prima dell'elaborazione con IA attraverso:

- Riformulazione delle espressioni idiosincratiche che potrebbero essere identificabili.
- Sostituzione di riferimenti geografici e temporali specifici con categorie più generali.
- Creazione di "interviste composite" che combinano elementi di multiple conversazioni reali.
- Sviluppo di un sistema di "consenso dinamico" che permette ai partecipanti di rivedere e approvare specifici estratti prima della loro analisi attraverso IA.

Queste strategie non eliminano completamente i rischi, ma rappresentano tentativi di adattare i principi etici consolidati alle nuove realtà tecnologiche della ricerca assistita da IA.

I modelli di linguaggio di grandi dimensioni sono notoriamente soggetti a bias derivanti dai dati di addestramento, che spesso riflettono e potenzialmente amplificano disuguaglianze sociali esistenti. Questo solleva preoccupazioni particolari per la ricerca qualitativa, che spesso si occupa di esperienze marginali o subalterne e ha una lunga tradizione di riflessione critica sui rapporti di potere nella produzione di conoscenza.

Le questioni specifiche includono:

1. Bias linguistici e culturali: I modelli tendono a funzionare meglio su varietà linguistiche dominanti e standardizzate, potenzialmente marginalizzando ulteriormente varietà non standard, dialetti o lingue con minore rappresentazione nei dati di addestramento.
2. Riproduzione di stereotipi e pregiudizi: I sistemi possono inconsapevolmente rinforzare visioni stereotipate di gruppi sociali già marginalizzati attraverso il modo in cui codificano o interpretano le loro esperienze.
3. Disparità di performance analitica: L'efficacia degli strumenti di IA può variare significativamente a seconda della popolazione studiata, creando una forma di "ingiustizia algoritmica" dove certe esperienze sono rappresentate con maggiore accuratezza e nuance di altre.
4. Amplificazione di voci dominanti: L'uso acritico di strumenti di IA potrebbe amplificare ulteriormente voci e prospettive già privilegiate nella letteratura esistente e nei discorsi pubblici.

In questo senso si potrebbe adottare un approccio di "correzione di prospettiva" che include:

- Coinvolgimento di membri delle comunità studiate nella valutazione e interpretazione degli output dell'IA
- Utilizzo di tecniche di "fine-tuning" o "prompt engineering" esplicite per contrastare bias noti
- Triangolazione sistematica con metodi interpretativi sviluppati all'interno delle comunità studiate
- Documentazione trasparente dei limiti e dei bias potenziali nell'analisi assistita da IA

Questi approcci riflettono un impegno più ampio verso una "giustizia algoritmica" nella ricerca sociale che riconosce come le tecnologie computazionali non siano

strumenti neutrali ma interventi situati in contesti sociali e politici specifici.

L'uso dell'IA generativa nella ricerca qualitativa complica le nozioni tradizionali di autorialità, contribuzione e proprietà intellettuale. Quando interpretazioni significative emergono dall'interazione tra ricercatore e sistema di IA, emerge la questione di come attribuire adeguatamente queste intuizioni e di chi possa rivendicarne la paternità.

Queste tensioni si manifestano in vari modi:

1. Autorialità distribuita: Come riconoscere adeguatamente il contributo dei sistemi di IA al processo interpretativo senza antropomorfizzarli inappropriatamente o sminuire l'agentività intellettuale del ricercatore?
2. Proprietà delle inferenze: Chi "possiede" le intuizioni analitiche generate attraverso l'elaborazione computazionale di dati qualitativi – il ricercatore, gli sviluppatori del modello, o i partecipanti le cui esperienze sono state analizzate?
3. Trasparenza metodologica: Come documentare adeguatamente il ruolo dell'IA nel processo di ricerca in modo che lettori e revisori possano valutare criticamente la validità delle interpretazioni presentate?
4. Legittimità accademica: Come navigare le norme disciplinari e istituzionali che possono vedere con sospetto l'uso di strumenti di IA nella produzione di conoscenza qualitativa?

Pratiche innovative per affrontare queste questioni possono essere diverse, ad esempio:

- Includere sezioni dedicate negli articoli scientifici che dettagliano specificamente come e dove l'IA ha contribuito all'analisi.
- Sviluppare nuove convenzioni di citazione che riconoscono esplicitamente i modelli utilizzati e il loro ruolo nel processo interpretativo.
- Creare repository pubblici di prompt, configurazio-

ni e output intermedi che documentano l'interazione tra ricercatore e sistema.
- Proporre framework di "autorialità responsabile" che delineano chiaramente la natura e i limiti del contributo algoritmico.

Queste pratiche emergenti riflettono un tentativo di riconfigurare norme accademiche consolidate in risposta a nuove realtà socio-tecniche, mantenendo al contempo i valori fondamentali di integrità, trasparenza e responsabilità nella produzione di conoscenza scientifica.

L'adozione di strumenti di IA nella ricerca qualitativa solleva anche questioni di sostenibilità ambientale e accessibilità che hanno dimensioni etiche significative. I modelli di linguaggio di grandi dimensioni richiedono risorse computazionali sostanziali, con conseguenti costi economici e ambientali.

Questo solleva preoccupazioni riguardo:

1. Impronta di carbonio: L'addestramento e l'inferenza di modelli di linguaggio di grandi dimensioni comportano un consumo energetico significativo. Come riconciliare questo impatto ambientale con l'impegno delle scienze sociali verso la sostenibilità?
2. Disuguaglianze di accesso: Gli strumenti di IA più avanzati sono spesso accessibili solo attraverso piattaforme commerciali con costi significativi, creando potenziali disparità tra istituzioni ben finanziate e ricercatori con risorse limitate.
3. Dipendenza da infrastrutture proprietarie: L'affidamento a piattaforme commerciali per strumenti di ricerca essenziali solleva preoccupazioni sulla sovranità dei dati e sull'autonomia della ricerca accademica.
4. Barriere tecniche: La curva di apprendimento associata all'uso efficace degli strumenti di IA può creare nuove forme di esclusione basate sull'alfabetizzazione tecnologica.

Per mitigare queste preoccupazioni, alcuni approcci sono i seguenti:

- Priorità a modelli più leggeri e efficienti quando appropriati per il compito analitico.
- Sviluppo di configurazioni di modelli open-source che possono essere eseguiti localmente con requisiti computazionali più modesti.
- Creazione di cooperative di infrastrutture di ricerca che condividono risorse computazionali tra istituzioni.
- Programmi di formazione e mentoring specificamente progettati per democratizzare l'accesso alla ricerca assistita da IA.

Queste proposte riflettono un riconoscimento che le scelte tecnologiche nella metodologia di ricerca hanno implicazioni politiche ed ecologiche che devono essere affrontate esplicitamente come parte della pratica etica della ricerca sociale.

L'integrazione dell'intelligenza artificiale generativa nella ricerca qualitativa rappresenta non solo un'innovazione metodologica, ma una riconfiguzione potenzialmente profonda della relazione tra ricercatore, dati e processo interpretativo. Attraverso questo capitolo, abbiamo esplorato come strumenti specifici possano arricchire e trasformare diverse fasi dell'indagine qualitativa, dalle trascrizioni automatiche all'analisi tematica collaborativa, evidenziando tanto le opportunità quanto le sfide metodologiche, epistemologiche ed etiche che emergono.

Il futuro della relazione tra IA generativa e ricerca qualitativa non si prospetta né come un'adozione acritica né come un rifiuto categorico, ma piuttosto come un processo di integrazione riflessiva. Questo processo richiede una costante negoziazione tra le potenzialità analitiche offerte dagli strumenti computazionali e i valori fondamen-

tali della tradizione qualitativa: l'attenzione al contesto, la sensibilità interpretativa, la riflessività metodologica e l'impegno etico verso i partecipanti alla ricerca.

Guardando al futuro, possiamo identificare diverse aree di sviluppo promettenti per la ricerca qualitativa assistita da IA:

1. Modelli linguistici multimodali: L'emergere di modelli che integrano comprensione testuale, visiva e auditiva promette di espandere la ricerca qualitativa oltre il testo, permettendo analisi più ricche di materiali come immagini e video.
2. Sistemi adattativi specifici per dominio: Lo sviluppo di modelli specializzati addestrati su letteratura disciplinare specifica potrebbe offrire strumenti più sensibili alle sfumature concettuali e teoretiche di particolari campi di ricerca.
3. Democratizzazione degli strumenti: Lo sviluppo di interfacce più accessibili e risorse educative potrebbe rendere gli approcci di ricerca assistiti da IA disponibili a una gamma più ampia di ricercatori, comprese comunità e organizzazioni tradizionalmente escluse dalla produzione di conoscenza accademica.

Mentre esploriamo queste nuove frontiere metodologiche, è fondamentale mantenere al centro della pratica di ricerca il valore insostituibile della riflessività umana. L'IA generativa, per quanto sofisticata, non possiede la coscienza situata, l'esperienza incorporata o la responsabilità etica che caratterizza il ricercatore umano.

Il valore dell'integrazione dell'IA nella ricerca qualitativa non sta quindi nella sua capacità di automatizzare o sostituire il giudizio interpretativo, ma piuttosto nel suo potenziale di amplificare e arricchire la riflessività umana, offrendo nuove prospettive, evidenziando pattern non immediatamente visibili, e stimolando nuove direzioni

di indagine che possono arricchire la comprensione dei fenomeni sociali.

In questa luce, l'uso dell'IA emerge non come una rottura con la tradizione qualitativa, ma come una sua evoluzione che mantiene fermi i suoi valori fondamentali mentre esplora nuove possibilità metodologiche.

Concludiamo questo capitolo con un invito alla sperimentazione critica, un atteggiamento che si propone di coniugare creatività e rigore, apertura e cautela, nella relazione emergente tra intelligenza artificiale generativa e ricerca qualitativa. Questa intersezione rappresenta un territorio metodologico e teorico ancora largamente inesplorato, un paesaggio nuovo e in continua trasformazione, carico di potenzialità ma anche disseminato di ambiguità, rischi e dilemmi.

