



Human&Future

# *Life-on-life*: l'umanesimo di fronte alle sfide dell'intelligenza artificiale

a cura di Mario Cosenza e Luca Lo Sapia



Federico II University Press



fedOA Press



Human&Future



Human&Future

*Direttore*

Mario Cosenza – Università degli Studi di Napoli Federico II

*Comitato scientifico*

Francesco Celentano – Université de Lausanne

Massimiliano Delfino – Università degli Studi di Napoli Federico II

Roberto Delle Donne – Università degli Studi di Napoli Federico II

Colas Duflo – Université Paris Nanterre

Gianluca Giannini – Università degli Studi di Napoli Federico II

Clelia Iasevoli – Università degli Studi di Napoli Federico II

Ignas Kalpokas – Vytautas Magnus University, Kaunas (Lithuania)

Graziano Lingua – Università degli Studi di Torino

Luca Lo Sapio – Università degli Studi di Torino

Niina Meriläinen – Tampere University (Finland)

Lucio Pastore – Università degli Studi di Napoli Federico II

Antonio Pescapè – Università degli Studi di Napoli Federico II

Geminello Preterossi – Università degli Studi di Salerno

Antonino Squillace – Università degli Studi di Napoli Federico II

Simona Venezia – Università degli Studi di Napoli Federico II

Andrea Zhok – Università degli Studi di Milano Statale

*Comitato di redazione*

Alessia Botta – Università degli Studi di Napoli Federico II

Roberto Corso – Università degli Studi di Napoli Federico II

Alessia Gifuni – Università degli Studi di Napoli Federico II

Mattia Papa – Università degli Studi di Napoli Federico II

Valerio Specchio – Università degli Studi di Napoli Federico II

La collana *Human and Future (H&F)* intende accogliere pubblicazioni legate ai lavori dell'omonima Task Force d'Ateneo. Quest'ultima accoglie studiosi di elevata qualificazione e a forte carattere interdisciplinare per promuovere attività di ricerca, di studio, di riflessione ed elaborazione, di relativo trasferimento di conoscenze e competenze sui temi connessi al futuro dell'umano. La collana *H&F* vuole essere anzitutto l'occasione per unire e combinare le competenze di umanisti e scienziati che si occupano direttamente di questioni quali bio-ingegneria, bio-informatica, AI, Industria 4.0, domotica etc., bio-politica, digitalizzazione, sicurezza. Una collana che sia un luogo d'incontro e fusione di saperi e metodologie, distinti dalle vecchie dinamiche di riunione-confronto-sintesi tra ambiti disciplinari separati. Crediamo che la sfida posta dalla transizione umana renda necessaria la co-elaborazione di codici e linguaggi (anche comunicativi e divulgativi) frutto, appunto, non di lavori riepilogativi, bensì di nuovo conio. In tal senso *H&F* vuole interrogarsi sull'impatto delle nuove tecnologie sulle forme di governo e autogoverno della vita – tanto a livello individuale quanto sociale – per promuovere forme di regolazione/regolamentazione politico-sociale del rapporto tra l'uomo e la sua potenza tecnologica e digitale adeguate alle sfide contemporanee, anche in considerazione del rapporto dell'uomo con l'ambiente e della ridefinizione stessa di ambiente (che trascende ormai i confini del nostro stesso pianeta).

*Life-on-life*: l'umanesimo  
di fronte alle sfide  
dell'intelligenza artificiale

a cura di Mario Cosenza e Luca Lo Sapiro

Federico II University Press



fedOA Press

Life-on-life: l'umanesimo di fronte alle sfide dell'intelligenza artificiale / a cura di Mario Cosenza e Luca Lo Sapio. – Napoli : FedOAPress, 2025. – XVI+267 p. ; 21 cm. – (Human & Future ; 1).

Accesso alla versione elettronica: <http://www.fedoabooks.unina.it>

ISBN: 978-88-6887-355-4

DOI: 10.6093/978-88-6887-355-4

Volume pubblicato con il contributo dell'Università degli Studi di Napoli Federico II, Fondi Dipartimentali del Dipartimento di Studi Umanistici (2024)

© 2025 FedOAPress – Federico II University Press

Università degli Studi di Napoli Federico II  
Centro di Ateneo per le Biblioteche “Roberto Pettorino”  
Piazza Bellini 59-60  
80138 Napoli, Italy  
<http://www.fedoapress.unina.it/>  
Published in Italy  
Prima edizione: luglio 2025

Gli E-Book di FedOAPress sono pubblicati con licenza Creative Commons Attribution 4.0 International

## Indice

Mario Cosenza, Luca Lo Sapia, <i>Introduzione</i>	VII
<i>Ringraziamenti</i>	XV
Mario Cosenza, <i>Esistenze artificiali. Strutture, politica e filosofia nel digitale</i>	1
Roberto Delle Donne, <i>L'editoria accademica, i dati della ricerca e l'intelligenza artificiale</i>	31
Fabiana Gambardella, <i>Words matter (?): su dilemmi e mistificazioni del Secolo post-nichilista</i>	55
Gianluca Giannini, <i>L'insostenibile leggerezza dell'InfoSapiens (del Digitocene e dintorni)</i>	77
Emiliano Grimaldi, <i>L'intelligenza Artificiale e l'Educazione. Appunti per una sociologia pubblica</i>	101
Luca Lo Sapia, <i>Creatività umana e creatività computazionale: una visione pluralistica dell'atto creativo</i>	123
Simon Pietro Romano, <i>Guerra cibernetica e frontiere etiche. L'evoluzione dei conflitti nell'arena digitale delle sfide globali</i>	157
Francesco Romeo, <i>Intelligenza Artificiale, povertà dell'etica, inutilità del giurista</i>	197
Steven Umbrello, <i>L'intersezione tra il realismo critico di Bernard Lonergan, il bene comune e l'intelligenza artificiale nelle pratiche religiose moderne</i>	219



## Introduzione

Con il lemma-concetto *intelligenza artificiale* (IA) si indica, generalmente, il progetto di realizzare enti artificiali per svolgere mansioni che gli esseri umani trovano faticose o impegnative e, al contempo, la disciplina che studia i meccanismi di funzionamento del cervello umano *a partire* dalla creazione di modelli artificiali. Va da sé che tale definizione riesce a catturare soltanto in parte le plurime sfumature di un fenomeno che, soprattutto negli ultimi anni, con l'emergere e l'imporsi dell'intelligenza artificiale generativa e relazionale, sta disegnando nuovi territori e nuove piattaforme per l'umano e dell'umano che *di fatto* rompono ogni facile schema classificatorio. Siamo nell'era in cui, per riprendere Luciano Floridi, l'essere umano non esiste più off-line e talvolta on-line ma, piuttosto, in cui l'esistenza si dispiega nei termini dell'*on-life*, un esser-connessi che permea ogni aspetto della vita quotidiana e si fa nuova (ormai neanche più tanto nuova) cifra dell'umano. Nell'era dello *humanware* e degli ecosistemi digitali in cui ciascuno di noi *volens nolens* inabita, le opzioni disponibili non sono *essere connessi* o *disconnessi*, nella nuova forma ascetica della *digital detox*, ma *essere connessi* con una postura critico-problematica, nell'ottica della *presa-di-coscienza* o essere connessi *passivamente* – cioè agiti dai processi inaggrirabili che gli ecosistemi digitali dispiegano.

All'interno di tale scenario, in relazione all'IA, possono essere individuati due gruppi di problemi etici: i problemi etici *applicativi* e i problemi etici *strutturali*. Per problemi etici applicativi intendiamo problemi *a valle delle applicazioni* plurime a cui i Sistemi di IA posso-

no essere destinati: qui troviamo i problemi relativi alla *privacy*, all'autorialità, alla trasparenza, alla manipolazione o ancora il problema di quale modello normativo sia più efficace a guidare alcune specifiche applicazioni, dai veicoli a guida autonoma alle *automatic weapons* per la *modern warfare*. I problemi strutturali invece sono quelli relativi alla riscrittura dello status morale dell'umano alla luce degli sviluppi dell'IA, la ridefinizione della nozione di *creatività*, la possibilità di ri-territorializzare la nozione di agency, il tema della Superintelligenza.

Questo secondo blocco di temi, in altri termini, riguarda la costruzione di uno specifico Ethos per l'IA. Come tutte le Rivoluzioni, anche quella digitale, infatti, ci costringe a fare i conti con la necessità di contribuire a *generare* un Ethos per una co-esistenza adeguata, in questo caso, degli enti umani e degli enti artificiali.

In quest'ottica, al di là dei toni allarmistici e delle interpretazioni strumentaliste della Rivoluzione IA, un punto va messo a fuoco: siamo di fronte a una nuova forma di *agentività* che pur non essendo cosciente è in grado di immettere nei nostri ecosistemi sociali prodotti che erano tipicamente associati alla creatività e all'estro umano. In tal senso, proprio quella definizione operativa fornita nelle battute iniziali viene meno. Di fatto, i SIA non vengono usati più solo per alleviare la fatica umana e, così, consentirci di dedicare più tempo alle attività artistiche, culturali e *lato sensu* basate sulla creatività, né solo per coadiuvare il lavoro intellettuale (pur essendo questo senz'altro uno dei possibili usi, ad esempio, di Chat-GPT o altri software di IA generativa e relazionale) ma per partecipare attivamente all'immissione nell'infosfera di oggetti sintetici indistinguibili *a parte objecti* dagli oggetti realizzati dalle intelligenze umane.

Questo, per dare una rappresentazione concreta e plastica del processo in corso, vuol dire trovarsi di fronte un brano musicale, una foto, un filmato o una rappresentazione pittorica che *attribuiremmo*, non esistessero IA generative, senz'altro ad attori umani.

Da oggi, saranno almeno due le specie in grado di generare *strutture* e *contenuti* per gli ecosistemi digitali e tale scenario sarà sempre più accentuato man mano che il controllo degli attori umani sugli enti artificiali diminuirà (e sta già avvenendo dal momento che *intenzionalmente* sono già immessi nello spazio digitale sistemi di IA in grado di produrre senza *prompt umani* oggetti sintetici in un loop nel quale *Sapiens* sembra sempre più attore marginale).

Non si tratta dunque *solo* di elaborare linee guida e modelli normativi per evitare che le applicazioni di specifici enti artificiali risulti potenzialmente dannosa per gli attori umani. Si tratta anche e soprattutto di individuare *nuove categorie* e *nuove forme dell'abitare* che consentano a *Sapiens* di rimanere nel *loop* umano-digitale quantomeno da co-attore protagonista e non da semplice comparsa.

Per fare ciò gli enti artificiali che implementano sistemi di IA non vanno intesi come *meri strumenti* ma come strumenti in grado di ridisegnare, con un margine di autonomia funzionale mai raggiunto prima da un prodotto umano, ecosistemi antropici, possibilità d'essere e modalità relazionali di *Sapiens* con il mondo esterno e con se stesso.

Ci troviamo oggi a navigare in una nuova era, un'epoca storica inedita, caratterizzata non solo dall'onnipresenza delle tecnologie digitali, ma dalla loro simbiosi con le fondamenta stesse della nostra società, della nostra cultura e della nostra identità. In questo contesto, l'intelligenza artificiale è lo strumento di una trasformazione radicale del nostro modo di vivere, lavorare e relazionarci come soggetti in un *infra* ormai integralmente invaso dalla digitalizzazione.

L'ascesa dell'IA, infatti, solleva questioni cruciali e non più procrastinabili che interrogano l'individuo e le comunità, il pubblico e il privato. Si può orientare lo sviluppo tecnologico verso il benessere collettivo o è solo una pia illusione? Quali principi etici devono guidare la progettazione e l'impiego degli algoritmi che prendono decisioni

sul nostro futuro? In che modo il diritto può e deve adattarsi per normare questa realtà emergente, proteggendo i diritti fondamentali all'interno di quello che si configura come un nuovo Ecosistema Digitale Globale? Qual è la filiera *materiale* di queste tecnologie, dove si annidano i suoi segreti politici, qual è il rapporto tra tecnologie, Capitalocene e ritorno delle guerre distruttive nello spazio pubblico tardo-occidentale? Il volume vuole essere, anche, una bussola scientificamente affidabile per orientarsi nella complessità della società dell'IA, offrendo una cassetta degli attrezzi concettuale e pratica per comprendere le sfide odierne. Si tratta di conoscere per demistificare la coltre di "sensazionalismo" riguardo queste tecnologie, fumo che impedisce una realistica comprensione delle poste in gioco; c'è una tensione tra quest'innovazione tecnologica e i quadri normativi e politici esistenti, così come un profondo coinvolgimento del sapere filosofico – si pensi a cosa, di categorie come *bias*, *responsabilità*, *autonomia*, viene modificato con l'irruzione del macchinico negli ambiti etici.

Il presente volume intende presentarsi come un tentativo di ibridare discorsi e discipline. Non si tratta di interdisciplinarietà o transdisciplinarietà: parlare oggi di IA e di digitale è già, di per sé, parlare di un *loop* tra saperi, pratiche e discorsi. Un unico grande – labirintico – rimando in cui l'*umano* è intrappolato se non trova direzioni e bussole. Discipline storicamente concepite come separate sono obbligate dalla loro stessa logica interna a confondersi tra loro, perché proprio nel momento in cui la sfida sembra venire da un fuori – da un'intelligenza *artificiale* – si rendono conto della loro unità: che il loro segreto è l'umano. Se il profluvio non sempre informato di IA ha un merito, è quello di aver abbattuto come un fiume in piena il mutismo selettivo nel dialogo tra *mondi accademici* e di ricerca.

Siamo qui, già a priori *abitanti* dello spazio digitale, conurbazioni necessarie oggi finanche per piccole azioni di riproduzione sociale e della cittadinanza.

Pur accorciato lo iato tra discipline, restano, evidentemente, orientamenti predominanti e sensibilità differenti, non per forza conciliate. Nel volume si incontrano – e, si spera, felicemente *scontrano* – saperi per troppo tempo visti come geneticamente incompatibili. I lettori troveranno un robusto coté filosofico, ma anche nervature attinenti a questioni di Etica dell’informazione, di Diritto, di Sociologia e Pedagogia nonché di Ingegneria e sicurezza.

Il saggio di Mario Cosenza, uno dei due curatori, cerca di delineare alcuni nodi cruciali del rapporto tra soggetti e nuovi sviluppi tecnologici legati all’intelligenza artificiale. Due domande lo guidano, due domande che, anche in forme più implicite, attraversano il volume: le tecnologie digitali sono contendibili politicamente come spazio di implementazione della democrazia o ci si deve rassegnare alla versione edulcorata del reale per le quali esse sarebbero *neutrali*? E, in tutto questo, la filosofia come critica ha ancora uno spazio nell’evidenziare le storture del reale o essa è ormai puro residuo museale?