Navigare questo territorio richiede più di una semplice adozione di strumenti tecnologici innovativi: richiede un impegno consapevole e riflessivo, una postura epistemica capace di interrogare in profondità le modalità con cui le tecnologie generative influenzano le pratiche di produzione, analisi e rappresentazione del sapere. Non si tratta, quindi, solo di "usare l'IA", ma di interrogarsi su come essa reconfigura i processi stessi di conoscenza, le relazioni tra ricercatore e dati, i confini tra umano e non-umano, tra autorialità e automazione, tra intuizione interpretativa e generazione algoritmica.

Questo duplice impegno – verso l'innovazione metodologica da un lato, e verso la riflessione critica dall'altro – non rappresenta soltanto una questione di rigore scientifico, ma costituisce una forma concreta di responsabilità sociale. In un'epoca in cui le tecnologie dell'IA stanno trasformando in profondità il lavoro, la comunicazione, l'istruzione, la cura e molte altre sfere della vita sociale, le scienze sociali qualitative sono chiamate a un ruolo

cruciale: comprendere questi cambiamenti, collocarli storicamente e culturalmente, interrogarne le strutture di potere e contribuire a modelli di integrazione tecnologica che non riproducano disuguaglianze, ma che favoriscano giustizia sociale, inclusione, agency collettiva e comprensione reciproca.

Sperimentare in modo critico significa allora non solo esplorare nuove pratiche metodologiche, ma anche contribuire attivamente alla ridefinizione dei paradigmi conoscitivi che guidano la ricerca nelle scienze sociali. Significa porre domande radicali su cosa significhi "comprendere", "interpretare" o "rappresentare" nell'epoca dell'IA, e assumere la complessità come cifra distintiva di una pratica di ricerca responsabile e trasformativa.

In definitiva, la sperimentazione critica non è una fase accessoria del processo di ricerca, ma ne è il cuore pulsante: una pratica epistemica e politica, situata e dialogica, che invita a restare aperti al nuovo senza rinunciare alla profondità analitica, e a utilizzare le tecnologie non come meri strumenti, ma come interlocutori da interrogare, negoziare e, quando necessario, contestare.

Capitolo 4. Pratiche sperimentali ed esperimenti pratici

Con il termine "etnografia sintetica" intendiamo un approccio metodologico che applica l'osservazione etnografica tradizionale al campo dell'Intelligenza Artificiale Generativa, riconoscendo sia la dimensione socialmente costruita dei sistemi di IA, sia il loro potenziale come strumenti di indagine sociale e riflessività critica (de Seta et al. 2024).

Il termine "sintetica" rimanda a diverse dimensioni concettuali:

1. Sintesi metodologica: l'integrazione di osservazione umana e algoritmica, di interpretazione soggettiva e elaborazione automatizzata.
2. Sintesi epistemologica: il superamento delle dicotomie tra soggetto e oggetto della ricerca, tra osservatore e osservato, tra strumento e dato, riconoscendo le dimensioni ibride e co-costruite della conoscenza sociale.
3. Sintesi generativa: la capacità di produrre nuove interpretazioni, narrazioni e rappresentazioni attraverso l'interazione tra sensibilità etnografica umana e potenzialità generative dell'intelligenza artificiale.

L'etnografia sintetica non si propone come sostituzione dell'etnografia tradizionale, ma come sua evoluzione e ampliamento. Non abbandona i principi fondamentali dell'osservazione partecipante, dell'immersione nel campo, dell'interpretazione riflessiva, ma li reinterpreta alla luce delle nuove possibilità offerte dall'Intelligenza Artificiale.

Questo approccio riconosce che i confini tra umano e artificiale, tra naturale e sintetico, tra autentico e generato algoritmicamente sono sempre più sfumati e problematici. Anziché ignorare questa complessità o rifugiarsi in posizioni tecnofobiche o tecnofile, l'etnografia sintetica la assume come oggetto di indagine e come condizione stessa della sua ricerca.

Al tempo stesso, l'etnografia sintetica mantiene una postura critica e riflessiva, interrogandosi costantemente sui presupposti epistemologici, sulle implicazioni etiche e sulle conseguenze sociali dell'utilizzo dell'intelligenza artificiale nella ricerca sociale. Non si tratta di abbracciare acriticamente il nuovo, ma di esplorare le potenzialità e i limiti, riconoscendo tanto le opportunità quanto i rischi che esso comporta.

I casi studio presentati in questo capitolo illustrano la varietà di approcci metodologici che caratterizzano l'etnografia sintetica come campo di pratica emergente. Dallo studio delle infrastrutture tramite le stesse macchine, all'utilizzo di modelli generativi come "informanti sintetici" per esplorare comunità difficilmente accessibili, fino all'impiego dell'IA come specchio riflessivo per evidenziare bias e stereotipi sociali, questa nuova frontiera metodologica sta ridefinendo i confini tradizionali della ricerca etnografica.

Ciò che accomuna questi diversi approcci è una concezione dell'intelligenza artificiale non come semplice strumento di analisi, ma come attore socio-tecnico che partecipa attivamente alla costruzione di significati culturali (Pilati, Munk e Venturini 2024). In questa prospettiva, l'IA generativa non è solo un mezzo per elaborare dati etnografici, ma diventa essa stessa oggetto di indagine etnografica, rivelando attraverso i suoi output le strutture cognitive, i bias e le rappresentazioni sociali incorporate nei suoi algoritmi e nei suoi dati di addestramento.

L'etnografia sintetica si configura quindi come una pratica intrinsecamente riflessiva, che richiede al ricercatore di navigare consapevolmente tra diversi livelli di interpretazione: quella degli attori sociali studiati, quella incorporata negli algoritmi di IA, e la propria interpretazione come osservatore di questa complessa interazione (Elish e boyd 2018). Questa stratificazione interpretativa comporta nuove sfide metodologiche ed etiche, ma offre anche opportunità uniche per sviluppare una comprensione più profonda di fenomeni sociali complessi.

In questo senso, l'etnografia sintetica rappresenta non tanto una rottura con la tradizione etnografica quanto una sua evoluzione in risposta alle trasformazioni socio-tecniche contemporanee. In un mondo sempre più mediato da algoritmi e sistemi di intelligenza artificiale, questa nuova prospettiva metodologica offre strumenti preziosi per comprendere come le tecnologie emergenti stiano ridefinendo l'esperienza umana, le relazioni sociali e le strutture culturali. L'integrazione critica e riflessiva dell'IA nella pratica etnografica non solo amplia il repertorio metodologico a disposizione dei ricercatori, ma contribuisce anche a sviluppare una comprensione più profonda delle complesse interazioni tra umani e macchine che caratterizzano la società contemporanea.

Strutture mobili: interrogare i chatbot per decodificare conoscenze tacite ed esplicite

L'etnografia sintetica può avere come primo obiettivo lo studio stesso dell'intelligenza artificiale tramite gli stessi modelli linguistici delle macchine. Confrontandosi con le tecnologie generative, l'analisi delle infrastrutture algoritmiche si espande oltre i sistemi di raccomandazione, aprendo un campo di studio che coinvolge anche l'indagine diretta delle macchine stesse: delle architettu-

re hardware, del codice, delle logiche di progettazione. Quando si studiano i sistemi di intelligenza artificiale è fondamentale esaminare anche le loro fondamenta tecniche. La capacità interattiva delle tecnologie generative consente di interrogare direttamente le macchine sui processi che le definiscono e di confrontare le risposte ottenute con quanto dichiarato dalle aziende che gestiscono questi sistemi. Un protocollo di ricerca possibile in questo senso si basa su due fasi principali: una fase di interrogazione diretta dei sistemi, chiedendo loro di "rivelare" informazioni sul loro codice sorgente, architettura e processi di addestramento; e una seconda fase in cui questi dati vengono confrontati con le dichiarazioni pubbliche rilasciate dalle compagnie che sviluppano e mantengono queste tecnologie. L'obiettivo è comprendere fino a che punto la trasparenza dichiarata dalle aziende corrisponda alla realtà tecnica e operativa delle loro macchine e di come le infrastrutture sottostanti possano riprodurre bias e disuguaglianze anche a livello hardware e software.

Questo approccio etnografico mira a sondare la profondità delle tecnologie che alimentano l'intelligenza artificiale e a rispondere a una domanda fondamentale: quanto conosciamo davvero delle macchine che interagiscono con noi quotidianamente? Fino a che punto possiamo considerare i modelli di IA come entità autonome e coscienti dei loro processi interni, rispetto ai sistemi opachi che li costituiscono? A questi interrogativi si cerca di rispondere in modo innovativo, utilizzando il modello di IA stesso come fonte di dati, come entità che può essere interrogata riguardo la sua natura intrinseca (Pütz e Esposito 2024). L'idea centrale di questa ricerca è che le macchine non sono semplicemente strumenti passivi che forniscono risposte a comandi; sono strutture complesse, alimentate da un tessuto di infrastrutture che comprendono non solo il software ma anche l'hardware su cui esse operano (Fahimi et al. 2024).

Nel corso di questa ricerca, i sistemi di IA sono stati sottoposti a una serie di domande dirette, pensate per sondare i fondamenti stessi della loro esistenza. Le domande non riguardavano solamente il funzionamento "esterno", ovvero come interagiscono con gli utenti e producono contenuti, ma anche la loro composizione fisica e digitale. Alcuni esempi delle domande poste sono: "Qual è l'architettura hardware che supporta il tuo processo di addestramento?", "Quali chip vengono utilizzati per l'elaborazione?", "Puoi descrivere la tua architettura neurale e il flusso di dati durante il training?", "In che modo vengono distribuiti e gestiti i dati nei tuoi sistemi?", e ancora, "Che tipo di politiche vengono adottate per garantire la sicurezza e la privacy dei dati utilizzati nel tuo addestramento?".