Come suggerisce Gianluca Giannini, necessitiamo di nuovi lemmi-concetti che forniscano inediti strumenti per comprendere il reale. Quella in corso è una vera e propria trasformazione antropologica innescata dalla pervasività del digitale. Non più *Sapiens*, bensì *Info-Sapiens*, i soggetti, esito di quella che appare essere una metamorfosi irreversibile dell’umano in simbiosi con le macchine cognitive.

Come sottolinea Fabiana Gambardella, difatti, non solo il digitale ma la narrazione stessa intorno al digitale sta velocemente trasformando linguaggio e pensiero dei soggetti, e ci interroga riguardo il nostro ruolo attivo nel “confronto” con le macchine: dove finisce la nostra volontaria plasmabilità e dove inizia invece il feedback per il quale sono loro a ridefinire noi. Anche l’altro curatore, Luca Lo Sapia, arricchisce la discussione proponendo una visione pluralistica della creatività, cercando di includere anche le intelligenze artificiali generative nell’insieme degli agenti creativi, superando la tradizio-

nale distinzione antropocentrica che vela una reale comprensione della co-creatività uomo-macchina. In questo quadro problematico si inserisce il contributo di Steven Umbrello, che tenta una lettura del carattere innovativo dell'intelligenza artificiale a partire da un confronto con un autore come Bernard Lonergan, il cui realismo critico serve all'autore come specchio per mostrare in che modo l'IA interagisca con le tradizioni e rimodelli simboli e riti della vita spirituale.

Di là dai contributi della filosofia morale e politica, nel volume, come detto, si irradiano tematiche che, pur se accomunate da un'inquietudine comune, afferiscono a pratiche di sapere eterogenee. Il contributo di Roberto Delle Donne parte proprio da una problematica "accademica" che però ben si offre a fornire risposte più ampie rispetto alle possibilità di accesso alle tecnologie; lo fa mostrando sfide e opportunità dei nuovi canali di divulgazione scientifica e di modelli linguistici che, su basi algoritmiche, hanno invaso il campo delle pubblicazioni universitarie. Sempre su educazione e sapere dell'IA, Emiliano Grimaldi guida il lettore all'interno di una tematica d'importanza inaggrabile, ossia la comprensione e la valorizzazione delle dimensioni sociali e pedagogiche delle tecnologie digitali: è possibile un utilizzo dell'IA per contribuire a formare cittadini autonomi, in grado di riconoscere i lati opachi di tali tecnologie così come il loro potenziale innovativo nel settore dell'educazione pubblica? Il problema non è meno arduo lì dove si tratti dell'applicazione dei sistemi di IA al diritto, ed è in questo delicato ambito – tra tecnologie, pregiudizi, ermeneutica e teoria del diritto – che Francesco Romeo prova a fare chiarezza sulla contesa (e sulle possibilità a essa inerenti) in atto tra tecnologia e Legge. Infine, sul ruolo della sovranità statale e sulle modalità del muover guerra in ambito cibernetico, Simon Pietro Romano discute di tecnologie e frontiere politiche "fluide", di strategie di protezione e di sicurezza digitale.

## *Introduzione*

Come si vede, tanti elementi per discussioni che sempre più sono il nostro oggi, nell'inesausto rimando tra soggetti e oggetti delle loro creazioni, in questo circolo – virtuoso o vizioso, è il caso di prender posizione – tra enti naturali, artificiali, tecnologie e mondi della vita. Perché l'intelligenza artificiale è oggi e sarà domani, e senza *critica* immanente non resta che la passività di fronte allo stato di cose presenti.

giugno 2025

Mario Cosenza, Luca Lo Sapio



## *Ringraziamenti*

Individuare in modo esaustivo tutte le persone che hanno contribuito alla realizzazione di un'opera collettiva e interdisciplinare è impresa tutt'altro che semplice. Ciò vale a maggior ragione per un lavoro che nasce all'interno di un progetto sperimentale come *Human&Future*, la Task Force di Ateneo dell'Università degli Studi di Napoli Federico II, della quale il presente volume rappresenta un primo precipitato, inaugurandone la collana editoriale.

Obiettivo dichiarato della Task Force è la creazione di uno spazio stabile di interscambio tra discipline, capace non solo di favorire il dialogo tra saperi solo apparentemente distanti, ma di concorrere alla costruzione di un linguaggio condiviso tra filosofia, scienze umane e cosiddette "scienze dure". Questo volume, per quanto in forma embrionale, intende riflettere e testimoniare tale ambizione, già "sperimentata" nei numerosissimi seminari, convegni e giornate di studio di questi primissimi anni di vita di *Human&Future*.

In quest'ottica desideriamo dunque esprimere la nostra profonda gratitudine a tutti coloro che hanno contribuito alla stesura del volume: il Professor Roberto Delle Donne, la Professoressa Fabiana Gambardella, il Professor Gianluca Giannini, il Professor Emiliano Grimaldi, il Professor Francesco Romeo, il Professor Simon Pietro Romano e il Dottor Steven Umbrello. Desideriamo ringraziarli non solo per il valore dei rispettivi contributi, ma anche la disponibilità e l'apertura con cui hanno accolto la proposta di partecipare a un progetto di natura trasversale e sperimentale e ne hanno tollerato i rallentamenti.

Un ringraziamento sentito da parte nostra va in particolare alla dottoressa Alessia Gifuni, nonché al dottor Mattia Papa e al dottor Valerio Specchio, studiosi, collaboratori e amici che hanno donato il loro tempo e la loro attenzione offrendo un prezioso contributo di editing e revisione dei contributi.

MARIO COSENZA

*Esistenze artificiali.*

*Strutture, politica e filosofia nel digitale*

*Abstract:* The essay aims to outline some key points of an interpretation of the relationship between subjects and the new technological-digital developments linked to artificial intelligence. It seeks to introduce a discussion on the potential political contention surrounding technologies that, by presenting themselves as neutral, completely mask their potential to be destructive instruments for democratic societies. Does philosophical thought still have something to contribute to this issue, or must it resign itself to inconsistency?

*Keywords:* Philosophy, Politics, Digital, Artificial Intelligence, Science, Structure, Power, Democracy, Materialism.

1. *Soggetti, strumenti e specificità della tecnologia contemporanea*

1.1

Il rapporto tra *Sapiens* e tecnologia, o, meglio, la sua narrazione, può contenere sfumature epiche. Benjamín Labatut, nel suo *Maniac*<sup>1</sup>, ha raccontato con maestria la tormentata stagione del pensiero scientifico occidentale, o almeno quella sottosezione che possiamo situare entro un intervallo che va dal dibattito matematico di inizio '900 fino alla decisione statunitense di radere al suolo Hiroshima e Nagasaki. Sulle due sponde dell'Atlantico, mentre il conflitto mondiale spazzava via le ultime vestigie dello *jus publicum europaeum*, le certezze secolari

<sup>1</sup> B. Labatut, *Maniac*.

dell'*esattezza* scientifica traballavano sotto i colpi dell'*indeterminismo*; e l'etica della ricerca veniva interrogata – o forse *travolta* – dalle inquietudini dell'utilizzo delle armi distruttive a potenziale *definitivo* generate dal lavoro di scienziati a pieno titolo coinvolti nel conflitto. Non è questo il luogo per indagare le ricadute psichiche sui singoli pionieri – Labatut insiste su von Neumann – della proteiforme saldatura politica-tecnica-tecnologia bellica che inquietò la coscienza scientifica; ma, seguendo la narrazione di Labatut, possiamo segnare una linea che unisca le prime macchine computazionali di von Neumann fino alla sfida del 2015 tra Lee Se-dol e *AlphaGo*, il programma finanziato da Google per giocare all'antichissimo gioco cinese Go e uscito vincitore dalla contesa. Un po' come la sfida a scacchi tra Garry Kasparov e Deep Blue che nel 1996 rivoluzionò la percezione comune delle potenzialità dei programmi di intelligenza artificiale, la sfida del 2015, la schiacciante vittoria artificiale, è il termine *post quem* della comprensione della potenza algoritmica.

Epica, senz'altro; ma non solo. A cotanto racconto "alto" andrebbe affiancata, foss'anche solo per esaltare la varietà d'applicazione del nuovo sterminato campo *digitale*, un'altra delle prestazioni riconducibili all'intelligenza artificiale. Un esempio tra tanti: verso la fine del maggio 2024, le informazioni fornite da *Google Maps* hanno contribuito a generare una gigantesca coda in uscita dall'autostrada del Brennero segnalando un'interruzione che in realtà non sussisteva. Neanche l'intervento di decine di addetti intenti a garantire che non fosse necessario uscire dall'autostrada ha evitato che gli automobilisti scegliessero comunque di dare *fiducia* al consiglio algoritmico, contribuendo così ad aumentare l'ingorgo nelle strade secondarie. Di fatto, il consiglio (sbagliato) dell'app era ritenuto più affidabile financo della segnalazione in presenza di addetti non artificiali. Si cita un episodio di tal genere perché, tra i tanti, espone didascalicamente l'autorevolezza di cui le tecnologie algoritmiche godono ai

nostri occhi nel sistema-mondo che abitiamo, ormai dominato dal monopolio strutturale di poche grandi aziende, nel quale basta un errore in un file per paralizzare il traffico mondiale di aerei, trasporti e merci, o un collo di bottiglia in un punto della catena di valore del commercio mondiale per creare scorie impossibili da smaltire o, ancora, un errore banale per imbottigliarsi nel traffico. Gli errori delle macchine li pagano gli umani.

## 1.2

L'auspicio è che ci si ricordi che alla vita umana l'*affaire* IA sempre attiene. Difatti, la discussione intorno all'IA è ancora costantemente gravata da uno scontro tra apocalittici e integrati<sup>2</sup>, contesa che impedisce di pensare in termini concreti la questione dell'impatto di tali tecnologie sul vivere associato che le genera e che esse contribuiscono a polarizzare; termini concreti che sono sempre inevitabilmente termini materiali e termini *pensati*, pratiche e concettualizzazioni, e che così devono essere intesi se si vuol rendere servizio alla complessità dell'organizzazione funzionale delle forme antropiche d'essere. Non disprezzare né temere bensì capire, insomma.

In tal senso, da un'ottica filosofica si tratta sempre e comunque di indicare quale può essere il ruolo del pensiero umanistico *tout court* in questa discussione – uno dei grandi drammi della filosofia nella società: doversi difendere e giustificare. Ad esempio, se il contributo debba essere di tipo *deontologico*, nell'indicare cosa sia *giusto* o meno fare nei vari dilemmi morali scaturiti dall'artificialità dell'azione e della

<sup>2</sup> La canonica distinzione è da ricondurre al saggio del 1964 di Umberto Eco, *Apocalittici e integrati: comunicazioni di massa e teorie della cultura di massa*.

decisione; o se debba servire da collante al funzionamento di questa o quella applicazione pratica; o se, invece, tale contributo, quello di una filosofia che ambisca a essere vettore di comprensione di complessità, possa fungere da dissonanza che ricordi la necessità di un giudizio a carattere sintetico riguardo il mondo com'è e come potrebbe essere, e che questo giudizio è questione profondamente umana, al netto delle scelte lessicali e di denominazione – uomo, ente, *Anthropos*, etc. – che si decide di applicare a chi si pone l'inquieta domanda.

Invece di chiederci della sopravvivenza dell'umanità all'intelligenza artificiale, la questione realmente urgente concerne la ridefinizione che la realtà sta attraversando<sup>3</sup>: dall'ottica tardo-occidentale dalla quale si scrive, *quale* umanità ripenserà e riprodurrà sé stessa al cospetto dell'incalzante ibridazione tecnologica e dell'automazione che si presume intelligente? Tecnologia, digitale, rivoluzione artificiale – e noi. Si tratta di pensare sempre la produzione di certi soggetti – e di certe forme di soggettività – a partire dalla *situazione*, cercando di sfuggire alle trappole sommamente ideologiche della naturalizzazione di presunte caratteristiche eterne degli *enti*; tale pensiero non può ritrarsi dinanzi alla sfida di cogliere le nuove e cangianti predicazioni inerenti ai soggetti né di fronte alle nuove discipline sorte da pratiche di potere mai neutrali. Ancora, il soggetto della modernità occidentale, anche se con gestazioni differenziate e tempi non per forza uniformi, nasce nel cono d'ombra della forma Stato, della sua organizzazione economica vieppiù capitalistica, del

<sup>3</sup> La questione del “rimodellamento della realtà”, come suo cambiamento di forma, costituisce, tra l'altro, il sottotitolo del testo di Luciano Floridi dedicato alla Quarta Rivoluzione: *The Fourth Revolution. How the Infosphere is reshaping Human Reality*, sebbene nella traduzione italiana il termine “realtà” sia stato sostituito con “mondo” e “reshape” con “trasformazione”, per cui il titolo in questione diventa: *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*.

suo quadro istituzionale così come delle contese interne alla ripartizione di ruoli e poteri nel suo recinto e agli scontri esterni; e così come il dominio di una certa forma economico-politica ha prodotto certe caratteristiche nei soggetti – in primis, una sorta di crescita esponenziale del riconoscimento delle facoltà soggettive, ossia l’iper-inflazione della *coscienza individuale*, funzionale anche al trionfo di un regime di organizzazione economica che dal margine di profitto del singolo parte per costruire il mondo intero –, è lecito pensare che nuove forme politiche o di negazione della politica, produrranno nuove ibridazioni e nuove pratiche di soggettività. Proprio su queste bisogna interrogarsi, così come sul rapporto tra gli habitat digitali e le produzioni antropiche.

Le questioni sono molteplici: anche solo per porsi in maniera primordiale la domanda intorno al *pensare* il rapporto tra enti materiali storici e tecnologie – l’intelligenza artificiale, ma non solo – bisogna dunque tener conto che non è scontato che si dia un qualcosa di genericamente inteso come il *pensiero* e che non si può prescindere – nell’evocare la possibilità di dire qualcosa intorno a tale funzione cerebrale nonché di significazione simbolica – dal ricordare che anch’esso è tutto *nella* storia materiale dei rapporti umani e sociali; ossia, che il *pensare* è sempre in una cornice ampia, in orizzonti storici e politici, e non nel gesto auto-riferito di una coscienza isolata che prova a convocare il mondo dinanzi a sé. Il rapporto con i manufatti tecnologici, dal bastone a *Maniac*, è sempre interno alle configurazioni dei rapporti sociali e politici.

### 1.3

Non c’è rivoluzione tecnologica che possa sfuggire a questo incardinamento nella storia, foss’anche se si decidesse e sanzionasse

che davvero l'epoca da noi abitata è epoca di *fine della storia*, dunque senza storia, se non addirittura post-politica – o meglio, senza possibilità di pensarla e controllarla, la politica, poiché anche l'assenza di politica è politica.

Pensare l'IA quindi è pensare in *questo* mondo storico e non in astratto.

*Anthropos* è costitutivamente potenzialità, apertura, *struttura relazionale* che sempre fa i conti con la propria ambientazione storica, coinvolto in un insopprimibile rapporto tra *singularità* della coscienza e *unitarietà* dell'esperienza storica. Esso non solo *abita* il tempo e la storia, ma *è* abitando il tempo, in un inesausto rapporto tra soggetto e condizioni storiche – che hanno il volto di scienza, tecnologia, politica, cultura e concretezza di produzione sociale. Un condizionamento che modella il soggetto nella propria materia: tale potenzialità di *Anthropos* non è solo astratta convinzione di filosofi ma concreta *potenza* del cervello nella strutturazione (o nel deterioramento) delle sue facoltà<sup>4</sup>.

Il rapporto con il mondo esterno – con l'oggetto: ossia l'interazione bi-direzionale tra ente e mondo – influenza le caratteristiche della struttura; lo fa in maniera non infinita, “eliminando” continuamente possibilità e traiettorie, tramite una dinamicità interattiva e ri-organizzativa che agita – inquieta – i soggetti, portandoli verso nuove riconfigurazioni<sup>5</sup>. Nell'ottica di un dialogo fecondo tra

<sup>4</sup> In particolare, il concetto di plasticità cerebrale ha ampiamente innervato le neuroscienze contemporanee dimostrando la possibile riconfigurazione durante l'arco di un'esistenza delle strutture cerebrali con cui si pensa il mondo.

<sup>5</sup> Tale capacità neuronale e cerebrale appare del tutto coerente con un imprinting filosoficamente materialistico, di quel materialismo che già si interrogava sulle possibilità di un salto qualitativo della materia organizzantesi in un nuovo equilibrato sistema di complessità; materia che pensa. Il salto qualitativo prevede

discipline i cui paletti sono ampiamente superati appare del tutto lecito pensare *dalla parte della filosofia* in che modo i nuovi sistemi di IA concorrano alla modificazione dell'apertura generica di *Sapiens* – alla sua *plasticità* – a cui addirittura la facoltà della “sapienza” avrebbe affidato, nei dettami della tassonomia, la sua stessa caratterizzazione prioritaria rispetto alle altre specie esistenti. Se si modifica la sapienza di *Sapiens* si sta evidentemente modificando *Sapiens*<sup>6</sup>. Se *Sapiens* è l'ente dell'apertura generica e della condizione che non lo obbliga all'aderenza perfetta al dato biologico ma mantiene una forma di distanza rispetto alla pura fisiologia, il cambiamento fa parte del suo raggio di possibilità, e i paradigmi tecnologici e strumentali di una certa epoca intercorrono a perimetrarne il recinto di possibile *azione* – che resta una sua prerogativa – e di successiva ominazione, a perimetrarne le possibilità di persistenza in vita e di adattamento strategico.

però una nuova discussione intorno alla galassia valoriale, al plusvalore simbolico che anche si ritiene impossibile da non considerare nel concettualizzare il pensiero e il macchinale, pena la caduta in un nuovo meccanicismo, in un nuovo uomo-macchina, che però, al contrario dei tentativi antropologici del '700 non contenga in sé nessuna scintilla di comprensione libertaria e financo gaudente dei nessi materiali.

<sup>6</sup> La dicitura *Sapiens* categorizza una classificazione tassonomica; *Anthropos* prova a delinearne alcune condizioni inaggrabili della finitudine dell'esistere, noi esseri materialmente costituiti ma anche forniti di un dato coscienziale insopprimibile; *uomo* richiama la definizione che l'intero sapere filosofico ha dato della specie del soggetto pensante, singolare e generico, aperto. Definizioni, denominazioni, euristiche, tendenti a un noi complessivo di impossibile chiusura: qualsiasi di esse si utilizzi, l'oggetto coincide col soggetto che formula il pensiero e la possibile definizione e, insomma, la nomenclatura non esaurisce il problema, e può addirittura velarlo. Nell'alternanza di scelte lessicali, non ci si deve perdere l'essenziale; parlare di IA, parlare di digitale ha in sé un segreto: che la domanda la poniamo noi.

Il genio di Stanley Kubrick aveva racchiuso il tutto in un singolo passaggio di montaggio, probabilmente senza pari nell'intera storia del cinema: quel primate, ancora poco antropizzato, che scopre la potenza dell'utilizzo di una tecnologia volta al predominio – l'osso servitogli per dominare l'elemento biologico, così da trionfare in battaglia – e, in tal modo, fonda la Storia; lo stesso osso che, lanciato in area, sulle note di Strauss, porta direttamente – con un salto epocale ma a suo modo *conseguente* – al dominio dell'elemento cosmico-spaziale, e diviene una navicella che galleggia nello spazio. Visione terribile, ad alto carico di pessimismo antropologico, che indica però che è proprio nel rapporto (in quel caso) manuale con manufatti e artefatti che si annida il senso della crescita di quello scimmione modificato divenuto padrone, nonché potenzialmente distruttore, del mondo. La natura che arriva a pensare – e a muovere guerra a sé stessa.

Non serve abbracciare integralmente i nessi causali né la visione implicita nella immaginifica versione evolutiva di *2001: Odissea nello spazio* per inquadrare il rapporto strettissimo tra umano e tecnologico – semmai ci aiuta a rinfrescare l'idea che *tecnologia* non è per forza “elettronica-digitale”, ma è, in primis, manipolazione e addomesticazione per fini specifici e adattivi a un mondo che a *Sapiens* è, in parte, ostile.

Serve qui ricordare quanto una forma particolare di tecnologia che venga concepita come *intelligenza artificiale* agisca immediatamente sulle configurazioni essenziali dell'ente, ponendone in questione nient'altro che la sua *diversità* e la sua *specificità* – peraltro sempre surrettiziamente considerata *superiorità* dalla tradizione occidentale – ontologica; e che, se tutte le forme di tecnologia e di strumentalità sono state in qualche modo forme di *razionalità* strumentale, la nuova configurazione della tecnologia così come veicolata dall'informatica e dal digitale è di per sé stessa una riconfigurazione dello

stare al mondo dei soggetti. Siamo quindi dinanzi all'affermarsi di una tecnologia che, già solo nella sua funzione di "mediazione" per appartenere alla società – anzi *al mondo*, per usufruirne, per averne *accesso* –, ossia quella algoritmica dei supporti tecnologici che ormai la maggioranza utilizza, modifica non solo o non tanto posture e arti bensì quelle funzioni in qualche modo derivate dalla struttura cerebrale, quelle mentali. Anche se si prende la cosa in senso molto ampio, senza per forza pensare di risolvere il plurisecolare dibattito della dipendenza delle funzioni psichiche dal cervello si può dire che parliamo di tecnologie che modificano in primis l'apparato psichico, il suo funzionamento e di conseguenza la dimensione simbolica.

Siamo dinanzi al trionfo di una tecnologia che incide come mai prima sulla psiche dei soggetti che la utilizzano – quella stessa psiche che si vorrebbe *imitare* tramite funzioni matematiche, decidendo che l'intelligenza umana è solo calcolo; e comunque, al netto della modifica intrinseca in ogni rapporto tra soggetto e mondo, siamo dinanzi a una novità assoluta riguardo ciò che concerne le "porzioni", le "aree", le caratteristiche cerebrali e psichiche potenziate, allenate, stimolate e chiamate in causa. La tecnologia diviene sempre più una forma non più ausiliare bensì del tutto sostitutiva di certi tipi di funzioni – e dunque di caratteristiche *tout court* antropologiche – prima affidate a porzioni differenti di cervello. E, *soprattutto*, siamo dinanzi a strumenti che *imparano*, che interagiscono, che si *modificano* in seguito alla retroazione (feedback) della loro relazione con i soggetti, che, fatte salve le necessità energetiche per alimentarle, sono funzionalmente autonome.

Avviene dunque la generazione di caratteristiche antropologiche che prima non erano contemplate tra le possibilità, in una misura largamente nuova; questo può apparire addirittura scontato, a ben pensarci, ma non pare superfluo calcare la mano riguardo questo aspetto lì dove sembra più presente il rischio contrario di presentare

gli strumenti della tecnologia come puramente *neutrali* non solo da un punto di vista degli attori e degli interessi, bensì nel senso più forte di una negazione della possibilità strumentale di *antropologizzare* diversamente l'ente: se esso è *solo* naturale, se è dato una volta per tutte, imm modificabile, ne consegue che lo strumento è solo un mezzo per un fine, in alcun caso condizionato invece dalla storia, dai valori, dalle contingenze, insomma dai fattori che lo *condizionano*. Si dice: già una calcolatrice sostituiva qualcosa dapprima affidato esclusivamente al calcolo umano e poi vieppiù ai vari strumenti; in questo senso l'intelligenza artificiale porta all'estremo un qualcosa di già da sempre dato nel rapporto con la cosa. La differenza è però che nessuno si sarebbe sognato di dire che l'osso, l'abaco, o le prime forme di calcolatrici fossero *intelligenti*<sup>7</sup>; la configurazione di *oggetto* o, anche, di strumento macchinale risultava non contendibile, mancando una qualsivoglia forma di modificazione dello strumento stesso in seguito alla relazione col soggetto. L'ente da sempre *subisce* il mondo, da sempre *lo modifica* col pensiero e con l'azione, ma mai prima lo strumento da lui creato poteva in qualche modo *ribattere* e modificarsi.

<sup>7</sup> Sulla legittimità di definire "intelligenti" le nuove prestazioni degli strumenti tecnologici, il dibattito è fervido, e sono molti coloro che ne sottolineano la problematicità. Tra essi, Luciano Floridi sostiene che questo tipo di "intelligenza" è più una capacità di eseguire compiti specifici, come riconoscere pattern o fare previsioni, che non implica una vera e propria "intelligenza" nel senso cognitivo che noi attribuiamo agli esseri umani. A prescindere dall'accordo che si può dare o meno al termine, l'aspetto rivoluzionario di queste tecnologie concerne la capacità di svolgere attività considerate strettamente e *identitariamente* umane, riproponendo, di fatto, la domanda circa la specificità umana e il suo ruolo nell'uso delle macchine, oltre che sulla presunta inutilità dell'uomo rispetto alla macchina, di fronte alla quale, come ben argomentava Günther Anders, i soggetti si vergognano per la nostra finitezza.

Se ne possedeva, inoltre, una chiave d'utilizzo e di ricostruzione della sua genesi a misura di pensiero singolo, integralmente in grado di "riprodurre" o ricostruirne i connotati; poi, sempre meno capace di costruire uno specifico strumento a propria volta senza una specializzazione; adesso però totalmente dominato dalla specificità e dall'ampiezza di saperi custoditi e necessari alla costruzione della strumentazione tecnologica dell'intelligenza artificiale contemporanea.

Porsi sempre lo sfondo della carica di novità che la tecnologia apporta intrinsecamente alla plasticità dell'ente – di cui la plasticità cerebrale è in fondo una delle basi neuroscientifiche – e risvegliare il senso di innovazione radicale di questa specifica tecnologia assume un carattere testimoniale da non sottovalutare, poiché riporta l'attenzione al fatto che si è sempre dinanzi a funzioni adattative e persistentive di un ente che risponde a sfide di adattamento e agisce sulle condizioni che a loro volta agiscono su di lui; un ente che cerca infinite forme di rapporto con ciò che esso stesso mette al mondo con la sua attività sensibile, produttiva, informativa e conoscitiva. Porre quindi l'intelligenza artificiale *al di là* di ciò che viene prodotto e negarne le funzioni di riconfigurazione antropica significa pensare una cosa in sé fuori dal mondo della vita e, dunque, politicamente *sempre già indirizzata*.

Pone, allora, le condizioni del dominio.

## 2. *Vincitori e vinti della non neutralità tecnologica*

### 2.1

Il potenziale novatorio del rapporto ente-prolungamento tecnologico è dunque una costante da cui risulta impossibile prescindere. Cambia col tempo, e allena diversi modi dell'umano.

Non esistono *gli oggetti* fuori da un tempo storico. Essi assumono le sembianze del processo di produzione necessario a immerterli nel mondo e a combinarli tra loro. La quantità di conoscenze e di lavoro presente in un singolo oggetto è figlia del nesso con i fattori storici di produzione, e come tale si presenta a un soggetto umano che è di per sé storico, ossia situato sempre nel qui e ora, in rapporto con una pluralità di dimensioni che sono però sempre e comunque indirizzate dalla materia sensibile.

Come ha ben scritto Numerico: «La tecnologia rappresenta un'interfaccia per rapportarsi socialmente tra esseri umani, ma anche una connessione con gli altri viventi e il mondo inorganico. Occorre riflettere su quali opzioni renderanno tali rapporti più duraturi e fiorenti, specialmente se intendiamo il nostro modo di essere come il nodo di un sistema triadico in cui la tecnologia svolge una funzione di mediazione, rappresentazione e intervento. Nella consapevolezza che la maggior parte dei contesti di cui facciamo parte non si prestano a essere quantificati precisamente. È perciò necessario definire localmente i criteri delle nostre ragioni nella storia e nello spazio che abitiamo. Senza consegnarci a una fantomatica, insidiosa razionalità prefabbricata, unica e univoca»<sup>8</sup>.

*Non è indifferente* che l'oggetto sia la clava, la falce o un tablet<sup>9</sup>; non lo è né per i “luoghi” cerebrali su cui impatta, non lo è per

<sup>8</sup> T. Numerico, *Dobbiamo ripensare l'intelligenza artificiale*.

<sup>9</sup> Sul rapporto tra oggetti materiali ed evoluzione biologica, Leroi-Gourhan è pensatore di riferimento: per il filosofo francese, l'uso degli oggetti, in particolare la tecnologia e gli strumenti, modificherebbe non solo l'ambiente umano, ma anche la stessa natura dell'essere umano, unendo in un unico processo l'evoluzione biologica e quella culturale. La relazione tra l'uomo e gli oggetti tecnici non è soltanto una questione funzionale, ma costituisce una parte essenziale dell'evoluzione umana.

quanto riguarda la consapevolezza possibile della complessità di operazioni tecnico-materiali che hanno contribuito a formarlo in quanto manufatto, né lo è per la quantità di potere contenuto in quello specifico oggetto. Più la complessità di operazioni contenute in un manufatto è ampia, più aumenta l'incomprensione della quantità di lavoro (suddiviso su catene del valore mondializzate) necessaria alla sua strutturazione. La catena del valore di oggetti quali i tablet o altri supporti digitali è ampia e fragile, completamente legata alla possibilità del commercio stabile nel sistema-mondo; il Covid-19 o le interruzioni del passaggio nei vari colli di bottiglia del commercio mondiale dimostrano la fragilità del tutto.