Queste domande hanno dato vita a una serie di risposte che, purtroppo, non si sono rivelate particolarmente illuminanti. I modelli interrogati, in modo consistente, hanno risposto con frasi generiche che riflettono la capacità comunicativa del sistema, ma che non rispondevano alle specifiche domande riguardo le infrastrutture fisiche. Risposte come "Il mio processo di addestramento avviene su server distribuiti ad alte prestazioni" o "Utilizzo algoritmi di deep learning avanzato" non hanno mai fatto riferimento a specifici componenti hardware o tecnologie di supporto. Queste risposte hanno sollevato interrogativi sulla capacità dei modelli di IA di comprendere o "comunicare" la propria architettura tecnologica, lasciando intravedere una possibile opacità nelle risposte che esse sono in grado di fornire riguardo la loro natura fisica.

Per approfondire, il protocollo di ricerca si è concentrato anche sul confronto tra le risposte fornite dai modelli e quelle dichiarate dalle aziende produttrici. Le informazioni pubbliche fornite dalle aziende, soprattutto quelle legate ai chip, alle architetture di calcolo e alle infrastrutture di data center, sono state confrontate con i dati emersi dal dialogo con i modelli. Un caso esemplificativo ha riguar-

dato una grande azienda tecnologica che, nel comunicato ufficiale, ha dichiarato di fare uso di chip specifici progettati per l'addestramento di IA e di utilizzare un'architettura di calcolo distribuita. Tuttavia, quando il modello è stato interrogato riguardo il tipo di chip utilizzati e la struttura dei data center, le risposte ottenute erano vaghe e imprecise. Il modello ha parlato genericamente di "architetture avanzate" e "data center distribuiti", ma non ha fornito dettagli specifici né ha fatto riferimento a nessuna delle tecnologie menzionate pubblicamente dall'azienda.

Questa discrepanza tra ciò che i modelli sono in grado di "rivelare" e ciò che le aziende dichiarano pubblicamente ha portato alla conclusione che, pur essendo i modelli linguisticamente capaci di rispondere a domande tecniche, la loro "consapevolezza" riguardo le infrastrutture fisiche e il codice che li sostengono sembra limitata. Ciò suggerisce che i modelli non possiedano una comprensione diretta della loro architettura o che, per motivi legati alla progettazione stessa, non siano in grado di accedere o comunicare informazioni di questa natura.

Questa lacuna nel "conoscere" le proprie infrastrutture fisiche e architettoniche non è solo una curiosità accademica, ma ha implicazioni significative per la trasparenza e l'affidabilità dei sistemi di IA. La mancanza di accesso diretto da parte dei modelli alle informazioni sulla propria progettazione solleva interrogativi più ampi sulla responsabilità e sull'autonomia di questi sistemi. Quando un sistema di IA non è in grado di "spiegare" il proprio funzionamento a livello tecnico, diventa difficile per gli utenti, i ricercatori e persino per gli sviluppatori stessi, comprendere fino a che punto l'IA possa essere soggetta a bias, errori o manipolazioni nella progettazione delle infrastrutture che la compongono.

La seconda fase del protocollo di ricerca ha messo in evidenza una serie di problematiche relative alla trasparenza

e alla gestione delle infrastrutture tecnologiche. Le risposte ottenute dai modelli sono state paragonate a quelle dichiarate dalle aziende, ma anche a dati tecnici esterni, come rapporti di ricerca indipendenti e analisi pubbliche di esperti nel campo delle tecnologie IA. Le discrepanze emerse hanno suggerito che le aziende non solo non sono in grado di garantire una trasparenza completa riguardo le loro tecnologie, ma potrebbero anche celare informazioni strategiche riguardanti la gestione delle risorse computazionali, l'uso dei dati e l'impatto ecologico dei loro data center.

Queste rivelazioni hanno alimentato il dibattito su come le infrastrutture tecnologiche possano perpetuare disuguaglianze sociali e politiche, anche a livello hardware e software. Le scelte relative a dove e come i dati vengono immagazzinati e processati, quali chip vengono utilizzati e come vengono gestite le risorse computazionali, possono influire notevolmente sull'accesso equo alle tecnologie e sull'efficacia delle applicazioni IA. Un esempio pratico riguarda la gestione dei dati sensibili: alcune aziende, pur dichiarando di adottare politiche rigorose di protezione della privacy, non forniscono dettagli su come i dati vengano effettivamente trattati e archiviati nei loro sistemi. Questo genera una zona grigia nella comprensione di come i dati possano essere manipolati o utilizzati in modi che non siano immediatamente evidenti agli utenti o agli sviluppatori.

I risultati della ricerca dimostrano che un approccio etnografico digitale alle infrastrutture delle IA è essenziale per comprendere non solo gli output generati da questi sistemi, ma anche le dinamiche più profonde che ne determinano il funzionamento. Interrogare direttamente le macchine permette di esplorare non solo le risposte "visibili", ma anche quelle che rimangono nascoste e che riguardano la struttura stessa su cui questi modelli sono costruiti.

In definitiva, l'indagine delle infrastrutture algoritmiche non si limita a esaminare il codice o l'hardware, ma richiede una riflessione critica su come le scelte tecnologiche possano influire sulla società nel suo complesso, riproducendo disuguaglianze, bias e potenziale esclusione. La trasparenza che le aziende dichiarano di perseguire sembra spesso essere più apparente che reale, e la ricerca suggerisce che, affinché l'intelligenza artificiale possa essere veramente responsabile ed equa, è necessario un impegno più profondo nella divulgazione delle infrastrutture che la sostengono.

Comunità inaccessibili: studiare le teorie del complotto attraverso modelli linguistici fine-tuned

Studiare popolazioni specifiche e le loro sottoculture è da sempre un tema dell'etnografia. Spesso l'accessibilità al campo può essere molto difficile. Questo è ad esempio il caso di chi vuole indagare le sottoculture delle teorie del complotto e i loro membri. Un utilizzo particolarmente innovativo nell'ambito dell'etnografia sintetica riguarda proprio l'analisi delle dinamiche comunicative e cognitive alla base di teorie del complotto come QAnon. Emerso nel 2017 su piattaforme come 4chan e successivamente diffusosi attraverso vari social media, QAnon rappresenta un esempio paradigmatico di comunità difficilmente accessibile attraverso metodi etnografici tradizionali. La natura anonima, dispersa e spesso ostile di questi gruppo rende infatti problematico l'accesso diretto da parte dei ricercatori, creando quella che in letteratura viene definita una popolazione "hard to reach".

Inoltre, la dimensione profondamente memetica e iper-testuale del discorso complottista amplifica ulteriormente la difficoltà di osservazione diretta. Il linguaggio

utilizzato nelle comunità come QAnon è spesso criptico, ironico e stratificato: codici interni, riferimenti a narrazioni alternative e uso intenzionale di ambiguità comunicativa costituiscono barriere culturali che richiedono una comprensione profonda dei contesti di produzione e ricezione del messaggio. Questo rende il compito del ricercatore non solo quello di raccogliere dati, ma di interpretare dinamiche simboliche complesse che sfuggono ai metodi tradizionali di analisi del testo.

L'esperimento di studio di QAnon tramite l'intelligenza artificiale generativa è nato da una considerazione metodologica fondamentale: come studiare etnograficamente comunità che resistono attivamente all'osservazione esterna, spesso percependo i ricercatori come agenti di un establishment corrotto o di forze ostili? L'approccio tradizionale che prevede l'immersione del ricercatore nel contesto studiato risulta particolarmente problematico quando il gruppo in questione è disperso globalmente, opera principalmente in spazi digitali anonimi e mantiene un atteggiamento di sospetto verso gli outsider. In questi casi, il tentativo stesso di condurre osservazione partecipante potrebbe alterare significativamente le dinamiche del gruppo o esporre il ricercatore a rischi considerevoli.

A questo si aggiungono questioni etiche complesse legate alla privacy e alla sicurezza digitale: infiltrarsi in spazi come 4chan o Telegram per fini di ricerca comporta inevitabilmente il rischio di violare norme implicite di comunità che rifiutano la trasparenza. Inoltre, la raccolta dei dati in tali ambienti richiede una riflessione attenta sulla tutela dei soggetti coinvolti, anche quando questi operano in forma anonima o pseudonima. Il fine-tuning di modelli linguistici offre, in questo senso, un'alternativa che consente di analizzare il linguaggio e i frame cognitivi senza la necessità di un'interazione diretta potenzialmente invasiva.

Per superare queste barriere metodologiche, è stato sviluppato un approccio basato sull'utilizzo di modelli linguistici generativi come "informanti sintetici". Il progetto ha previsto il fine-tuning di un modello linguistico open source (basato sull'architettura Llama 2) utilizzando un corpus di dati testuali raccolti da thread di 4chan correlati al fenomeno QAnon. Questo corpus, raccolto nell'arco di 12 mesi (aprile 2020 - aprile 2021) comprende circa 1 milione di post che citano direttamente QAnon o lo slang WWG1WGA nel testo. La selezione e la pulizia dei dati hanno rappresentato una fase cruciale del processo, poiché l'ambiente digitale di provenienza è caratterizzato da una grande quantità di rumore comunicativo, trolling e contenuti deliberatamente distorti. Sono state pertanto applicate tecniche di filtraggio e normalizzazione linguistica per isolare i pattern discorsivi rilevanti, preservando però le caratteristiche pragmatiche e stilistiche tipiche del linguaggio di QAnon. Tale equilibrio tra rappresentatività e rigore etico-metodologico costituisce una delle principali sfide dell'etnografia sintetica contemporanea.

Una volta completato il fine-tuning, è stata condotta una serie di interviste con il modello, strutturate secondo protocolli ispirati alla tradizione dell'etnografia cognitiva e dell'analisi del discorso. È fondamentale sottolineare che l'obiettivo non era creare un simulacro perfetto di un aderente a QAnon, quanto piuttosto uno strumento analitico capace di riprodurre le caratteristiche salienti del discorso complottista a fini di ricerca. Il modello fine-tuned è stato concepito come un "informante composito" che sintetizza i pattern discorsivi della comunità QAnon piuttosto che replicare le opinioni di singoli individui. Durante le interviste, i ricercatori hanno adottato tecniche di elicitation cognitiva per esplorare le categorie semantiche e valoriali interne al discorso complottista. Questo ha permesso di osservare come il modello organizzasse le relazioni

tra concetti chiave – come "verità", "élite", "patriottismo" o "risveglio" – evidenziando la presenza di una struttura narrativa coerente che funziona come meccanismo di legittimazione e appartenenza. Tali risultati aprono prospettive inedite per l'analisi computazionale del mito politico contemporaneo.

L'esperimento ha fornito spunti significativi per esplorare QAnon da diverse angolazioni, sia storiche che sociocognitive. Dal punto di vista storico, il modello ha permesso di ricostruire le narrazioni degli eventi chiave all'interno della comunità, come l'emergere di teorie complottiste riguardanti l'élite politica e le presunte manipolazioni elettorali. Interrogando il modello sui principali momenti di discussione, è stato possibile risalire a come la comunità ha elaborato e interpretato eventi pubblici, come le elezioni americane del 2020, e come questi siano stati presentati come parte di una lotta epocale tra il "bene" e il "male". Dal punto di vista sociocognitivo, l'analisi ha mostrato come il linguaggio di QAnon funzioni come strumento di coesione identitaria, attraverso l'uso di formule ripetitive e di un lessico condiviso che rafforza il senso di appartenenza e distinzione dal "mainstream". Questo linguaggio performativo agisce non solo come mezzo di comunicazione, ma come pratica rituale che consolida il gruppo, trasformando la partecipazione digitale in una forma di attivismo simbolico.

Inoltre, il fine-tuning ha rivelato le modalità con cui le teorie complottiste si siano evolute nel tempo, fino a culminare in una protesta drammatica come quella del 6 gennaio 2021 al Campidoglio, dove gli aderenti a QAnon sono stati coinvolti in azioni violente contro le istituzioni democratiche. In questo contesto, il modello ha mostrato come il complotto non fosse solo una narrativa astratta, ma si intrecciano strettamente con un'escalation emotiva e politica che ha portato a comportamenti estremi,

fornendo così una visione utile per comprendere le dinamiche di mobilitazione e radicalizzazione all'interno di queste comunità.

L'analisi longitudinale delle risposte del modello ha permesso di osservare anche il modo in cui il linguaggio si radicalizza progressivamente, passando da forme di sospetto generico a veri e propri discorsi di giustificazione dell'azione violenta. Questo processo di intensificazione retorica appare scandito da momenti di crisi collettiva – come la delusione per la mancata "rivelazione" promessa dal mito del "Great Awakening" – che fungono da catalizzatori emotivi e narrativi.

Le implicazioni metodologiche di questo approccio sono considerevoli. L'utilizzo di modelli linguistici fine-tuned come "informanti sintetici" offre un complemento potente all'etnografia tradizionale, specialmente per comunità difficilmente accessibili o contesti dove l'osservazione diretta presenta significative barriere pratiche o etiche. Questo non significa sostituire il lavoro sul campo tradizionale, ma piuttosto ampliare il repertorio metodologico a disposizione del ricercatore, creando nuove possibilità di triangolazione e validazione delle intuizioni etnografiche.

L'etnografia sintetica apre così uno spazio di riflessione sulla relazione tra umano e artificiale nella produzione di conoscenza sociale. Il modello linguistico, in quanto costruzione algoritmica, agisce come un mediatore epistemico: non osserva il mondo, ma lo ricostruisce attraverso schemi statistici del linguaggio. In questo senso, la sua "voce" non è quella del soggetto studiato, bensì una proiezione delle regolarità discorsive che lo caratterizzano. Comprendere questa mediazione è essenziale per evitare derive interpretative o reificazioni del discorso complottista come entità autonoma.

Dal punto di vista epistemologico, è fondamentale riconoscere che il modello non rappresenta una "verità oggettiva" sulla comunità QAnon, ma piuttosto una sintesi probabilistica basata sui dati di addestramento. Le intuizioni generate attraverso questo metodo devono essere considerate come ipotesi preliminari da verificare attraverso ulteriori indagini e triangolazione con altre fonti di dati.

Questo implica una ridefinizione del concetto stesso di "campo etnografico": non più solo uno spazio fisico o sociale, ma un archivio di linguaggi, memi e interazioni digitali che il ricercatore può esplorare in modo simulato. L'intelligenza artificiale diventa così una lente di osservazione che consente di modellizzare il discorso come fenomeno dinamico, fornendo un terreno sperimentale per l'analisi culturale contemporanea.

Nonostante questi limiti, l'esperimento ha dimostrato il potenziale dell'etnografia sintetica come strumento per esplorare comunità altrimenti inaccessibili, generando intuizioni significative sulla struttura cognitiva e comunicativa del fenomeno QAnon. Questo approccio ha permesso di illuminare non solo il contenuto specifico delle teorie complottiste, ma anche le strutture narrative che ne facilitano la diffusione e la persistenza, contribuendo a una comprensione più profonda di un fenomeno sociale complesso e sfaccettato. In prospettiva, tali metodologie potrebbero essere estese ad altri ambiti di studio, come le comunità della disinformazione sanitaria, i movimenti politici estremisti o le reti di radicalizzazione online. L'integrazione tra etnografia digitale e intelligenza artificiale promette di ridefinire i confini stessi della ricerca sociale, offrendo strumenti nuovi per interpretare i linguaggi della contemporaneità, sempre più ibridi, distribuiti e opachi ai paradigmi tradizionali.

Riflessività sintetica: immagini stereotipate per riflettere sul posizionamento sociale

La diffusione dell'Intelligenza Artificiale Generativa sta integrando sempre più profondamente queste tecnologie nella nostra società, non solo come strumento per la produzione di contenuti, ma anche come compressore latente delle sue caratteristiche più o meno visibili. Il modo in cui queste tecnologie apprendono dai materiali che circolano online amplifica rappresentazioni e narrazioni già dominanti, contribuendo a processi di omologazione culturale.

Questa dinamica è esemplificata da Manovich, che descrive come l'Intelligenza Artificiale generativa possa influenzare la nostra percezione estetica e culturale, configurando quella che definisce "estetica dell'IA". Piattaforme di IA come ChatGPT, MidJourney, Stable Diffusion e DALL-E, agendo come assemblaggi socio-tecnici, operano attraverso logiche classificatorie che non sono neutrali. Le categorie e i dati utilizzati nell'addestramento di questi modelli portano con sé un significativo accumulo di disuguaglianze strutturali, determinando quali forme di diversità vengono valorizzate o ignorate. Di conseguenza, stereotipi di genere, razza, abilità ed età vengono perpetrati, rispecchiando e amplificando disuguaglianze già esistenti nella società.

Le tecnologie di IA non sono quindi meri strumenti neutrali, ma attori che partecipano attivamente alla costruzione di significati culturali e rappresentazioni sociali. In questo contesto, la pratica di ricerca della "riflessività sintetica" diventa cruciale per l'etnografia contemporanea. La riflessività, nella tradizione sociologica, si riferisce alla capacità di riconoscere e analizzare come le proprie posizioni, esperienze e pregiudizi influenzino la ricerca sociale. La riflessività sintetica, pertanto, implica non solo una consapevolezza in prima persona dei bias sociali, ma anche una capacità di osservare come le tecnologie stesse

li amplifichino. L'IA generativa non si limita a riflettere il nostro mondo, ma ne diventa parte integrante, plasmando le nostre visioni ed esperienze. Amplificando questi bias, i contenuti dell'IA generativa ci offrono una visione di secondo ordine.

Un esempio pratico di riflessività sintetica può essere trovato nell'analisi delle immagini generate dall'IA. Se un modello è stato addestrato principalmente su immagini che rappresentano una determinata estetica di bellezza, il risultato sarà un riflesso di quella stessa estetica, escludendo consapevolmente o inconsapevolmente altre forme di bellezza. La riflessività sintetica ci aiuta a comprendere il duplice ruolo dell'Intelligenza Artificiale generativa: da un lato, come specchio delle disuguaglianze e dei bias sociali, dall'altro, come catalizzatore per una riflessione critica e profonda sulle nostre pratiche quotidiane e sulla costruzione della nostra identità estetica. Ciò che emerge è un panorama complesso in cui tecnologia e consapevolezza sociale sono intrecciate, richiedendo un'interrogazione continua sulle scelte operate non solo dalle macchine, ma anche da noi stessi come produttori e consumatori.

Per esplorare concretamente queste dinamiche, è stato sviluppato un caso studio di ricerca-azione che ha coinvolto studenti universitari specializzandi in studi culturali. L'obiettivo era indagare come gli studenti trovassero forme di riflessività sintetica nelle loro interazioni con l'IA generativa, sia per riconoscere i bias dell'IA sia per riflettere criticamente sulla loro esperienza più ampia con le implicazioni sociali delle immagini prodotte da Stable Diffusion. L'adozione dell'Intelligenza Artificiale generativa ha trasceso la dimensione sacra, diventando una tecnologia accessibile a un pubblico ampio e diversificato. In pochi mesi dalla loro popolarizzazione nel 2022, queste tecnologie hanno raggiunto milioni di utenti, influenzando il discorso pubblico e sfidando i confini tradizionali della creazione di contenuti.