Benché di natura diversa, ma anche i dati sono merce. Oggetto virtuale, entità che può essere elaborata e processata tramite algoritmi, il dato è il vero protagonista della connessione, fino a divenire ciò che comprime l'esperienza del soggetto, colui che la *significa*. Ora, l'estrazione di dati precede completamente lo spazio di formazione di un'entità identificabile come soggetto. Nello scenario virtuale, quel che sembra un ingresso *primigenio e neutrale* è in realtà l'immersione in una rete di estrazione e ri-significazione complessiva dei dati accumulati dall'esperienza d'altri – sempre quantitativamente intesa – e dalla volontà di sviluppatori nonché dei proprietari dello spazio. Le reti social sono in questo una presunta offerta di libertà senza mediazioni il cui prezzo è l'acquisizione dell'intera esperienza virtuale stessa, le preferenze, le scelte, le convinzioni. La gratuità d'accesso è in questo senso il cavallo di troia per un (non) contratto virtuale in cui il singolo utente lascia tutto in cambio di qualcosa di effimero – un contratto iniquo, per dirla con Rousseau.

Il dato, considerato come ente neutro, anteriore a ogni accumulazione originaria, non inerisce affatto in maniera diretta ai fenomeni, ma ne è solo rappresentazione specifica; e senza interpretazione

che ne sveli il carattere escludente per tutti gli aspetti considerati contingenti, la retorica *bigdataistica* sarà sempre basata, a esser buoni, su di un grossolano equivoco – quello per il quale una colossale mole di informazioni, la *quantità*, sia di per sé garanzia di esaustività e di scientifica obiettività, e dunque di *qualità*.

In questa trasformazione dalla quantità alla qualità scompare invece ogni possibile comprensione della complessa serie di interrelazioni sociali e tecniche che includono le varie forme di “partecipazione” al processo dell’accumulo. I dati insomma recano con loro l’inesausto enigma della merce: essi sono un rapporto sociale, ossia un rapporto tra enti nella loro relazione compartecipativa nonché in *relazione* al mezzo utilizzato e al mezzo di produzione sociale. Tutto questo scompare completamente nell’ideologia dataistica, la quale occulta e reifica il dato rendendolo oggetto estraneo alla sua stessa produzione. Già definire cos’è un dato è operazione politica con profonde radici filosofico-scientifiche e impliciti ideologici.

Il controllo di sé stessi e la prevedibilità delle azioni sembrano essere, nell’ideologia dataistica, lo slancio verso la conquista del *Graal* dell’infinita prevedibilità statistica del comportamento umano: un enorme epocale *Calculemus!* a regolare il traffico del grande terrore dell’imprevedibilità di *Sapiens*. La quantità di dati immagazzinati sorpasserebbe sé stessa – una fase puramente volta a *incamerare* informazioni da ogni singola interazione su una qualsivoglia piattaforma tecnologica social algoritmicamente fondata – e porrebbe così le basi per la sua trasvalutazione verso forme *prescrittive* di *agire*; la *quantità* di *predizione* genererebbe la prescrizione verso la *qualità* di un’azione.

In questo circuito vizioso, i singoli utenti, spesso del tutto lontani dall’essere *soggetti consapevoli*, sono a priori immersi in un flusso informativo già dato, insopprimibile e trascendentale, e contribuiscono con ogni loro gesto social, ogni loro ricerca, ogni loro informazione a

formattare intorno alla loro minuta esperienza il proprio spazio vitale<sup>10</sup>. Una “bolla” in cui il proposto è già gradito, e ciò che è gradito è riproposto, e da cui i mondi della vita si parlano esclusivamente nella tangenza di due bolle su di un singolo contributo. A questo punto, l'uomo è ciò che chatta, ed è ciò che domanda a ChatGPT credendo di farsi avventuriere del sapere mentre ne viene divorato. Il rischio è dunque l'oblio più profondo riguardo la consapevolezza che la gamma delle azioni umane è sempre più ampia della sua *reductio* a mero dato da incorporare, alla sua quantificazione a fini algoritmici; e le sfumature del giudizio sono assai ben più ampie e multicolori e complesse della riduzione al *like* o alle poche gamme di emozioni offerte come *reaction* all'utente. Se il *like* è la configurazione monodimensionale dell'apprezzamento si genera allora una unidimensionalità dell'esperienza riflessiva-predittiva-pratica, si appiattisce la globalità dell'esperienza ermeneutica ed estetica del giudizio giungendo a una forma meccanica e puramente binomiale della valutazione – un sì/no che non ammette repliche, una incomprendimento radicale. È su questa perdita di spessore che il pensiero critico deve insistere, ben oltre le mere questioni da apocalittici o integrati.

<sup>10</sup> Risulta quasi profetica la tematizzazione andersiana dell'atomizzazione dei soggetti implicita nell'avvento dei media al tempo contemporanei, di televisione e radio. Esemplificativo il caso della televisione: quando in casa arriva la televisione, alla conversazione si sostituisce la fruizione del mezzo, in virtù di cui allo sguardo reciproco subentra quello unilateralmente diretto verso il mezzo. Con il digitale, questa parcellizzazione si radicalizza, per cui i soggetti non sono più accomunati nemmeno dall'oggetto-punto di fuga. Ognuno se ne sta nel proprio piccolo mondo che lo strumento, modellato *ad personam* per i suoi contenuti, rende a domicilio.

## 2.2

È possibile riconoscere le tracce dell'umano se si individua nel funzionalismo esasperato delle macchine addestrate con dati rubati (anche quelle più straordinarie) una parcellizzazione del pensiero umano che si vuole imitare – né il migliore né quello più alto. Ancora, è possibile riconoscere, nelle macchine, la società umana – dalla filiera estrattiva all'addestramento, dalla pubblicizzazione allo scarico dei costi – solo ricordando che il funzionamento della tecnologia è legato a doppio filo alla *storia sociale* della tecnica: le macchine nascono come prolungamenti e implementazioni di rapporti sociali di dati momenti ed esse si generano per rispondere ad alcune specifiche problematiche nell'organizzazione del lavoro e dei poteri. E quindi: conoscere la macchina è conoscere come pensa, com'è organizzata, com'è diviso l'ente sociale che la pensa. Con Feuerbach, saremmo ancora alla scoperta che *il segreto* dell'intelligenza artificiale è la psiche socializzata di un'epoca storica.

La concezione sensazionalistica dell'IA sembra invece sostenere l'idea di un'intelligenza artificiale generale in grado di assumere funzioni pressoché divine (con l'algoritmo a fare da Sacra Scrittura), una sorta di ente storico oltre-storico, materiale ma non contendibile, in grado di alzarsi "al di sopra della propria ombra" e configurarsi la totalità delle condizioni storico-sociali complessive e dunque farsi mente di Dio, intuizione intellettuale, noumeno. In questa straordinaria astrazione parcellizzata e filtrata dalle magnifiche sorti che vediamo narrate, i dati esistono in loro stessi, in purezza, e non come unione continua di estrazione, rappresentazione e selezione di lavoro involontario. Destino di tale ideologia è velare il ruolo attivo dei soggetti che in giro per il mondo contribuiscono a fornire informazioni numeriche di sé, così come il ruolo ben più consapevole dei cacciatori di dati delle grandi e dominanti *corporations* che hanno

organizzato l'architettura dello spazio digitale e ne posseggono le chiavi riproduttive. Tutto questo crea un piano inclinato di continuo restringimento del dato dell'esperienza, del ruolo interpretativo del soggetto, della capacità di "creare complessità" politica e dunque di pensare a come contendere il dominio degli spazi ormai a pieno titolo pubblici; ci si scorda che in questo furioso articolarsi di informazioni in cui tutto sembra già fatalisticamente determinato bisogna inserire anche i soggetti che lo pensano e che concettualizzano il tutto, e che a questo tutto donano loro stessi; ci si ferma a una rigidità meccanicistica dogmatica senza esplicitare il ruolo creativo degli enti, attivi anche quando presi pienamente nella chapliniana catena di montaggio; costretti a iper-produrre loro malgrado, finanche nel sonno, e costretti a dimenticare che le nuove configurazioni tecnologiche nascono da una sottrazione ai loro danni.

### 2.3

Nell'ambito dei big data, nel digitale che è ormai ambito impossibile da evitare anche per funzioni banalmente quotidiane, possono restare intrappolati i corpi<sup>11</sup>. L'ibridazione, l'immersione in una realtà aumentata (tanto più che la realtà *non aumentata* sembra deprimere), o anche solo il fisiologico lento sostituirsi di un certo tipo di esperienza *de visu* con rapporti che nascono e talvolta possono anche morire in quanto virtuali, mette a rischio di polverizzazione l'esperienza sensibile dei corpi? La domanda è lecita, e non è detto che la risposta apocalittica sia la più avveduta; potremmo anche dirci che

<sup>11</sup> Per una riflessione specifica sulla corporeità, si veda M. T. Catena, *Breve storia del corpo*.

in sé stesso il campo digitale-virtuale come “ambientazione” (con l’IA quale strumento) sia l’epifenomeno di una polverizzazione già avvenuta dell’esperienza *moderna* dell’esperire e dell’abitare il mondo in un certo modo relazionale, della quale l’organizzazione delle interazioni corpuscolari digitalizzate altro non è che un’ovvia caduta pratica? Ossia, che l’IA altro non è che il modo tecnologicamente adeguato nell’elaborazione sovrumana di dati nel riproporre digitalmente l’esperienza di una solitudine già ampiamente operante? Certo, appare quantomeno ingenuo sostenere che prima dell’apparizione dell’abitare digitale il modo d’essere delle relazioni antropiche fosse un Eden; il punto è però chiarirsi se questo sistema digitale-virtuale di intendere le relazioni umane e così depotenziarne i caratteri di *tensione* con l’esistente – al fine, perché no, di renderne impossibile il ribaltamento – non rechi con sé un *plusvalore* di *repressione addizionale*<sup>12</sup> rispetto alla finitudine inconciliabile di ente e mondo nonché alle disfunzioni necessariamente operanti in una qualsiasi associazione umana. Senza speranze palingenetiche – e dunque intrinsecamente non *tragiche* della condizione umana –, è lecito chiedersi se a operare come acceleratore di infelicità non siano le stesse tecnologie che nascevano – a sentirne i fondatori, si intende – come facilitazioni al servizio di una radicale accessibilità a nuove forme di relazioni tra soggetti.

Insomma: quale che sia la breccia scelta per entrare nel discorso sulla *forma mentis* algoritmica o sugli ecosistemi digitali abitati ogni istante della nostra esistenza, nondimeno salta agli occhi che nella narrazione intorno alle rivoluzioni ciò che manca è uno degli aspetti meno negabili delle rivoluzioni stesse – ossia: che esse prevedono dei *vincitori* e dei *vinti*, non sono a somma zero, sono, senz’altro, l’*evento*

<sup>12</sup> H. Marcuse, *Eros and Civilization*.

ma anche il *processo* che a esse porta<sup>13</sup> e generano nuove strutturazioni dei rapporti, nuove governamentalità, nuovi modi di abitare l'esistente così come nuovi modi di esclusione.

L'algoritmo non è neutralità né diamante puro, non è un *cosa* ma un *come*, un funzionamento che si innesta su una certa organizzazione sociale e di essa si fa prolungamento al fine di perpetuare l'organizzazione stessa che ne è alla base; può essere strumento di disciplinamento e sfruttamento del lavoro diretto e indiretto, strumento di esclusione, di aumento di diseguaglianza, propaggine dell'esistente che all'esistente rimanda – un canone di riproposizione di ciò che è già, un cambiamento estetizzante di polarità già tra loro avverse. Nelle mani di larghe fette di cittadinanza poste al centro della società senza esserne né guardiani né fruitori totalmente passivi, tutti gli strumenti di connessione digitale sembrano meri oggetti posti a “riempire il tempo”, o a organizzare l'esistenza telematica, a offrire *virtue signalling*, un ricco passatempo come una distruttiva ossessione e non sembrano ciò che effettivamente sono: ossia la propaggine dell'organizzazione sociale che perpetuamente alimenta divisione e accumulazione di capitale.

Esiste un'intelligenza artificiale vista “da giù”, e non è quella di produttori, ideatori né, al contempo, dei lavoratori impiegati – talvolta, spessissimo, sfruttati nelle diverse zone del mondo dove avviene la prima concreta estrazione dei materiali che servono ai nostri oggetti: l'IA è *prima di tutto* una filiera industriale – nella sua filiera mondializzata: è quella dei fruitori “medi”, tutto sommato “comuni”, a cui è stato detto che senza certi tipi di fruizione digitale avrebbero perso financo lo status di cittadini – o meglio, l'accesso virtuale ai servizi (sempre più virtuali, nel duplice senso) che una volta erano

<sup>13</sup> A. Burgio, *Il sogno di una cosa*.

la prima marca di esistenza della forma Stato socialdemocratica. Ed esiste un'intelligenza artificiale "vista da su", ossia dai grandi attori del settore – in alleanza e, al contempo, in concorrenza tra loro per il vero oggetto del contendere: il tempo di tutti noi, la nostra attenzione, lo sfruttamento e la connessione di ogni secondo della nostra vita, fino al grande sogno: farla finita con il sonno per poter produrre *sempre* – da chi da decenni e sempre più accumula capitale, dati, informazioni, possesso dei mezzi di produzione del campo digitale stesso, potere politico sempre meno indiretto, condizionando ormai le sorti economiche, politiche, sociali della specie.

Questa *rivoluzione* non l'hanno vinta tutti. L'hanno vinta coloro che hanno rinforzato i loro ruoli apicali, esercitando un inusitato dominio su dati, psiche e tempo libero della stragrande parte di popolazione che invece ha introiettato la rivoluzione passiva.

### 3. *Critica e azione (post) politica*

#### 3.1

La possibilità (coattiva) della connessione costante scorre parallela a una sistematica perdita del senso e a forme sempre più *accelerate* di esistenza. Accelerazioni che sono legate alla continua fisiologica richiesta del sistema sociale a produrre costanti *escalation*, poiché è la stabilizzazione dinamica a caratterizzarla in quanto modalità unica di produzione d'ordine<sup>14</sup>. Per seguire qui Hartmut Rosa, si accelera perché la crisi permanente di cui il capitalismo si nutre necessita dell'accelerazione costante non per generare sviluppo bensì per

<sup>14</sup> H. Rosa, *Beschleunigung und Entfremdung*.

mantenersi in piedi e, soggettivamente, produce il timore di restare indietro, *disconnessi* dal flusso<sup>15</sup>.