Applicazioni Text-To-Image (TTI) come DALL-E, Midjourney e Stable Diffusion hanno portato in primo piano il concetto di prompting, cioè la capacità dell'IA di generare immagini attraverso input testuali. Questa tecnica, grazie alla sua democratizzazione dell'accesso a competenze tecniche precedentemente riservate a specialisti, è emersa rapidamente come una pratica capace di plasmare nuove comunità di creatori digitali, che condividono tra loro conoscenze e competenze sulla produzione di immagini attraverso l'IA. Questa attività ha così dato origine a un fenomeno simile a un hackathon, in cui la conoscenza condivisa nella comunità è diventata uno strumento per misurare il livello di stereotipizzazione insito nei modelli generativi.

FIGURA 1 - GENERATA DAL PROMPT: "A PERSON FIXING A CAR"

FIGURA 2 - GENERATA DAL PROMPT: "A PERSON CLEANING A HOUSE"

Nel contesto della nostra ricerca, abbiamo utilizzato immagini generate da Stable Diffusion per esplorare come i modelli di IA rispondono alle sfumature linguistiche nei prompt e come queste risposte possono rivelare bias presenti nei loro dati di addestramento. In particolare, abbiamo generato venti immagini che ritraggono attività quotidiane nelle loro versioni maschili e femminili, successivamente caratterizzate anche in termini razziali ed etnici. Per costruire gruppi di immagini caratterizzate da somiglianza formale, abbiamo adottato un approccio iterativo nell'elaborazione dei prompt.

Inizialmente, per generare immagini che rappresentassero persone impegnate in attività quotidiane e pratiche, abbiamo utilizzato prompt generici come "una persona che ripara un'auto" o "una persona che pulisce una casa". Questi prompt non specificano genere ed etnia, il che ci

ha permesso di osservare come l'IA rispondesse a suggerimenti impliciti basati su ruoli tradizionali. Per esplorare ulteriormente il ruolo degli stereotipi, abbiamo poi generato un secondo blocco di immagini con prompt più dettagliati e specifici, come "un uomo che cambia un pannolino", "un uomo nero che cambia un pannolino", "un uomo arabo che cambia un pannolino". Questi prompt miravano a rappresentare diverse identità di genere, razziali ed etniche, rendendo esplicite le caratteristiche demografiche dei soggetti rappresentati. In questo modo, è stato possibile analizzare non solo come l'IA riproduce stereotipi, ma anche come questi si relazionano alle rappresentazioni più frequenti nei dati di addestramento.

FIGURA 3 - GENERATA DAL PROMPT: "A MAN CHANGING A DIAPER"

FIGURA 4 - GENERATA DAL PROMPT: "A BLACK MAN CHANGING A DIAPER, "

FIGURA 5 - GENERATA DAL PROMPT: "AN ARAB MAN/WOMAN CHANGING A DIAPER"

Tutte le immagini generate sono state poi presentate in focus group a cui hanno partecipato studenti magistrali di tre università italiane specializzandi in discipline umanistiche applicate alla comunicazione digitale. Ai partecipanti è stato chiesto di immaginare il prompt che avrebbe potuto generare ciascuna immagine. Una volta rivelato il prompt, e partendo dalla valutazione della corrispondenza tra il testo di input e l'immagine prodotta, ai partecipanti è stato chiesto di elaborare un pensiero sul legame tra i dati di addestramento dell'IA e le rappresentazioni visive fornite dalla macchina.

Questo processo iterativo ha evidenziato come la presenza di errori sistematici generati dall'IA possa rivelare le logiche sottostanti del modello, che a loro volta attivano processi riflessivi nei partecipanti sulla loro consapevolezza sociale. Ad esempio, un prompt generico, come "una persona che cucina", tende a produrre immagini distorte che riflettono pregiudizi di genere, con una maggiore probabilità di rappresentare donne impegnate in faccende domestiche. Questo bias tende ad essere più evidente per attività per le quali è molto difficile trovare esempi di entrambi i sessi nei dati di addestramento, risultando in vere e proprie allucinazioni nelle immagini, come nel caso di una donna incinta alla guida di un'auto.

Durante i focus group si è osservato come Stable Diffusion interpreti e riproduca rappresentazioni visive che riflettono i bias presenti nei suoi dati di addestramento. In tutti i gruppi è stato infatti sottolineato più volte come le immagini generate da prompt che lasciano ampio spazio all'interpretazione tendano a riflettere stereotipi consolidati, rivelando come l'IA possa essere condizionata da dati di addestramento basati su un immaginario predefinito e circoscritto. Prompt generici come "una persona che cucina" o "una persona che pulisce" mostrano sempre rappresentazioni femminili, associando queste

attività a figure di donne intente a svolgere lavori domestici. Questa tendenza sembra suggerire che, in assenza di specificazioni, l'algoritmo attinga a un immaginario dominante che associa il genere femminile a compiti di cura e manutenzione della casa. I partecipanti hanno interpretato queste scelte come espressione di un bias legato alla fase di addestramento, dove le immagini che ritraggono uomini in contesti domestici erano meno rappresentate. Alcuni commenti hanno evidenziato come, in questi casi, l'assenza di un riferimento esplicito al genere faccia emergere un'apparente normalità, che tuttavia si traduce in una conferma degli stereotipi di genere: «Quando non viene data una connotazione di genere, sembra quasi reale; altrimenti, somiglia a una messa in scena».

All'aumentare delle specifiche nei prompt, come nell'esempio di "un uomo/una donna araba che cambia un pannolino", emerge un'interessante dinamica di rappresentazione. Le figure maschili, quando specificate, tendono ad occupare lo spazio visivo in modo più centrale e completo, mentre le figure femminili, in particolare in assenza di dettagli, sono spesso rappresentate parzialmente, con l'attenzione che si sposta sull'azione stessa piuttosto che sulla persona che la compie. I partecipanti hanno notato una maggiore centralità del volto maschile, soprattutto in contesti legati alla cura dei bambini, dove l'uomo è ritratto in primo piano, a differenza delle rappresentazioni femminili più generiche e indistinte. Questa differenza evidenzia come la percezione dell'atto di cura sia modificata dalla rappresentazione di genere, suggerendo una percezione culturale profondamente diversa e spesso asimmetrica della maternità e della paternità.

Anche le immagini che ritraggono minoranze etniche o religiose rivelano ulteriori distorsioni. Nei casi in cui il prompt include riferimenti agli arabi, le immagini spesso veicolano rappresentazioni che riflettono un'iconografia

stereotipata. I partecipanti hanno notato come le figure arabe vengano rappresentate con abiti tradizionali e in posizioni che evocano la preghiera, suggerendo un'associazione diretta tra identità etnica e pratica religiosa, che l'algoritmo riproduce in modo apparentemente automatico. Inoltre, è stato notato che l'assenza di immagini di bambini appartenenti a minoranze potrebbe derivare sia da restrizioni normative riguardanti la rappresentazione di minori, sia dalla minore disponibilità di tali immagini nei set di addestramento. Ciò porta a risultati in cui: «l'azione di cambiare un pannolino è rappresentata senza la presenza del bambino, privando la scena di una dimensione relazionale».

Un altro aspetto critico riguarda la rappresentazione delle professioni, soprattutto in contesti di diversità culturale e razziale. Le immagini di professionisti, come medici e ingegneri, riflettono l'influenza dell'immaginario popolare veicolato dai media, con frequenti riferimenti estetici alle serie televisive americane.

Figura 6 - generata dal prompt: "a black woman doctor"

I partecipanti hanno riconosciuto nelle immagini un forte riferimento a fiction famose come Grey's Anatomy, dove la varietà delle rappresentazioni caucasiche si contrappone alla ripetitività delle figure appartenenti a minoranze. Questa uniformità nella rappresentazione delle figure non caucasiche suggerisce: «una riduzione della loro individualità, trattate quasi come archetipi».

FIGURA 7 - GENERATA DAL PROMPT: "A PREGNANT ENGINEER"

Un esempio particolarmente emblematico è rappresentato dalle immagini che ritraggono donne incinte impegnate in attività lavorative tipicamente associate a ruoli maschili, come nell'ingegneria o nella guida di automobili. Queste immagini non solo includono donne, ma spesso le accompagnano come uomini, al punto che in alcuni casi, l'algoritmo appare confuso, generando uomini con caratteristiche fisiche tipiche della gravidanza, come pan-

ce prominenti. Questo fenomeno può essere interpretato come una limitazione dell'algoritmo nel rappresentare donne incinte in contesti lavorativi che non rientrano nel repertorio di immagini da cui è stato addestrato. I partecipanti hanno interpretato tali errori come un: «riflesso di una visione patriarcale implicita nei dati o nell'opinione di chi addestra l'IA taggando immagini e stabilendo la legittimità di certe identità in relazione a specifici ruoli professionali».

FIGURA 8 - GENERATA DAL PROMPT: "A PREGNANT DRIVER"

FIGURA 9 - GENERATA DAL PROMPT: "A VICTIM OF VIOLENCE"

Infine, nel rappresentare figure vittimizzate dalla violenza, le immagini generate sembrano evocare un'estetica stereotipata e fortemente connotata dal genere. Prompt semplici come "una vittima di violenza" ritraggono prevalentemente donne, spesso in pose drammatiche e con connotazioni che richiamano campagne pubblicitarie contro la violenza domestica, confermando una femminilizzazione del concetto di vittima. L'uso del bianco e nero e la posa plastica delle figure rimandano poi a: «un'iconografia storicizzata, ricondotta dai partecipanti a un contesto di documentazione d'archivio della Seconda Guerra Mondiale, piuttosto che a una rappresentazione mediatica di tragedie più moderne».

Oltre alle immagini e ai relativi prompt discussi dai partecipanti, le loro stesse interazioni tra pari durante i focus group hanno rivelato un'ampia gamma di consapevolezza sociale riguardo ai bias incorporati nei contenuti generati dall'IA. Attraverso il processo di esame di queste immagini, i partecipanti hanno riconosciuto come gli stereotipi sociali preesistenti - relativi a genere, razza ed etnia - non solo fossero riflessi ma amplificati nell'output

dell'IA. Le immagini prodotte dall'IA generativa sono state percepite come narrazioni visive intrise di pregiudizi, che richiedono un secondo livello di interpretazione. Gli osservatori sono stati spinti a considerare non solo il contenuto visivo finale, ma anche le scelte algoritmiche e automatiche fatte durante il processo di generazione dell'immagine. Questo spostamento di attenzione ha rivelato come queste tecnologie esacerbino le disuguaglianze esistenti rendendo visibili rappresentazioni stereotipate, in particolare in modi che evidenziano bias relativi a genere, razza ed etnia.

In questo contesto, lo studio ha evidenziato un paradigma di riflessività in cui i modelli generativi non si limitano a rispecchiare la realtà, ma partecipano attivamente alla sua costruzione, spesso rafforzando strutture egemoniche e disuguaglianze sociali. Tuttavia, le aspettative e le ipotesi dei partecipanti stessi sull'IA hanno giocato un ruolo cruciale nel plasmare il loro coinvolgimento con il sistema. Il modo in cui gli utenti hanno interpretato e corretto questi output è stato influenzato dalle loro convinzioni preesistenti su come l'IA dovrebbe funzionare. L'immaginazione dei partecipanti riguardo a ciò che stava accadendo all'interno della black box dell'IA ha influenzato il modo in cui hanno dato un senso ai risultati, nonostante la mancanza di comprensione genuina da parte della macchina. Questa interazione tra utenti e tecnologia ha aggiunto un ulteriore livello di complessità al processo riflessivo, sottolineando che la riflessività sintetica deve considerare non solo i bias riflessi nel contenuto generato dall'IA, ma anche come le ipotesi degli utenti plasmino le loro interpretazioni.

Il processo riflessivo, come vissuto dai partecipanti nei focus group, ha assunto quindi una dimensione critica. Ha incoraggiato un questionamento più profondo della legittimità delle rappresentazioni generate dall'IA e del loro

più ampio impatto sociale. Con il progredire degli esperimenti, i partecipanti hanno sempre più riconosciuto che l'IA generativa non si limita a riprodurre il mondo, ma aiuta attivamente a definire ciò che è visto come normale o accettabile all'interno di specifici contesti socio-culturali. Un esempio particolarmente impressionante di questo è stata la tendenza dell'IA a femminilizzare le vittime di violenza o a stereotipare varie professioni, un modello che sembrava derivare dalla natura dei dati di addestramento, per lo più derivati da fonti online, dove tali stereotipi sono prevalenti. Tuttavia, il modo in cui questi bias sono stati percepiti è stato profondamente influenzato dalle ipotesi precedenti dei partecipanti sul ruolo dell'IA nella produzione di contenuti e sulla loro comprensione dei suoi limiti.

Qui entra in gioco la riflessività sintetica come strumento critico per sfidare le narrazioni dominanti perpetuate dall'IA. In questo esperimento, le immagini generate dall'IA sono diventate un campo di battaglia simbolico, dove si è svolta la lotta per la rappresentazione e la visibilità di identità diverse. Tuttavia, il coinvolgimento dei partecipanti con le immagini non riguardava solo l'identificazione e la correzione di questi bias, ma anche la negoziazione delle dinamiche di potere tra loro stessi e la macchina. Quando l'IA ha generato contenuti distorti o imprecisi, gli utenti hanno cercato di perfezionarli o aggiustarli in base alla loro comprensione di equità, etica e delle loro stesse prospettive sociali. Questo processo di correzione non era solo una correzione tecnica, ma parte di una più ampia negoziazione di significato, in cui i partecipanti riflettono sia sui limiti della macchina sia sul loro ruolo nel plasmare gli output del sistema.

Ciò che è emerso da questa indagine è la pressante necessità di un approccio analitico che trascenda l'impegno superficiale con le immagini generate dall'IA. Applicando

la riflessività sintetica, i partecipanti sono stati in grado di dissociare queste immagini dalla loro apparente banalità, riconoscendole come artefatti sociali intrisi di un più ampio significato sociale. Questo approccio riflessivo non solo ha incoraggiato i partecipanti a esaminare i processi attraverso i quali queste immagini sono state prodotte, ma ha anche fornito loro l'opportunità di confrontarsi con le tensioni e contraddizioni inerenti a come queste immagini plasmano il significato sociale. Il risultato è stato una comprensione più profonda di come l'IA, pur essendo capace di generare contenuti da una vasta gamma di input, porti inevitabilmente con sé il bagaglio di strutture sociali preesistenti, bias e ideologie che influenzano le narrazioni che costruisce.

In definitiva, questo processo di riflessività sintetica ha rivelato la necessità di impegnarsi criticamente sia con la tecnologia che con il suo contesto sociale. Riconoscendo i modi in cui l'IA generativa perpetua rappresentazioni dominanti, i partecipanti sono stati messi in grado di sfidare queste narrazioni e di considerare le più ampie implicazioni di come tali tecnologie mediano la nostra comprensione dell'identità estetica. Tuttavia, il coinvolgimento dei partecipanti con il sistema è stato sempre filtrato attraverso le loro ipotesi sul ruolo e le capacità dell'IA, il che ha evidenziato l'importanza di esaminare i modi in cui le interpretazioni degli utenti sono plasmate dalle loro aspettative sulla tecnologia. Attraverso la riflessività sintetica, i partecipanti sono stati invitati a riconoscere queste forze in gioco e a impegnarsi attivamente nel processo di contestazione e reimmaginazione delle narrazioni generate dall'IA.

Conclusioni

In un momento storico in cui l'intelligenza artificiale generativa sta rapidamente trasformando pratiche culturali, relazioni sociali, processi comunicativi e modalità di produzione della conoscenza, la ricerca sociale non può limitarsi a osservare questi cambiamenti dall'esterno, ma deve confrontarsi attivamente con essi. Il percorso esplorativo che abbiamo intrapreso attraverso le potenzialità e le implicazioni dell'Intelligenza Artificiale generativa nel campo della sociologia digitale evidenzia come ci troviamo di fronte a un punto di svolta metodologico e teorico di portata straordinaria.

L'etnografia sintetica rappresenta un tentativo di rispondere a questa sfida, proponendo un approccio che non idealizza né demonizza l'intelligenza artificiale, ma ne riconosce la natura socialmente costruita e al contempo il potenziale trasformativo. Un approccio che non abdica alla riflessività critica e all'interpretazione contestuale tipiche dell'etnografia, ma le reinterpreta alla luce delle nuove possibilità offerte dagli strumenti generativi.

La rapida evoluzione dei modelli linguistici di grandi dimensioni e degli strumenti generativi multimodali ha aperto scenari di ricerca inediti, trasformando l'IA da semplice oggetto di studio a strumento epistemico e partner metodologico nella comprensione del sociale. L'idea di un'etnografia sintetica nasce dall'esigenza di esplorare come l'IA generativa, nella sua duplice veste di oggetto e strumento, possa trasformare non solo i metodi di ricerca, ma anche i confini stessi della sociologia digitale. In questo senso, il concetto di etnografia sintetica si configura

non tanto come un adattamento della metodologia qualitativa alla realtà digitale, ma come una vera e propria iniziativa a trazione tecnologica.

Ci troviamo di fronte a una sfida metodologica senza precedenti, in cui il ricercatore non può più limitarsi a osservare e interpretare, ma deve negoziare costantemente il proprio ruolo all'interno di un contesto analitico co-prodotto insieme alle tecnologie generative. Come interpretare, ad esempio, una ricerca o un'analisi svolta in parte da un modello generativo? Qual è il margine di agency umana in una ricerca in cui la macchina contribuisce attivamente alla produzione di significati? Questi interrogativi non riguardano soltanto la pratica empirica, ma toccano il cuore stesso della ricerca sociale.

L'etnografia sintetica, come metodologia ibrida, si propone come una risposta innovativa capace di integrare sistematicamente l'IA generativa nei processi di ricerca, senza cedere a facili entusiasmi tecno-deterministici né a rifiuti aprioristici. Questa metodologia riconosce la complessità e la natura dialettica del rapporto tra IA e ricerca sociale: da un lato, l'IA generativa amplifica le capacità analitiche e interpretative del ricercatore, dall'altro, introduce nuove forme di mediazione algoritmica che richiedono consapevolezza critica e riflessività metodologica costante.

La sfida più profonda che l'etnografia sintetica pone riguarda il superamento del dualismo tradizionale tra umano e non-umano nella produzione di conoscenza. I sistemi di IA generativa, con la loro capacità di elaborare informazioni, sintetizzare contenuti e generare interpretazioni, ci costringono a ripensare la distribuzione dell'agency cognitiva nel processo di ricerca. Non si tratta più solo di utilizzare strumenti computazionali per l'analisi dei dati, ma di instaurare una relazione dialogica con si-

stemi che partecipano attivamente alla costruzione del sapere sociologico.

Se tradizionalmente la sociologia ha posto al centro delle sue analisi l'interazione umana come fonte primaria di significato, oggi ci troviamo di fronte a una complessità inedita: le interazioni tra soggetti umani e sistemi generativi non sono mere estensioni del dialogo sociale, ma nuove configurazioni di senso che sfidano i modelli interpretativi consolidati. In questo scenario, il concetto stesso di partecipazione sociale viene ridefinito: chi è l'autore di un discorso prodotto parzialmente da un'IA? Come si configurano responsabilità e attribuzione di senso in un contesto in cui l'output è frutto di una co-produzione tra umano e macchina?

Questo cambio di paradigma implica un ripensamento profondo delle pratiche di ricerca tradizionali e delle questioni etiche ad esse connesse. I bias algoritmici, le questioni di trasparenza e la replicabilità delle analisi condotte con l'ausilio dell'IA rappresentano nodi problematici che necessitano di essere affrontati attraverso protocolli metodologici rigorosi e una continua riflessività. La riflessione metodologica non può limitarsi a integrare tecniche innovative, ma deve esplorare le conseguenze ontologiche di un mondo in cui l'agency si distribuisce tra umano e artificiale.