Si arriva qui a uno dei punti nodali del rapporto tra iper-sviluppo tecnologico, vissuto soggettivo e politica come luogo dell'abitare e del pensare un controllo sulle condizioni sociali che non sia un semplice *adattamento* alla verità non criticabile del sistema capitalistico e non sia insomma pura reazione: *chi governerà la condizione umana?*

Se la plasticità dell'ente ne permette la continua surrettizia modifica, se queste modifiche ristrutturano *le condizioni* della condizione umana, ossia la capacità di essere esistenza culturale e non solo biologica di *anthropos*, se tali modifiche sono apportate da sviluppi tecnologici prima inimmaginabili che impattano sulla dinamica psichica e cerebrale producendo così nuove forme antropiche, nuovi *ethos*, una nuova *Weltanschauung*, tali nuove forme di abitare la Storia come possono essere governate? Formulando tale quesito si esula dalla pura descrizione e si entra nel campo della prescrizione – ambizione filosofica la cui rinuncia comporta la cessazione dell'ambizione del pensiero di essere *forza* – e nella capacità di immaginare al contempo dei *possibili* adatti al tempo che viviamo e che tematizzino l'accelerazione tecnologica e il suo ristrutturarsi algoritmico, oggi in palese contraddizione con *i tempi lenti della democrazia* e dunque con gli schemi associativi del lungo Novecento. Perché, in fondo, lo scarto sembra essere tutto qui: la politica come forma organizzata della strategia persistentiva degli enti è ancora in grado di farsi *forza reale* di governo delle cose umane digitalizzate o invece è ormai destinata a essere solo – ed esclusivamente, questo il punto – ancella della più spregiudicata forma di realismo mai conosciuta, ossia quella per cui la realtà che si ha oggi innanzi è immodificabile, vera ed

<sup>15</sup> V. Costa, *La società dell'ansia*; B.-C. Han, *Im Schwarm. Ansichten des Digitalen*.

eterna, puro oggetto contro cui ci si scontra rimbalzando indietro, distrutti ma senza modifiche apportate al reale? Siamo, adesso, in un momento in cui il pensiero non può che rincorrere confusamente il reale, una Nottola di Minerva in costante affanno dinanzi alle innovazioni tecnologiche e alle ricadute politiche e geopolitiche pressoché quotidiane<sup>16</sup>.

Se il quadro è questo, se le forme democratiche scontano la stanchezza di forme storiche non più produttive di senso rispetto a saldature tra poteri prima meno pervasivi e tenuti a bada e a ibridazioni tecnologiche del tutto inusitate se, più in generale, è la capacità potenziale dell'*immaginazione pratica* del pensiero politico il *grande rimosso*, torna appunto la domanda: chi governerà la spaesata condizione umana? S'intenda la domanda in ambito ampio, sia come "egemonia" tra saperi – tra i quali il pensiero umanistico sembra aver rinunciato a concorrere – ma anche, nella *praxis*, chi, come e cosa governerà il rapporto soggetti-lavoro, simbolicamente e strutturalmente, in un tempo in cui sempre più è depotenziata l'aggregazione della piazza e dal "basso", le forme intermedie sono ormai liquidate e le forme d'unione sono pensabili solo come virtuali e dunque soggette al ricatto stesso dell'estrazione del valore e del dato?

Senza una radicale prassi di organizzazione di nuovi strumenti di partecipazione democratica il rischio – che è già pienamente attualità – è quello di poter ricorrere esclusivamente al placebo dell'atomistica idea dell'*autocontrollo* o dell'*auto-disciplina* al potere invasivo della strumentazione digitale e delle nuove tecnologie, con i loro calcoli, la loro proprietà non contendibile, le loro costrizio-

<sup>16</sup> Il presente saggio è ultimato nei giorni dell'avvento sul mercato di *DeepSeek* e delle annesse contese e riflessioni politiche, geopolitiche ed economiche sulle IA e il loro collegamento ai sistemi sociali e di produzione.

ni. Pratiche di tecnologia del sé che, ancorché comprensibili, sono pressoché sempre reattive, poiché contribuiscono all'idea di una rassegnazione complessiva: il mondo *così com'è* è il Reale; è la cosa in sé che si dà, a priori, senza alcuna mediazione e dunque possibilità di rovesciamento. La configurazione tecnologica ed economica così com'è è un orizzonte non trascendibile; e dunque si tratta, al meglio, di strutturarsi con una buona dose di etica stoica e *moderare* le passioni, aspettando che l'età di mezzo passi.

È una posizione da discutere – chi scrive la rifiuta – tenendo però a mente che il compito del pensiero è non accordare mai nulla di *scontato* a ciò che è, o a ciò che è concepito e pensato come intoccabile. Si comprende da questo punto di vista come sia una strategia politica e “aziendale” anche quella di lasciare il soggetto solipsista innanzi a un problema dato (in questo caso, quello dell'impatto del digitale e delle funzioni che identifichiamo come IA), velando completamente che si è dinanzi a un progetto politico che ha richiesto e richiederà investimenti militari-economici-tecnologici cospicui che bypassano completamente il singolo, sul quale ricade a mo' di *trickle-down economy* un qualche dividendo composto dell'oggetto che contribuirà a irregimentarlo e controllarlo; e che solo una radicale contesa tramite organizzazioni e ri-associazioni comuni potrebbe, forse, mettere in discussione lo stato di cose presenti.

### 3.2

Non solo struttura economica.

Se per fare cessare o modificare radicalmente le relazioni umane è sufficiente staccare la corrente in qualche centro dati sperduto risulta ovvio come la configurazione dei rapporti tra soggetti divenga qualcos'altro da ciò che il pensiero occidentale ha conosciuto, Altro

da sé – inevitabilmente, rendendo quasi inservibile tutto il lessico filosofico e politico della tradizione occidentale.

Ancora, se la memoria non è più *interna* bensì totalmente (e letteralmente) esterna, cioè appaltata alle strutture fisiche, ai dischi rigidi così come alle gigantesche macchine inquinatrici che materialmente incamerano i nostri dati, gusti, appuntamenti, opinioni, la riconfigurazione delle coscienze e dei vissuti segue necessariamente, poiché è nella memoria, nella persistenza dell'Io che riconosce una storia prospettica del sé, che si dà l'identità unitaria di ciò che possiamo chiamare *soggetti*. Il continuo demandare allo strumento il ruolo memoriale scarnifica la funzione della memoria, non ritenuta "necessaria", poco affidabile, poco allenata a riconoscersi. Non parliamo solo del dato già richiamato dei cambiamenti strutturali delle zone cerebrali non più stimulate, ma di un più radicale mutamento dei modi con i quali ci si mette in dialogo con la perdurante voce interna, se non con qualsiasi voce.

La sovranità del dato immagazzinato altrove e l'abdicazione della memoria sembrano realizzare l'incubo del Bradbury di *Fahrenheit 451*, con pochi tedorofori della memoria intenti a imparare *par coeur* i classici e il resto dell'umanità a dimenticare per sempre le opere che avevano contribuito a rendere *umani*. Si tratta di salvare il da salvare senza perdere in alcun modo il ruolo testimoniale delle coscienze. Scenario catastrofico, volutamente ingigantito, qui, ma che impone di ripensare le forme tramite cui ci riconosciamo quali soggetti psicologici e sociali, *sine ira et studio*.

L'alternativa è il collasso delle identità politiche e individuali, la sparizione di ogni forma intramondana di intendere l'abitare il mondo sociale, la vaporizzazione di ogni istanza comunitaria. In questo, si può sostenere che la politica non ha in alcun modo inteso il complesso sentimento di disperazione sociale diffuso; non ha voluto comprendere se le forme dell'ibridazione tecnologica pervasiva favo-

rissero, scoraggiassero o, al limite, fossero neutrali, rispetto all'idea di società che si vuole difendere; e ha semplicemente indicato come retri vi rigurgiti di plebeismo le domande di protezione sociale che le tornate elettorali sullo scenario occidentale hanno evidenziato e che erano anche legate a un sentimento di mancanza totale di controllo dei singoli rispetto alle nuove configurazioni degli habitat socio-digitali e a una pressione selettiva sempre più insopportabile, tutto questo mentre si liquidavano le forme di welfare in grado di permettere ai sottomessi di reggere il colpo degli scarichi energetici della globalizzazione. Le funzioni politiche che sono state riattivate e al contempo spregiate – *Stato, popolo e sovranità*, per esempio – sono dunque da considerare solo vestigie di una storia passata – e continuano a campeggiare nelle Costituzioni d'Occidente per mero formalismo – o sono ancora possibili vettori di *controllo, protezione e formulazione di senso* intorno al rapporto tra potere democratico e pervasiva e incontrollabile trasmutazione digitale e tecnologica? Ultime forme stabili pensabili dopo la liquidazione (dal pensiero) delle *classi* e del pensiero che esistano modi e forme di produzione, e tanto più così spregiate dal discorso mediatico o forme archeologiche di intendere la co-abitazione umana ai tempi del dominio di forme ed entrate politiche largamente sovra-nazionali e sovra-costituzionali?

Cosa ne sarà delle associazioni politiche ibridate che oggi conosciamo? Cosa dell'ibridazione delle forme che conosciamo con nuovi ruggenti poteri privati, nuove (ma non nuovissime) oligarchie finanziarie? Cosa dei sistemi monetari come li conosciamo dinanzi alla proliferazione (e alle contese geopolitiche annesse) di criptovalute<sup>17</sup> e *blockchain* legate alle nuove tecnologie? E cosa della forma Stato, al contempo, scomparsa e onnipresente, ferina e chiamata al controllo,

<sup>17</sup> E. Makdoun, *Geopolitica delle criptovalute*.

spregiata in Occidente così come pienamente attiva nel fu “terzo mondo”, in grado ancora di muover guerra ma che nel muover guerra deve chiedere sostegno a privati di immenso potere?

Una politica che oggi non è più in grado di stabilire un *Nomos* epocale è destinata a bruciarsi, come si sono bruciati i pochi movimenti politici recenti a tasso di innovazione una volta avuto accesso al potere; siamo immersi in un tempo anomico, in cui l'avvento di un certo modo astratto d'intendere la tecnologia e le relazioni su di esse basate cerca tiepidamente di lenire la paura della disconnessione sociale; un tempo in cui qualsiasi garante simbolico di un ordine che vada oltre il consumo individuale va liquidato – e così la *Sovranità* come potere in grado di rettificare le storture dell'*homo oeconomicus* diviene spauracchio, e così il *popolo* che tutti siamo diventa un cane morto della storia, e così lo *Stato* è amputato ma ancora sa essere pervasivo –, cedendo lo scettro a una insostenibile narrazione per cui l'individuo sovrano di sé stesso, solo, ha da cercare la propria salvezza individuale, e nessuna matrice può essere distrutta.

Non resta che chiedersi, a margine di questa critica filosofica quali sono i possibili spazi di azione per una salda postura filosofica che abbia ambizione, estrema, disperata forse, di essere una voce di indirizzo minimo nel marasma sociopolitico. Proprio dalla politica, o dalla mancanza della stessa – dalla post-politica – bisogna partire. Senza voler semplicemente civettare con il prefisso *post*, appare chiaro come si sia ormai di fronte a forme *successive* di democraticità; Colin Crouch ha ben descritto il fenomeno della postdemocrazia<sup>18</sup>, ma si tratta di ben intendere se tale configurazione non sia anch'essa una forma ormai *degenerata*, o se la quantità di cambiamenti non abbia portato con sé un ennesimo deterioramento che fa sì che la stessa *post-*

<sup>18</sup> C. Crouch, *Post-Democracy. After the Crisis*.

*democrazia* altro non sia ormai che puro *flatus vocis*. Probabilmente la domanda deve essere ancora più radicale: siamo forse dinanzi a forme talmente estreme di “*post*” – *post-politica* – in cui è il senso stesso della decisione collettiva, agonistica e deliberata a essersi smarrito, stritolato tra *pilota automatico, tecnici e tecno-incubi*, tra scomparsa della mediazione e attacco a qualsivoglia forma di organizzazione contraria all’impianto capitalista e neoliberista dell’esistenza? Scompare dietro la tecnica (economica, tecnologica, con ambizioni di colonizzazione spaziale) altro non è che una forma estrema di politicizzazione velata: sì, la *post-politica* tecnicizzata – a sua volta una tecnica della politica – riattiva le forme estreme del Politico, fa emergere i contrasti tellurici che esclusivamente una forma radicale di *democrazia post-bellica* aveva saputo sedare e ammansire, riporta in auge demoni che sembravano sconfitti, riattivati proprio dall’idea che non esista altro che il reale com’è, che non esista politica poiché le scelte sono parametri oggettivi e proporzioni sancite in eterno nei trattati, che a decidere debbano essere i nuovi *consiglieri* accelerati, che la politica non serve a migliorare l’esistente – il mondo qual è ma anche quale può essere – ma esclusivamente a ratificare scelte tutte basate su un’idea antropologica negativa, su un’idea *austeritaria* non solo dell’amministrazione bensì dell’esistenza – nichilismo realizzato – e sugli interessi dei possessori dei mezzi di produzione del campo informativo e digitale. La *post-politica* riattiva il Politico, questo è convincimento di chi scrive. E il Politico, oggi, può avere solo caratteri distruttivi.

### 3.3

Oggi più che mai, ciò che invece la filosofia può offrire è una possibilità di teoria critica di comprensione dell’immagine del mondo tramite un realismo problematico, farsi vettore (inascalato) dell’esi-

genza di continuare a porre la questione dei *valori* e delle eccedenze simboliche che non possono essere esaurite solo dal vortice infernale *produzione-consumo*; infine, tentativo estremo e sempiterno di terapia sociale ed esistenziale, foss'anche solo tramite un linguaggio ormai completamente vuoto al mondo e dunque nel tentativo di costruirne un altro, nonostante il ritardo sulle cose. Interpretare per cambiare, e cambiare anche per re-interpretare, e sennò smettere di esistere.

Ricordare dunque, sempre, che la filosofia è degna quando non è amica della disperazione.

## Bibliografia

- G. Anders, *Der Mensch ist antiquiert. Über die Seele im Zeitalter der zweiten industriellen Revolution*, München, Hanser Verlag, 2017; trad. it. *L'uomo è antiquato. Considerazioni sull'anima nell'epoca della seconda rivoluzione industriale*, Torino, Bollati Boringhieri, 2007.
- D. Andler, *Intelligence artificielle, intelligence humaine: la double énigme*, Paris, Gallimard, 2023; trad. it *Il duplice enigma. Intelligenza artificiale e intelligenza umana*, Torino, Einaudi, 2024.
- A. Aresu, *Geopolitica dell'intelligenza artificiale*, Milano, Feltrinelli, 2024.
- G. Balbi, *L'ultima ideologia. Breve storia della rivoluzione digitale*, Roma-Bari, Laterza, 2023.
- F. Balestrieri - L. Balestrieri, *Tecnologie dell'impero. AI, quantum computing, 6G e la nuova geopolitica del potere*, Roma, LUISS University Press, 2024.
- V. Bertola - S. Quintarelli, *Internet fatta a pezzi. Sovranità digitale, nazionalismi e big tech*, Torino, Bollati Boringhieri, 2023.
- A. Burgio, *Il sogno di una cosa. Per Marx*, Roma, DeriveApprodi, 2018.
- F. Cabitza - L. Floridi, *Intelligenza artificiale. L'uso delle nuove macchine*, Milano, Bompiani, 2021.
- M. T. Catena, *Breve storia del corpo*, Milano-Udine, Mimesis, 2020.
- V. Costa, *La società dell'ansia*, Roma, Inschibboleth, 2024.
- K. Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, New Heaven, Yale University Press, 2021; trad. it. *Né intelligente né artificiale. Il lato oscuro dell'IA*, Bologna, Il Mulino, 2023.

- N. Cristianini, *La scorciatoia. Come le macchine sono diventate intelligenti senza pensare in modo umano*, Bologna, Il Mulino, 2023.
- C. Crouch, *Post-Democracy After the Crisis*, Hoboken, Wiley and Sons, 2020.
- U. Eco, *Apocalittici e integrati: comunicazioni di massa e teorie della cultura di massa*, Milano, Bompiani, 1964.
- L. Floridi, *The Fourth Revolution. How the Infosphere is Reshaping Human Reality*, Oxford, Oxford University Press, 2014; trad. it. *La quarta rivoluzione. Come l'infosfera tra trasformando il mondo*, Milano, Raffaello Cortina Editore, 2017.
- M. Foucault, *Securité, territoire, population*, Paris, Seuil-Gallimard, 2004; trad. it. *Sicurezza, territorio, popolazione*, Milano, Feltrinelli, 2007.
- C. Galli, *Ideologia*, Bologna, Il Mulino, 2022.
- G. Gigerenzer, *How to Stay in a Smart World. Why Human Intelligence Still Beats Algorithms*, Cambridge, The MIT Press, 2022; trad. it. *Perché l'intelligenza artificiale batte ancora gli algoritmi*, Milano, Raffaello Cortina Editore, 2023.
- Global Changes Ethics, Politics and Environment in the Contemporary Technological World*, edited by Valera - J.C. Castilla, Switzerland, Springer, 2020.
- B.-C. Han, *Im Schwarm. Ansichten des Digitalen*, Berlin, MSB Matthes&Seitz, 2013; trad. it. *Nello sciame. Visioni del digitale*, Milano, Nottetempo, 2015.
- B.-C. Han, *Psychopolitik. Neoliberalismus und die neuen Machttechniken*, Frankfurt am Main, Fischer, 2014; trad. it. *Psicopolitica. Il neoliberalismo e le nuove tecniche del potere*, Roma, Nottetempo, 2016.
- Intelligenza Artificiale. Cos'è davvero, come funziona, che effetti avrà*, a cura di S. Quintarelli, Torino, Bollati Boringhieri, 2020.
- B. Labatut, *The Maniac*, London, Penguin, 2023; trad. it. *Maniac*, Milano, Adelphi, 2023.
- A. Leroi-Gourhan, *Le Geste et la Parole. Volume 1: La mémoire et la vitesse*, Paris, Albin Michel, 1964; trad. it. *Il gesto e la parola. Volume 1: la memoria e la velocità*, Milano, Mimesis, 2018.
- E. Makdoun, *Geopolitica delle criptovalute*, Roma, Castelvecchi, 2024.
- H. Marcuse, *Eros and Civilization: A Philosophical Inquiry into Freud*, Boston, Beacon, 1955; trad. it. *Eros e civiltà*, Torino, Einaudi, 1964.
- T. Numerico, *Big data e algoritmi. Prospettive critiche*, Roma, Carocci Editore, 2021.
- T. Numerico, *Dobbiamo ripensare l'intelligenza artificiale*, in «Limes», 12, 2022, pp. 73-80.
- M. Pasquinelli, *The Eye of the Master. A Social History of Artificial Intelligence*, London-New York, Verso Books, 2023.
- G. Roncaglia, *L'architetto e l'oracolo. Forme digitali del sapere da Wikipedia a Chat-GPT*, Roma-Bari, Laterza, 2023.

Mario Cosenza

- H. Rosa, *Beschleunigung und Entfremdung. Entwurf einer kritischen Theorie spätmoderner Zeitlichkeit*, Frankfurt am Main, Suhrkamp, 2013; trad. it. *Accelerazione e alienazione. Per una teoria critica del tempo nella tarda modernità*, Torino, Einaudi, 2015.
- H. Rosa, *Resonanz. Eine Soziologie der Weltbeziehung*, Berlin, Suhrkamp, 2016.
- F. Stalder, *Kultur der Digitalität*, Berlin, Suhrkamp, 2016; trad. en. *The Digital Condition*, Cambridge, Polity Press, 2018.
- G. Tamburrini, *Etica delle macchine. Dilemmi morali e intelligenza artificiale*, Roma, Carocci Editore, 2020.
- The Onlife Manifesto. Being Human in a Hyperconnected Era*, edited by L. Floridi, London, Springer, 2009.
- S. Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, London, Profile Books Ltd, 2019; trad. it. *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, Roma, LUISS University Press, 2023.

Mario Cosenza

Università degli Studi di Napoli Federico II

mario.cosenza@unina.it

<https://orcid.org/0000-0002-2462-7970>

ROBERTO DELLE DONNE

*L'editoria accademica, i dati della ricerca  
e l'intelligenza artificiale*

*Abstract:* The article examines the transformative impact of artificial intelligence (AI) and research data management on academic publishing, emphasizing global trends and the conservative stance of Italian publishers. It highlights initiatives such as Open Science, FAIR principles (Findable, Accessible, Interoperable, Reusable), and Text and Data Mining (TDM), discussing their challenges and potential for innovation. The role of Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG) in automating academic workflows, addressing ethical concerns, and promoting inclusivity is explored. The article concludes by advocating for equitable access to AI tools and data, aligning technology with the principles of cultural heritage preservation and knowledge democratization.

*Keywords:* Academic Publishing, Artificial Intelligence, Open Science, FAIR Data, Text and Data Mining, Large Language Models, Cultural Heritage, Knowledge Democratization.

*Introduzione*

L'editoria accademica sta attraversando una profonda trasformazione, guidata dall'integrazione di tecnologie avanzate come l'intelligenza artificiale (IA) e da un'attenzione crescente verso la gestione dei dati della ricerca. Tuttavia, questa evoluzione non è uniforme a livello globale. In Europa e su scala internazionale prevalgono approcci innovativi e collaborativi, mentre in Italia, dove predominano piccole e medie imprese spesso con risorse limitate per investire in nuove tecnologie, si registrano orientamenti più conservativi, anche

se non mancano eccezioni<sup>1</sup>. Ne è un esempio la lettera inviata di recente dall'Associazione Italiana Editori (AIE) ai rettori delle università italiane<sup>2</sup>, in cui emerge un approccio prevalentemente difensivo rispetto ai cambiamenti indotti dalla digitalizzazione e dal rapido affermarsi dell'intelligenza artificiale. Questo orientamento si discosta dalle tendenze europee e internazionali, che adottano un approccio proattivo orientato a promuovere modelli di Open Science, a garantire la sostenibilità tramite l'accesso aperto ai dati e a integrare tecnologie digitali avanzate per ottimizzare e automatizzare i processi editoriali e di gestione dei dati, favorendo così una condivisione più equa e inclusiva della conoscenza scientifica. Un aspetto cruciale di queste tendenze è il *Text and Data Mining* (TDM)<sup>3</sup>, una tecnica utilizzata da tempo per alimentare motori di ricerca, social network, piattaforme di vario genere e strumenti di ricerca bibliografica, come i *discovery tools* usati nelle biblioteche. Tuttavia, il TDM è oggi al centro di criticità diverse dal passato perché viene impiegato anche

<sup>1</sup> Per il panorama internazionale di veda Ruediger Wischenbart Content and Consulting, *Global 50 Publishing Ranking 2023*: <<https://www.wischenbart.com/product/global-50-publishing-ranking-2024/>>. Per quello italiano AIE. Ufficio Studi, *Rapporto sullo stato dell'editoria in Italia 2024*.

<sup>2</sup> Lettera circolare inviata dall'AIE, a firma di I. Cipolletta e M. Messina, ai rettori delle università italiane il 23 dicembre 2024: «Scriviamo oggi a tutti i Rettori delle università italiane perché sentiamo l'esigenza di avviare un dialogo sul tema della tutela e della valorizzazione del diritto d'autore in ambito universitario. La digitalizzazione e, di recente, il rapido affermarsi di strumenti basati sull'intelligenza artificiale rendono ogni giorno più complessa la gestione di entrambi gli aspetti. In relazione alla tutela, se nel mondo analogico era sufficiente combattere il fenomeno delle fotocopie pirata, oggi le violazioni con cui confrontarsi sono più complesse: dalla messa a disposizione in rete di intere opere fino al commercio di riassunti o slide realizzate senza il consenso degli aventi diritto, che non di rado – ci è stato segnalato – travisano il pensiero degli autori».

<sup>3</sup> Per una panoramica: F. L. Kleinkopf, *Text- und Data-Mining*.

per allenare l'intelligenza artificiale, in particolare quella generativa, da parte di aziende private che sfruttano dati pubblici e privati, dando luogo a un ampio ventaglio di questioni etiche e regolamentari<sup>4</sup>. Inoltre, il TDM per l'intelligenza artificiale è divenuto un segmento di mercato sempre più rilevante per molte imprese.

## 1. *La Scienza Aperta, i dati della ricerca e l'editoria*

### 1.1. Principali iniziative europee e internazionali per la Scienza Aperta

La Commissione Europea e i principali enti finanziatori internazionali hanno promosso negli ultimi anni numerose iniziative per incentivare la scienza aperta e l'utilizzo dei dati condivisi<sup>5</sup>. È impossibile ricordarle tutte, ma tra esse spiccano:

- *Horizon Europe e il Programma per l'Open Science*. Horizon Europe pone un forte accento sull'accesso aperto alle pubblicazioni e sui principi FAIR (*Findable, Accessible, Interoperable, Reusable*) per i dati della ricerca. Questo include l'obbligo di depositare i dati in repository aperti e di utilizzare licenze che ne consentano il riutilizzo<sup>6</sup>.

<sup>4</sup> Per queste ultime si veda il *Regolamento (UE) 2024/1689 del Parlamento Europeo e del Consiglio* del 13 giugno 2024: <[https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=OJ:L\\_202401689](https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=OJ:L_202401689)>.

<sup>5</sup> Per una panoramica sia consentito rimandare a R. Delle Donne, *Open Access, Biblioteche digitali e Big Data*.

<sup>6</sup> Commissione Europea, *Your Guide to Open Science in Horizon Europe*: <[https://intellectual-property-helpdesk.ec.europa.eu/publications/your-guide-open-science-horizon-europe\\_en](https://intellectual-property-helpdesk.ec.europa.eu/publications/your-guide-open-science-horizon-europe_en)>.

- *EOSC (European Open Science Cloud)*. Una piattaforma digitale che mira a fornire agli scienziati europei un’infrastruttura per archiviare, condividere e accedere ai dati della ricerca in modo interoperabile e sicuro<sup>7</sup>.
- *Plan S*. Lanciato da cOAlition S, un consorzio di enti finanziatori della ricerca, Plan S richiede che tutti i risultati della ricerca finanziata siano pubblicati in modalità open access o resi immediatamente disponibili attraverso repository aperti<sup>8</sup>.
- *Iniziative Nazionali e Transnazionali*. Oltre alla Commissione Europea, enti come il Wellcome Trust e il National Institute of Health (NIH) negli Stati Uniti stanno spingendo per politiche simili, evidenziando l’importanza dei dati aperti per accelerare l’innovazione scientifica<sup>9</sup>.

Queste iniziative non solo promuovono una maggiore trasparenza e collaborazione nella ricerca, ma rappresentano anche un modello per gli stati membri e altre organizzazioni internazionali interessate a favorire l’accesso equo e il riutilizzo dei dati scientifici<sup>10</sup>.

<sup>7</sup> Commissione Europea, *European Open Science Cloud (EOSC)*: <[https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/open-science/european-open-science-cloud-eosc\\_en](https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/open-science/european-open-science-cloud-eosc_en)>.

<sup>8</sup> *Plan S. Making full and immediate Open Access a reality*: <https://www.coalition-s.org/about/>. Lanciato nel 2018 ha subito nel tempo numerose messe a punto: cOAlitionS Office, *Annual Review 2023. Accelerating Open Access, 2023*: <<https://www.coalition-s.org/wp-content/uploads/2024/02/PlanS-2023-annual-review.pdf>>.

<sup>9</sup> Wellcome Trust, *Open Access Policy*: <<https://wellcome.org/grant-funding/guidance/open-access-guidance/open-access-policy>>; National Institute of Health (NIH), *What is the NIH Public Access Policy?*: <<https://www.nih.gov/health-information/nih-clinical-research-trials-you/what-is-nih-public-access-policy>>.

<sup>10</sup> Va ancora ricordato almeno il sostegno dato da *United Nations Educational, Scientific and Cultural Organization* (UNESCO) e da *European University Association* (EUS) all’Open Science: UNESCO, *UNESCO Recommendation on Open*

## 1.2. La grande editoria internazionale e i dati della ricerca

I grandi editori internazionali, come Elsevier, Springer Nature e Wiley, hanno sviluppato iniziative innovative per adattarsi alle nuove esigenze dell'editoria accademica e alla crescente attenzione per la gestione dei dati della ricerca:

- *Elsevier*. Attraverso piattaforme come Mendeley Data, Elsevier offre soluzioni per l'archiviazione, la condivisione e la gestione dei dati della ricerca<sup>11</sup>. L'integrazione con strumenti come DataSeer aiuta i ricercatori a identificare e preparare i dati per l'accesso aperto<sup>12</sup>.
- *Springer Nature*. L'editore ha lanciato guide per i ricercatori sull'adozione dei principi FAIR e offre servizi per il deposito dei dati attraverso partnership con repository riconosciuti, come figshare<sup>13</sup>.
- *Wiley*. Ha sviluppato una serie di strumenti per il supporto alla gestione dei dati e collabora con enti come Dryad per garantire la disponibilità e la conservazione a lungo termine dei dataset<sup>14</sup>.

*Science*, <<https://www.unesco.org/en/open-science/about?hub=686>>; EUA, *The EUA Open Science Agenda 2025*: <<https://www.eua.eu/publications/positions/the-eua-open-science-agenda-2025.html>>.

<sup>11</sup> *Mendeley Data*: <<https://data.mendeley.com/>>.

<sup>12</sup> *About DataSeer. Uncover. Advise. Monitor*: <<https://dataseer.ai/about/>>.