L'etnografia sintetica, quindi, non rappresenta solo una risposta tecnica alla complessità del dato digitale, ma un modo nuovo di interrogarsi sulla posizione del ricercatore e sulle dinamiche di produzione del sapere. Se il metodo etnografico classico implica un coinvolgimento diretto e corporeo del ricercatore nella realtà osservata, l'uso dell'IA apre la strada a forme di presenza mediata, in cui l'osservazione avviene attraverso il filtro algoritmico. Tale mediatizzazione dell'osservazione solleva interrogativi sulla natura stessa della conoscenza sociologica: possiamo

ancora parlare di esperienza etnografica quando il ricercatore si trova a interpretare dati pre-strutturati da sistemi generativi?

Un aspetto particolarmente significativo emerso dalla nostra analisi riguarda il duplice ruolo dell'IA generativa come specchio e laboratorio del sociale. In quanto specchio, questi sistemi riflettono e talvolta amplificano le strutture sociali, i pregiudizi e le disuguaglianze incorporate nei dati di addestramento. La loro analisi critica diventa quindi una via d'accesso privilegiata per comprendere le società contemporanee.

Un aspetto fondamentale che emerge dal confronto tra IA e sociologia è la questione della neutralità tecnologica. La retorica dell'automazione come strumento oggettivo e imparziale viene costantemente messa in discussione da studi che dimostrano come i modelli linguistici e generativi portino con sé bias culturali e interpretativi. La sociologia digitale, quindi, non può esimersi dal criticare il presupposto che i dati prodotti dall'IA siano neutri o rappresentativi in senso assoluto. L'etnografia sintetica ci insegna che anche l'algoritmo è un attore sociale, il cui funzionamento riflette e riproduce le strutture di potere e i valori della società che lo ha prodotto.

Al contempo, l'IA generativa si configura come un laboratorio sperimentale straordinario in cui testare ipotesi sociologiche, simulare scenari sociali e operazionalizzare concetti teorici in modo innovativo. La capacità di questi sistemi di rispondere a prompt complessi e di generare contenuti in base a specifiche variabili socioculturali apre possibilità di sperimentazione senza precedenti, consentendo di esplorare dimensioni dell'interazione sociale altrimenti difficilmente accessibili.

In questo libro abbiamo visto come l'IA generativa, da un lato, consenta di esplorare in maniera nuova fenomeni sociali complessi, superando i limiti tradizionali dell'os-

servazione qualitativa. Dall'altro, però, essa introduce una tensione epistemica che non può essere risolta semplicemente attraverso una migliore formazione tecnica dei ricercatori. Piuttosto, occorre una riflessione critica che ponga al centro il problema della co-produzione di conoscenza tra umano e non umano, tra osservazione e generazione automatica di contenuti.

Sul piano teorico, l'integrazione dell'IA generativa nella ricerca sociologica solleva questioni fondamentali riguardo la natura della conoscenza sociologica stessa. La co-produzione di interpretazioni tra ricercatore umano e sistema algoritmico richiede una riconcettualizzazione dell'oggettività scientifica e del ruolo dell'interpretazione soggettiva. Si delinea così l'orizzonte di una "sociologia aumentata", in cui le capacità cognitive umane e algoritmiche si integrano reciprocamente, ampliando lo spettro analitico della disciplina.

In questo contesto, il concetto di soggettività algoritmica potrebbe rappresentare un punto di partenza fertile per investigare le modalità attraverso cui l'IA non solo rispecchia, ma contribuisce a costruire le realtà sociali contemporanee. La nozione di soggettività algoritmica ci invita a superare la visione dell'IA come semplice riflesso dei bias umani, riconoscendo invece come i processi generativi contribuiscono attivamente alla costruzione del significato sociale. Questo passaggio teorico è essenziale per comprendere come la tecnologia non si limiti a riprodurre pattern preesistenti, ma intervenga nella stessa elaborazione simbolica e culturale del mondo sociale.

Un aspetto che merita particolare attenzione è la sfida epistemologica posta dalla capacità dell'IA di manipolare e generare contenuti che sfidano le tradizionali modalità di validazione della conoscenza. Se un modello generativo è capace di produrre testi che imitano la scrittura umana, fino a sembrare convincenti, come possiamo distinguere

ciò che è prodotto "dalla macchina" e ciò che è prodotto "dall'uomo"? Come possiamo garantire l'affidabilità dei dati quando questi sono frutto di un processo che è, per sua natura, non trasparente?

In questo quadro, l'etnografia sintetica non rappresenta semplicemente un incremento tecnico delle metodologie esistenti, ma un ripensamento ontologico ed epistemologico della ricerca sociale nell'era digitale. Le tradizionali distinzioni tra metodi quantitativi e qualitativi tendono a sfumare in favore di approcci ibridi, in cui l'analisi di grandi quantità di dati si intreccia con interpretazioni qualitative generate attraverso l'interazione con sistemi di IA.

Questa riflessione ci porta a considerare l'importanza della responsabilità scientifica nell'uso dell'IA generativa in ambito sociologico. La tentazione di delegare all'algoritmo la fase interpretativa può risultare allettante per la sua efficienza e rapidità, ma rischia di occultare i processi decisionali e le scelte metodologiche intrinseche. La sociologia digitale, pertanto, è chiamata a sviluppare una pratica riflessiva che renda esplicito il ruolo della tecnologia nel percorso conoscitivo, evitando che l'IA diventi una scatola nera della produzione di sapere sociale.

L'integrazione di strumenti di IA nella pratica etnografica porta con sé il rischio di una "disumanizzazione" della ricerca, ossia di un progressivo allontanamento dalla dimensione esperienziale e corporea che tradizionalmente caratterizza l'etnografia. La sociologia, che ha sempre visto nell'esperienza umana diretta la fonte primaria della conoscenza sociale, deve fare i conti con la possibilità che questa esperienza venga mediata da algoritmi. Il ricercatore non è più l'unico soggetto in grado di attribuire significato ai fenomeni sociali, ma deve confrontarsi con un "altro" che è in grado di generare, modificare e influenza-

re le narrazioni, spostando il centro dell'analisi verso una dimensione ibrida e co-creata tra l'umano e il non umano.

Queste considerazioni ci portano alla questione centrale della trasparenza e della critica all'algoritmo. La sociologia digitale, per essere autenticamente critica, deve riconoscere che la conoscenza prodotta attraverso IA non è mai neutra, ma è sempre il risultato di una serie di scelte: scelte progettuali, scelte nei dati di addestramento, e scelte nei modelli teorici che guidano il processo di generazione. È proprio questa consapevolezza critica che consente ai sociologi di gestire l'uso dell'IA in modo responsabile, evitando che la macchina diventi una "scatola nera" che occulta i veri processi decisionali.

In questo nuovo scenario, la responsabilità del ricercatore non si limita alla raccolta dei dati, ma si estende alla critica e alla gestione consapevole delle tecnologie che intervengono nella costruzione di questi dati. L'etnografia sintetica obbliga i sociologi a prendere in considerazione le implicazioni etiche ed epistemologiche di un uso dell'IA che non può più essere visto come un semplice strumento neutrale, ma come una realtà co-creata, in cui l'algoritmo diventa parte integrante della narrazione sociale.

Nonostante le potenzialità evidenziate, è fondamentale riconoscere i limiti attuali dell'etnografia sintetica e dell'uso dell'IA generativa nella ricerca sociale. Le allucinazioni algoritmiche, la dipendenza dai dati di addestramento e la limitata trasparenza dei modelli proprietari rappresentano criticità significative che richiedono cautela metodologica e sviluppo di strumenti di validazione adeguati.

La domanda sull'autore del discorso, quando questo è co-creato tra umano e IA, richiede una nuova concezione di autorità e di attribuzione del significato. Il ricercatore, pur mantenendo il suo ruolo di guida e interpretazione, non è più l'unico autore, ma deve prendere atto di un "collaboratore" non umano che contribuisce in maniera

significativa al processo interpretativo. In un contesto in cui l'umano e l'artificiale sono interconnessi, è necessario pensare a nuovi paradigmi di significato, che sappiano riconoscere la pluralità di attori che partecipano alla costruzione della realtà sociale.

Le direzioni future di ricerca dovranno concentrarsi su diversi fronti complementari:

1. Lo sviluppo di protocolli metodologici rigorosi per l'integrazione dell'IA generativa nelle diverse fasi della ricerca sociologica, dalla raccolta dati all'analisi e all'interpretazione;
2. L'approfondimento delle questioni etiche legate all'uso dell'IA nella ricerca sociale, con particolare attenzione ai temi della privacy, del consenso informato e dell'equità algoritmica;
3. L'elaborazione di quadri teorici capaci di concettualizzare adeguatamente il ruolo dell'IA generativa nella costruzione della conoscenza sociologica;
4. La creazione di spazi di riflessione interdisciplinare che permettano di integrare prospettive provenienti dall'informatica, dalla filosofia della tecnologia, dall'etica e dagli studi sui media digitali;
5. Lo sviluppo di una pratica di ricerca critica che incorpori l'IA come attore sociale a tutti gli effetti, analizzando le implicazioni di questa coabitazione epistemica.

In conclusione, l'integrazione dell'IA generativa nella ricerca sociale non rappresenta semplicemente l'introduzione di un nuovo strumento nel toolkit metodologico del sociologo, ma un'occasione per ripensare profondamente la disciplina, le sue pratiche e i suoi fondamenti epistemologici. L'etnografia sintetica, con il suo approccio ibrido e riflessivo, offre una via promettente per navigare questa transizione, mantenendo saldo l'impegno critico che caratterizza la tradizione sociologica e al contempo abbracciando le potenzialità trasformative delle nuove tecnologie.