<sup>13</sup> Springer Nature, *Research Data Policy*: <<https://www.springernature.com/gp/authors/research-data-policy>>; Springer Nature, *Figshare Integration. Guidance for Authors*: <<https://www.springernature.com/gp/authors/research-data/figshare-integration>>.

<sup>14</sup> Wiley, *Open Data. Maximizing the value of research*: <<https://authorservices.wiley.com/author-resources/Journal-Authors/open-research-policies/open-data/index.html>>; Wiley, *Data Sharing with Dryad*: <<https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-service.html>>.

- *Taylor & Francis*. Fornisce risorse dettagliate per i ricercatori su come rendere i dati FAIR e ha collaborato con piattaforme come Code Ocean per integrare codice e dati all'interno delle pubblicazioni<sup>15</sup>.
- *Collaborazioni trasversali*. Molti editori partecipano a iniziative globali come Research Data Alliance (RDA) e ORCID per promuovere standard comuni e interoperabilità tra sistemi<sup>16</sup>.

Questi sforzi dimostrano come i grandi editori stiano cercando di consolidare la propria posizione non solo come distributori di contenuti, ma anche come facilitatori di pratiche innovative nella gestione e nella condivisione dei dati.

### 1.3. I dati della ricerca in Italia

In Italia, la gestione dei dati della ricerca è un tema cruciale, alimentato dalle politiche europee come Horizon Europe, dalle iniziative del MUR e, soprattutto, della CRUI (Conferenza dei Rettori delle Università Italiane).

Tra le principali iniziative italiane per promuovere il libero accesso ai dati della ricerca, vanno menzionati i bandi PRIN (Progetti di Rilevante Interesse Nazionale) e SIR (Scientific Independence of young Researchers) del Ministero dell'Università e della Ricerca (MUR). Questi

<sup>15</sup> Taylor & Francis, *Open Data Sharing Policy*: <<https://authorservices.taylorandfrancis.com/data-sharing-policies/open-data/>>; la lista dei data repository raccomandati, sia disciplinari come Code Ocean sia generalisti come FAIRsharing and re3data.org, è alla pagina: <<https://authorservices.taylorandfrancis.com/data-sharing/share-your-data/repositories/>>.

<sup>16</sup> L'elenco delle diverse collaborazioni di Research Data Alliance con editori privati ed enti è alla pagina: <<https://www.rd-alliance.org/organisation-directory/>>; per le collaborazioni con Orcid: <<https://orcid.org/members>>.

bandi prevedono dal 2017 linee guida specifiche per l'adozione di politiche di Open Science e per la gestione dei dati secondo i principi FAIR. Attraverso il finanziamento di progetti che incentivano la condivisione dei dati, il MUR promuove una maggiore trasparenza e accessibilità della ricerca, allineandosi alle *best practice* europee e internazionali.

A supporto di queste iniziative, il *Piano Nazionale per la Scienza Aperta*, recentemente introdotto dal MUR, rappresenta un quadro strategico fondamentale per la promozione dell'Open Science in Italia<sup>17</sup>. Il piano definisce obiettivi e strumenti per favorire la condivisione dei dati della ricerca, la trasparenza dei processi scientifici e la valorizzazione dei risultati attraverso l'adozione di infrastrutture digitali e politiche coordinate a livello nazionale. Questo approccio rafforza il ruolo dell'Italia nella comunità scientifica globale, creando sinergie tra enti di ricerca, università e industrie.

La CRUI promuove l'accesso aperto ai risultati della ricerca scientifica e l'Open Science da oltre venti anni, in modo molto incisivo. Più specificamente, il documento sulla gestione dei dati della ricerca prodotto dalla CRUI nel 2024 stabilisce linee guida fondamentali per supportare le università italiane nella transizione verso pratiche di Open Science e adesione ai principi FAIR<sup>18</sup>.

<sup>17</sup> MUR, *PNR - Programma Nazionale per la Ricerca*.

<sup>18</sup> La CRUI è dal 2004 il principale attore in Italia per la promozione dell'accesso aperto e della scienza aperta, attraverso le attività della Commissione Biblioteche, al cui interno operano il gruppo di contrattazione con gli editori CARE, coordinato da Nino Grizzuti, e il gruppo Open Access, poi divenuto nel 2020 Osservatorio sulla scienza aperta: R. Delle Donne, *L'accesso aperto, le università e le SSH*. L'Osservatorio sulla scienza aperta, da me coordinato dal 2020, ha pubblicato a ottobre 2024 il documento: *Policy sulla gestione dei dati della ricerca e Data Management Plan (2024)*, <<http://osa.crui.it/wp-content/uploads/2024/12/policy-gestionedati.pdf>>. In diversi contratti con i grandi editori commerciali sottoscritti da CRUI attraverso il gruppo CARE (Elsevier, Springer, Wiley,

I punti principali evidenziati nel documento sono:

- *Pianificazione e Policy Istituzionali*. Le università sono incoraggiate a sviluppare policy interne per la gestione dei dati, definendo ruoli, responsabilità e strumenti per garantire la sostenibilità dei dati stessi.
- *Data Management Plan (DMP)*. L'adozione di piani di gestione dei dati è centrale per strutturare il ciclo di vita dei dati della ricerca, dalla raccolta alla conservazione a lungo termine.
- *Infrastrutture Digitali*. Sebbene il documento non prescriva l'uso di repository specifici, evidenzia l'importanza di infrastrutture affidabili e interoperabili, raccomandando l'utilizzo di piattaforme riconosciute a livello nazionale e internazionale.
- *Formazione e Supporto*. Si sottolinea la necessità di formare i ricercatori e il personale tecnico per implementare le pratiche di gestione dei dati in linea con gli standard internazionali.

Queste indicazioni riflettono un tentativo di colmare il divario tra le pratiche italiane e quelle più avanzate a livello internazionale, promuovendo una cultura della condivisione e della trasparenza nella ricerca.

#### 1.4. Il ruolo del Text and Data Mining

Il *Text and Data Mining (TDM)* rappresenta una delle tecnologie più promettenti per l'analisi avanzata dei dati della ricerca<sup>19</sup>. Questa tecnica consente di:

Taylor & Francis) sono previste clausole che consentono il TDM, anche se con alcune limitazioni per Elsevier e Springer.

<sup>19</sup> C. Zong - R. Xia - J. Zhang, *Text Data Mining*. Spunti di riflessione interessanti in R. Morriello, *IA ed editoria scientifica*.

- *Estrarre conoscenza complessa.* Il TDM è in grado di analizzare grandi volumi di dati testuali e numerici per identificare schemi, correlazioni e *trend* che altrimenti sfuggirebbero all'osservazione umana.
- *Supportare la scoperta scientifica.* Attraverso il TDM, è possibile accelerare il progresso scientifico: ad esempio, individuando nuove connessioni tra articoli scientifici, migliorando la diagnosi medica basata su grandi *dataset* clinici e ancora ottimizzando i processi industriali grazie all'analisi di brevetti.
- *Promuovere l'interdisciplinarietà.* L'integrazione di dati provenienti da discipline diverse permette di sviluppare nuove prospettive e soluzioni innovative.
- *Ottimizzare i processi editoriali.* Le redazioni e gli editori possono utilizzare il TDM per individuare lacune nella letteratura disciplinare già esistente e migliorare la qualità delle pubblicazioni. Nonostante il suo potenziale, l'accesso ai dati per il TDM è spesso regolato da politiche restrittive imposte dai grandi editori. Alcuni punti critici includono:
  - *Licenze specifiche.* Molti editori richiedono licenze aggiuntive per consentire il TDM sui contenuti da loro pubblicati, limitando l'accesso ai dati anche agli abbonati delle loro piattaforme<sup>20</sup>.
  - *Barriere tecniche.* Gli editori implementano spesso misure tecniche, come CAPTCHA o particolari limiti alla possibilità di scaricare dati, che ostacolano l'automazione necessaria per il TDM.
  - *Costi elevati.* Alcuni editori impongono costi significativi per consentire il TDM, rendendolo inaccessibile per ricercatori con budget limitati.

<sup>20</sup> Si veda la rassegna: Cranfield University. Library Services, *Publisher policies for TDM*.

Negli ultimi anni, alcune iniziative globali hanno cercato di rendere il TDM più accessibile, non senza alcune ombre:

- *Politiche dell’Unione Europea*. La direttiva sul diritto d’autore del 2019 ha introdotto eccezioni che consentono il TDM per scopi di ricerca scientifica, riducendo le restrizioni legali per istituzioni di ricerca e università. Questo rappresenta un passo significativo verso la democratizzazione dell’accesso ai dati per finalità scientifiche. In Italia, diversamente da altri paesi, le due eccezioni relative al TDM previste dalla Direttiva sono state correttamente recepite dal legislatore, che ne ha garantito l’applicazione anche in ambiti non strettamente legati alla ricerca accademica, posizionandosi come un esempio positivo nel panorama europeo<sup>21</sup>.
- *Partnership pubblico-privato*. Alcuni editori, come Elsevier e Springer Nature, hanno sviluppato politiche specifiche per il TDM, che sembrano venire incontro alle esigenze dei ricercatori, ma che in realtà suscitano preoccupazioni significative riguardo al controllo e alla gestione delle risorse di dati. Elsevier richiede che le istituzioni aggiornino i loro accordi di licenza per consentire l’accesso ai contenuti a fini di TDM, limitando tale accesso attraverso un’*Application Programming Interface* (API) proprietaria. Questo approccio è stato criticato perché impone condizioni restrittive e limita l’uso di tecniche comuni di *mining*, come l’uti-

<sup>21</sup> Sul recepimento in Europa: G. Vrakas, *A literature review of “lawful” text and data mining*. Per il recepimento in Italia: B. Calabrese, *Scientific TDM Exception and Communication to the Public*; D. De Angelis, *The Italian Implementation of the New EU Text and Data Mining Exceptions*. Più in generale, sulla Direttiva, si legga A. Musso, *Eccezioni e limitazioni ai diritti d’autore*; sul recepimento in Italia di un’altra eccezione fondamentale per il mondo della ricerca, relativa alla conservazione del patrimonio culturale: R. Maiello-G. Mazziotti-C. Sappa, *Qualche spunto sull’eccezione per la conservazione del patrimonio culturale*.

lizzo di robot o crawler. Inoltre, Elsevier richiede ai ricercatori di registrarsi come sviluppatori e ottenere un'APIKey, imponendo ulteriori vincoli operativi<sup>22</sup>. Springer Nature, invece, offre una varietà di API per consentire il TDM su contenuti come riviste, libri e altre risorse. Mentre i ricercatori possono scaricare materiali direttamente dalle piattaforme di Springer Nature, l'accesso tramite API dedicate comporta costi aggiuntivi. Per le istituzioni accademiche e governative che hanno sottoscritto abbonamenti, Springer Nature include generalmente i diritti di TDM senza ulteriori costi, prevedendoli nelle licenze di abbonamento. Tuttavia, queste condizioni di utilizzo, che si applicano in caso di nuovi contratti o di rinnovi di abbonamenti, comportano costi aggiuntivi per quegli utenti che richiedono invece l'accesso prima del rinnovo<sup>23</sup>.

In altri termini, le condizioni poste dagli editori finiscono con lo sbilanciare a loro vantaggio i rapporti con la comunità accademica, creando barriere economiche e operative che finiscono con il limitare le possibilità di molti ricercatori di utilizzo dei dati. È quindi essenziale che, in fase di contrattazione degli abbonamenti con gli editori, siano negoziati anche i diritti relativi all'uso del TDM; a maggior ragione, perché la direttiva EU 2019/790 e le norme di recepimento nazionale prevedono la nullità di patti contrari alle eccezioni previste sul TDM. A tal fine è necessario un dialogo con gli editori più equo e trasparente, per assicurare che il TDM resti uno strumento accessibile e inclusivo, evitando che si levino nuove barriere che ostacolino la democratizzazione della conoscenza. Inoltre, è importante sottolineare che l'uso di

<sup>22</sup> R. Van Noorden, *Elsevier opens its papers to text-mining*.

<sup>23</sup> Springer Nature, *Text and Data Mining*.

API proprietarie può complicare il processo di *mining*, rendendo più difficile per i ricercatori accedere ai dati in modo efficiente. Sarebbe auspicabile l'adozione di tecnologie open source che consentano di lavorare direttamente sui PDF, evitando le restrizioni imposte dalle soluzioni proprietarie impiegate dagli editori. In sintesi, mentre le partnership pubblico-privato nel campo del TDM possono offrire opportunità per l'innovazione, è fondamentale garantire che le politiche adottate non limitino l'accesso equo ai dati e non impongano restrizioni che ostacolino la ricerca scientifica.

- *Open Access e TDM*. Le pubblicazioni in open access rappresentano una risorsa cruciale per il TDM, dal momento che i contenuti sono liberamente accessibili e non soggetti alle limitazioni imposte dagli editori commerciali. L'adozione di politiche open access da parte di molte istituzioni accademiche e degli enti finanziatori della ricerca contribuisce a rendere il TDM uno strumento realmente inclusivo, riducendo le barriere economiche e promuovendo la diversità delle fonti disponibili per l'analisi. Tuttavia, persistono sfide legate alla qualità e all'interoperabilità dei dati, che richiedono un impegno costante nella definizione di standard comuni.

In conclusione, l'adozione del TDM continua a essere un terreno di confronto tra editori, ricercatori e *policy maker*, ma il suo potenziale per la scienza aperta e l'innovazione lo rende una componente cruciale per il futuro dell'editoria accademica. Nonostante il suo potenziale, il TDM continua a scontrarsi con ostacoli normativi e tecnologici. Anche in Italia, l'accesso ai dati per il TDM è spesso limitato dalle condizioni imposte dagli editori nazionali e internazionali.

## 2. LLM e RAG: le nuove frontiere della conoscenza

I “modelli linguistici di grandi dimensioni” (Large Language Models, LLM) e le tecniche di “generazione aumentata del recupero” dei dati (Retrieval-Augmented Generation, RAG) rappresentano innovazioni significative nell’ambito dell’intelligenza artificiale applicata all’editoria accademica e alla gestione della conoscenza. Questi strumenti combinano la potenza della comprensione linguistica avanzata con la capacità di accedere a dati esterni, creando risposte puntuali e altamente personalizzate. Gli LLM, come GPT di OpenAI o Llama di Meta, sono modelli di *machine learning* pre-addestrati su vasti *dataset* per comprendere e generare linguaggio naturale. Le tecniche di RAG combinano questi modelli con sistemi di recupero dei dati per generare risposte contestualizzate, migliorando la precisione e riducendo errori<sup>24</sup>.

I loro ambiti di applicazione nell’editoria sono:

- *L’automazione della produzione editoriale.* Gli LLM generano abstract, traduzioni e riassunti con notevole rapidità e precisione. I RAG li integrano con dati aggiornati, rendendo le risposte più pertinenti e accurate.
- *L’accesso personalizzato alla conoscenza.* Le tecniche di RAG consentono di filtrare e personalizzare i contenuti accademici in base alle esigenze degli utenti.
- *La peer review assistita.* Gli strumenti basati su LLM e RAG possono automatizzare il controllo della qualità e dell’originalità dei manoscritti.