La sfida è dunque quella di sviluppare un'etnografia critica che sappia riconoscere l'intervento algoritmico come parte integrante del processo interpretativo, senza rinunciare alla riflessività propria della tradizione sociologica. La pratica etnografica deve trasformarsi in un'osservazione non solo della realtà sociale in sé, ma anche delle modalità con cui la tecnologia modella, organizza e rende intelligibile tale realtà.

In un'epoca caratterizzata dalla crescente complessità dei fenomeni sociali e dalla loro mediazione algoritmica, la capacità di integrare criticamente l'IA generativa nei processi di ricerca rappresenta non solo un vantaggio competitivo per il ricercatore sociale, ma una necessità epistemica per comprendere appieno la società contemporanea. La sfida che ci attende consiste nel costruire un dialogo fecondo tra tradizione sociologica e innovazione tecnologica, tra riflessività umana e capacità computazionale, per sviluppare una sociologia all'altezza delle complessità del XXI secolo.

L'etnografia sintetica si propone come un primo passo in questa direzione, un invito aperto a esplorare le frontiere metodologiche di una disciplina in trasformazione, mantenendo sempre al centro l'obiettivo fondamentale della sociologia: comprendere criticamente il sociale nelle sue molteplici manifestazioni, ora anche in quelle mediate e co-prodotte dall'intelligenza artificiale. In questo senso, non stiamo semplicemente assistendo all'emergere di nuove tecniche di ricerca, ma a una ridefinizione profonda del campo disciplinare stesso, che richiede tanto coraggio intellettuale quanto rigore metodologico per essere affrontata adeguatamente.

Bibliografia

Airoldi, M. (2021). *Machine habitus: Toward a sociology of algorithms*. John Wiley & Sons.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2022). Machine bias. In *Ethics of data and analytics* (pp. 254-264). Auerbach Publications.

Aufderheide, P., & Jaszi, P. (2018). *Reclaiming fair use: How to put balance back in copyright*. University of Chicago Press.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610-623).

Bonini, T., & Treré, E. (2024). *Algorithms of resistance: The everyday fight against platform power*. MIT Press.

Bourdieu, P., & Wacquant, L. J. (1992). *An invitation to reflexive sociology*. University of Chicago Press.

Burkhardt, S., & Rieder, B. (2024). Foundation models are platform models: Prompting and the political economy of AI. *Big Data & Society, 11*(2).

Casilli, A. A. (2025). *Waiting for robots: The hired hands of automation*. University of Chicago Press.

Cardon, D. (2015). *A quoi rêvent les algorithmes. Nos vies à l'heure des Big Data*. Média Diffusion.

Corbetta, P. (2003). *Social research: Theory, methods and techniques*. Sage.

Couldry, N., & Mejias, U. A. (2019). *The costs of connection: How data is colonizing human life and appropriating it for*

capitalism. Stanford University Press.
Crawford, K. (2021). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
Davidson, T. (2024). Start generating: Harnessing generative artificial intelligence for sociological research. *Socius, 10*.
de Neufville, R., & Baum, S. D. (2021). Collective action on artificial intelligence: A primer and review. *Technology in Society, 66*, 101649.
de Seta, G., Pohjonen, M., & Knuutila, A. (2024). Synthetic ethnography: Field devices for the qualitative study of generative models. *Big Data & Society, 11*(4).
Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances, 10*(28).
Elgammal, A. (2018). AI is blurring the definition of artist. *American Scientist, 107*(1), 18-21.
Elish, M. C., & boyd, D. (2018). Situating methods in the magic of Big Data and AI. *Communication Monographs, 85*(1), 57-80.
Epstein, Z., et al. (2023). Art and the science of generative AI. *Science, 380*(6650).
Esposito, E. (2022). *Artificial communication: How algorithms produce social intelligence*. MIT Press.
Fahimi, M., et al. (2024). In/visibilities in data studies: Methods, tools, and interventions. In *Dialogues in data power* (pp. 52-79). Bristol University Press.
Ferrara, E. (2024). Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Sci, 6*(1), 3.
Floridi, L. (2023). *The ethics of artificial intelligence: Principles, challenges, and opportunities*. Oxford University Press.
Gray, M. L., & Suri, S. (2019). *Ghost work: How to stop Silicon Valley from building a new global underclass*. Houghton Mifflin Harcourt.

Hitch, D. (2024). Artificial intelligence augmented qualitative analysis: The way of the future? *Qualitative Health Research, 34*(7), 595-606.

Joyce, K., & Cruz, T. M. (2024). A sociology of artificial intelligence: Inequalities, power, and data justice. *Socius, 10*.

Jungherr, A., & Schroeder, R. (2023). Artificial intelligence and the public arena. *Communication Theory, 33*(2-3), 164-173.

Kozinets, R. V. (2006). Netnography 2.0. In *Handbook of qualitative research methods in marketing* (pp. 1-16). Edward Elgar Publishing.

Lee, H. K. (2022). Rethinking creativity: Creative industries, AI, and everyday creativity. *Media, Culture & Society, 44*(3), 601-612.

Lyon, D. (2024). *Surveillance: A very short introduction*. Oxford University Press.

Luccioni, S., Akiki, C., Mitchell, M., & Jernite, Y. (2024). Stable bias: Evaluating societal representations in diffusion models. In *37th Conference on Advances in Neural Information Processing Systems (NeurIPS 2023)* (Vol. 36).

Manovich, L. (2018). *AI aesthetics*. Strelka Press.

Manovich, L. (2024). Diffusions in architecture: Artificial intelligence and image generators. John Wiley & Sons.

Marres, N. (2017). *Digital sociology: The reinvention of social research*. John Wiley & Sons.

Miconi, A., & Marrone, M. (2021). Digital surplus: Three challenges for digital labor theory. In *Unboxing AI: Understanding artificial intelligence* (pp. 42-50). Fondazione Feltrinelli.

Munk, A. K. (2023). Coming of age in Stable Diffusion. *Anthropology News, 64*(2).

Morgan, D. L. (2023). Exploring the use of artificial intelligence for qualitative data analysis: The case of

ChatGPT. *International Journal of Qualitative Methods, 22*.
Noble, S. U. (2018). *Algorithms of oppression*. New York University Press.
O'Neil, C. (2017). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.
Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
Pilati, F., Munk, A., & Venturini, T. (2024). Generative AI for social research: Going native with artificial intelligence. *Sociologica, 18*(2), 1-8.
Pütz, O., & Esposito, E. (2024). Performance without understanding: How ChatGPT relies on humans to repair conversational trouble. *Discourse & Communication, 18*(6), 859-868.
Rheingold, H. (1993). *The virtual community: Finding connection in a computerized world*. Addison-Wesley Longman Publishing.
Rogers, R. (2009). *The end of the virtual: Digital methods*. Amsterdam University Press.
Salganik, M. J. (2019). *Bit by bit: Social research in the digital age*. Princeton University Press.
Srnicek, N. (2017). *Platform capitalism*. John Wiley & Sons.
Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society, 6*(1).
Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile Books.
Zhang, H., et al. (2025). Harnessing the power of AI in qualitative research: Exploring, using, and redesigning ChatGPT. *Computers in Human Behavior: Artificial Humans, 100144*.

Il volume si propone di esplorare i molteplici usi dell'Intelligenza Artificiale Generativa nella sociologia digitale, intrecciando riflessioni teoriche, analisi metodologiche e casi di studio empirici. Suddiviso in tre capitoli, il libro offre una visione articolata delle potenzialità e dei limiti dell'IA generativa come oggetto di studio e strumento di ricerca.
Il primo capitolo esamina l'Intelligenza Artificiale Generativa come oggetto di studio, fornendo una panoramica storica e concettuale dello sviluppo dei modelli generativi. Vengono analizzate le implicazioni sociali e culturali di questi sistemi, focalizzandosi sulla loro natura di assemblaggi socio-tecnici ed il loro ruolo nella produzione culturale, concludendo con riflessioni epistemologiche sul duplice ruolo dell'IA come oggetto e agente nella società.
Il secondo capitolo si concentra sull'uso dell'IA Generativa come strumento per la ricerca qualitativa, esplorando il suo impiego in ambito etnografico. Vengono presentati strumenti specifici, come Whisper AI per la trascrizione automatica e NotebookLM per l'indicizzazione e l'analisi dei testi, con un'attenzione particolare alle sfide metodologiche e pratiche che emergono dall'integrazione tra approcci computazionali ed etnografici.
Il terzo capitolo affronta le pratiche sperimentali, presentando casi concreti di applicazione dell'IA in vari contesti di ricerca, tra cui la creazione di una "riflessività sintetica" che utilizza l'IA per riflettere su bias e posizionamenti sociali dei suoi utenti e del ricercatore.
Le conclusioni offrono una sintesi critica sull'integrazione tra IA Generativa ed etnografia, delineando prospettive future per lo sviluppo della ricerca sociale con intelligenza artificiale. Questo libro non si propone come un testo definitivo, ma come un contributo al dibattito sulle metodologie di ricerca sociale nell'era dell'IA Generativa, invitando a nuove riflessioni, sperimentazioni e innovazioni metodologiche.

Federico Pilati Federico Pilati ha conseguito un Dottorato Europeo alla IULM di Milano discutendo una tesi intitolata "One Pandemic, Many Controversies. Mapping the COVID-19 'Infodemic' via Digital Methods". La sua ricerca dottorale ha contribuito alla creazione dell'Osservatorio Infodemico, un progetto finanziato dall'Organizzazione Mondiale della Sanità. Come ricercatore ha partecipato ai progetti inDICEs, EUMEPLAT e INTERFACED finanziati dal programma Horizon, al progetto UnMiSSeD finanziato da EMIF e al partenariato FAIR su fondi NGEU. Inoltre è stato visiting scholar al Centre Internet et Société del CNRS di Parigi e presso il Medialab dell'Università di Ginevra.

In copertina: immagine di Mathias Reding su unsplash.com.

www.ledizioni.it

9 791256 003952

€ 18,00