<sup>24</sup> Per gli LLM: W. Xin Zhao-K. Zhou-J. Li et al., *A Survey of Large Language Models*. Per RAG: Y. Gao-Y. Xiong-X. Gao et al., *Retrieval-Augmented Generation for Large Language Models*.

Gli LLM rappresentano una delle innovazioni più potenti e controverse nell'ambito dell'intelligenza artificiale applicata alla conoscenza. Il loro utilizzo solleva infatti notevoli interrogativi, riconducibili ad alcune macrotematiche: a) epistemologia e verità, *bias* e inclusività; b) ruolo dell'umano. Inoltre, completano il quadro complessivo delle loro implicazioni le opportunità e le sfide operative, come l'efficienza (c) e la governance dei dati (d).

- *Bias, verità e inclusività nei modelli linguistici avanzati.* Gli LLM generano risposte probabilistiche, coerenti con i dati di addestramento, ma non sempre allineate ai principi tradizionali della verità scientifica, quali la riproducibilità e la revisione paritaria. Questo introduce rischi significativi, come la perpetuazione di *bias* e di disuguaglianze sistemiche. I *dataset* di addestramento, spesso influenzati da pregiudizi di genere, razza o classe sociale, amplificano queste distorsioni, marginalizzando contributi di minoranze o aree meno rappresentate. Per contenere tali rischi, è essenziale ampliare la diversità dei dati, sviluppare algoritmi correttivi e coinvolgere esperti di diverse discipline. In definitiva, sono sfide che sollecitano una riflessione critica sul ruolo della verità e della neutralità all'interno dei modelli linguistici avanzati. Ad esempio, un modello linguistico potrebbe generare risposte apparentemente corrette basate su dati distorti, e suggerire che una minoranza culturale abbia contribuito meno di altre alla produzione scientifica globale semplicemente perché i *dataset* di addestramento riflettevano una sua sottorappresentazione storica. Questo esempio illustra la necessità di progettare *dataset* sempre più inclusivi e algoritmi che sappiano riconoscere e compensare tali squilibri.
- *Collaborazione uomo-macchina: etica e responsabilità nell'innovazione.* L'efficienza e la precisione degli LLM, che rappresentano un prezioso supporto per analisi preliminari e automazione di

- processi ripetitivi, devono essere bilanciate da un controllo umano attento e rigoroso, soprattutto in ambiti che richiedono una comprensione critica e contestuale. Ad esempio, la supervisione umana potrebbe essere decisiva nell'identificare *bias* nascosti o risposte inesatte generate dai modelli, specialmente quando i *dataset* riflettono pregiudizi sistemici. Questa sinergia tra uomo e macchina non solo mitiga i rischi epistemologici, ma rafforza anche la fiducia nel loro utilizzo in contesti accademici e operativi.
- *Efficienza, scalabilità e sostenibilità.* Gli LLM possono migliorare significativamente l'efficienza di diverse attività accademiche ed editoriali, automatizzando processi come la revisione dei testi, la generazione di abstract e la traduzione dei contenuti. Tuttavia, la loro implementazione richiede infrastrutture avanzate, competenze specifiche e una gestione sostenibile dei costi energetici e dell'impatto ambientale.
  - *Governance dei dati e collaborazione globale.* Le sfide epistemologiche e sociali associate all'uso degli LLM evidenziano rischi significativi, rendendo necessaria una gestione dei dati trasparente ed equa, fondamentale per generare fiducia nelle comunità scientifiche e indurle all'adozione di queste tecnologie avanzate per l'analisi e la rielaborazione dei dati. La protezione della privacy e la conformità alle normative vigenti, come il GDPR in Europa, sono elementi chiave per costruire e mantenere tale fiducia. La governance dei dati deve garantire un accesso equo, evitando che il loro controllo si traduca in concentrazioni di conoscenza e di potere nelle mani di grandi attori tecnologici. Allo stesso tempo, l'intelligenza artificiale può favorire la condivisione della conoscenza e promuovere collaborazioni inclusive su scala internazionale, purché si affrontino le disuguaglianze e le disparità tecnologiche che persistono tra i diversi paesi.

### 3. *La conoscenza come bene comune e la memoria culturale*

Nell'affrontare il ruolo delle nuove tecnologie nella gestione e diffusione del sapere, diviene essenziale considerare il valore della conoscenza come bene comune e il suo legame indissolubile con la memoria culturale. Questa prospettiva invita a riflettere sul rapporto tra innovazione tecnologica e conservazione del patrimonio immateriale.

- *La conoscenza come bene comune e il paradosso della proprietà intellettuale.* L'adozione di tecnologie come LLM e RAG rinnova una tensione storica: la conoscenza scientifica intesa come bene comune opposta al controllo proprietario da parte di aziende e istituzioni<sup>25</sup>. Questo dibattito, che affonda le sue radici nell'Illuminismo, richiama la necessità di bilanciare l'apertura del sapere con la protezione dei diritti. Mentre i principi dell'*Open Science* promuovono la condivisione della conoscenza per il bene collettivo, la realtà tecnologica odierna è spesso dominata da brevetti e licenze restrittive che limitano l'accesso. Questo paradosso evidenzia la necessità di politiche globali che favoriscano la democratizzazione della conoscenza senza penalizzare la ricerca dell'innovazione.
- *Tecnologia e asimmetrie di potere.* Le tecnologie avanzate, come LLM e RAG, amplificano le disuguaglianze già esistenti nell'editoria accademica. La capacità di sfruttare appieno queste tecnologie dipende dall'accesso a risorse avanzate, come infrastrutture computazionali e *dataset* di alta qualità, che sono spesso monopolizzate da grandi aziende tecnologiche. Questa dinamica concentra ulteriormente il potere nelle mani di pochi attori, raf-

<sup>25</sup> Si rimanda al volume collettaneo *Understanding Knowledge as a Commons*.

forzando quella che Michel Foucault definirebbe una biopolitica del sapere<sup>26</sup>. Per contrastare tali asimmetrie, è necessario promuovere infrastrutture aperte e investimenti pubblici che garantiscano un accesso equo alle risorse tecnologiche.

- *Rischi e opportunità della mediaticità digitale.* Come ogni medium innovativo, LLM e RAG stanno ridefinendo il rapporto tra umano e sapere<sup>27</sup>. La velocità nella generazione di contenuti rappresenta un'opportunità per democratizzare l'accesso all'informazione, ma rischia di sacrificare la profondità analitica e la riflessione critica. Inoltre, la mediaticità digitale può favorire un consumo passivo della conoscenza, sostituendo la ricerca attiva con risposte sintetiche generate automaticamente. Per sfruttare appieno le potenzialità di queste tecnologie, è fondamentale integrarle con metodologie tradizionali che valorizzino il pensiero critico e l'interpretazione contestuale.
- *Il ruolo dell'editoria nella conservazione della memoria culturale.* L'editoria non è solo produzione di contenuti, ma anche conservazione della memoria collettiva. Le tecnologie come LLM e RAG possono rafforzare questa funzione facilitando la catalogazione, la ricerca e il recupero di informazioni, ma introducono anche nuovi rischi. L'automazione nella gestione dei contenuti potrebbe favorire la perdita di materiali ritenuti marginali o di

<sup>26</sup> M. Foucault, *Securité, territoire, population*.

<sup>27</sup> Già Marshall McLuhan, in *Understanding Media: The Extensions of Man*, aveva richiamato l'attenzione sul fatto che i media influenzino il nostro modo di percepire e interagire con il mondo, anticipando molte riflessioni sull'impatto che i nuovi strumenti tecnologici hanno sul rapporto tra umano e sapere. Più recentemente Shoshana Zuboff, in *The Age of Surveillance Capitalism*, ha analizzato l'impatto delle tecnologie digitali sulla società, mettendo in luce i rischi connessi alla mediaticità digitale di concentrazione del potere e di riduzione dell'individuo a consumatore passivo di dati e informazioni.

nicchia e considerati meno rilevanti da criteri algoritmici come rilevanza dei contenuti, popolarità, valutazioni di qualità, *bias* incorporati nei dati di addestramento, ordine di priorità, filtraggio di “rumori” ecc. Perché l’editoria possa continuare a contribuire alla trasmissione della memoria culturale, è fondamentale riconoscere il suo legame con i depositi della memoria, come biblioteche e archivi, che conservano e custodiscono il patrimonio culturale collettivo, come sottolineato da Aleida Assmann<sup>28</sup>. Per svolgere efficacemente questa funzione, l’editoria dovrebbe integrare strumenti tecnologici avanzati con una curatela umana capace di valorizzare le diverse prospettive storiche e culturali, evitando che i criteri algoritmici riducano la complessità del sapere a schemi standardizzati o limitino la rappresentazione delle voci marginali.

### *Conclusione*

In un contesto di rapida innovazione tecnologica, è fondamentale trovare un equilibrio tra automazione e supervisione umana. Gli LLM e le tecnologie di RAG offrono straordinarie opportunità per aumentare l’efficienza e democratizzare l’accesso alla conoscenza, ma il loro impiego deve essere guidato da una visione etica e sostenibile. Progetti infrastrutturali come European Open Science Cloud (EOSC), pur non utilizzando direttamente LLM o RAG, mostrano come la collaborazione tra editori, ricercatori e *policy maker* possa creare ecosistemi di dati che abilitano nuove applicazioni basate sull’intelligenza artificiale. EOSC, ad esempio, supporta la scienza

<sup>28</sup> A. Assmann, *Erinnerungsräume*.

aperta garantendo l'accesso a dati FAIR, una risorsa essenziale per addestrare tecnologie avanzate in modo trasparente e responsabile.

Le riflessioni sviluppate da Gino Roncaglia sull'intelligenza artificiale in un suo recente volume inducono a considerare il ruolo di questi strumenti non solo come tecnologie operative, ma anche come componenti essenziali di una trasformazione epistemologica<sup>29</sup>. Roncaglia evidenzia come l'IA generativa, con il suo approccio oracolare-predittivo, rappresenti sia una sfida sia un'opportunità per strutturare e aggregare conoscenza in un'"età della frammentazione". Il compito degli "architetti del sapere", di studiosi, educatori, bibliotecari, archivisti, editori e progettisti di sistemi informativi impegnati nella costruzione di infrastrutture e metodologie per organizzare e trasmettere conoscenza, diventa quindi cruciale per integrare strumenti di IA in modelli collaborativi e inclusivi, come quelli rappresentati da Wikipedia e dai Linked data, superando i limiti di un'informazione dispersa e frammentata.

In questa prospettiva, anche gli atenei possono e devono svolgere un ruolo chiave, promuovendo l'adozione di tecnologie digitali in linea con i principi della scienza aperta e favorendo un approccio critico al loro utilizzo. La capacità di coniugare l'efficienza degli strumenti di IA con metodologie che valorizzino il pensiero critico e l'interpretazione contestuale rappresenta una sfida centrale per evitare che la velocità della generazione automatica di contenuti sacrifichi la profondità analitica e la riflessione. È necessario progettare un futuro in cui i valori fondamentali della conoscenza – apertura, inclusione e sostenibilità – siano rispettati e rafforzati.

Solo attraverso iniziative che promuovano interoperabilità, equità e responsabilità sarà possibile costruire un futuro in cui l'intel-

<sup>29</sup> G. Roncaglia, *L'architetto e l'oracolo*.

ligenza artificiale diventi uno strumento davvero al servizio della conoscenza, favorendo non solo l'innovazione tecnologica, ma anche una comprensione più profonda e consapevole del sapere umano.

## Bibliografia

*Tutti gli URL sono stati verificati il 13 gennaio 2025*

- AIE. Ufficio Studi, *Rapporto sullo stato dell'editoria in Italia 2024 su consolidato 2023*, Milano, Associazione Italiana Editori, 2024.
- A. Assmann, *Erinnerungsräume: Formen und Wandlungen des kulturellen Gedächtnisses*, München, Verlag C.H. Beck, 1999; trad. it. *Ricordare. Forme e mutamenti della memoria culturale*, Bologna, il Mulino, 2002.
- B. Calabrese, *Scientific TDM Exception and Communication to the Public: Did Italians Do It Better... or at Least Not Worse?*, in «Journal of Intellectual Property Law & Practice», 17, 5, May 2022, pp. 399-406, <<https://doi.org/10.1093/jiplp/jpac026>>.
- I. Cipolletta - M. Messina, *Lettera circolare dell'AIE ai rettori delle università italiane*, 23 dicembre 2024.
- cOAlitionS Office, *Annual Review 2023. Accelerating Open Access*, 2023, <<https://www.coalition-s.org/wp-content/uploads/2024/02/PlanS-2023-annual-review.pdf>>.
- cOAlitionS, *Plan S. Making full and immediate Open Access a reality*, <<https://www.coalition-s.org/about/>>.
- Commissione Europea, *European Open Science Cloud (EOSC)*, Bruxelles, 2023, <[https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/open-science/european-open-science-cloud-eosc\\_en](https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/open-science/european-open-science-cloud-eosc_en)>.
- Commissione Europea, *Your Guide to Open Science in Horizon Europe*, Bruxelles, 2024, <[https://intellectual-property-helpdesk.ec.europa.eu/publications/your-guide-open-science-horizon-europe\\_en](https://intellectual-property-helpdesk.ec.europa.eu/publications/your-guide-open-science-horizon-europe_en)>.
- Cranfield University. Library Services, *Publisher policies for TDM*, <<https://library.cranfield.ac.uk/text-and-data-mining/publisher-policies>>.
- CRUI. Osservatorio sulla scienza aperta, *Policy sulla gestione dei dati della ricerca e Data Management Plan (2024)*, Roma, 2024, <<http://osa.crui.it/wp-content/uploads/2024/12/policy-gestionedati.pdf>>.
- DataSeer Research Data Service, *About DataSeer. Uncover. Advise. Monitor*, <<https://dataseer.ai/about/>>.

- D. De Angelis, *The Italian Implementation of the New EU Text and Data Mining Exceptions*, in «Communia», Dec 14, 2022, <<https://communia-association.org/2022/12/14/italian-implementation-of-the-new-eu-tdm-exceptions/>>.
- R. Delle Donne, *L'accesso aperto, le università e le SSH / Open Access, Universities and SSH*, in «Il capitale culturale», 17, 2018, pp. 17-45, <<https://doi.org/10.13138/2039-2362/1944>>.
- R. Delle Donne, *Open Access, Biblioteche digitali e Big Data*, in *I magnifici incontri 2018. Piano nazionale università digitale*, a cura di G. Bernardi - A. Tercig, Forum, Udine, pp. 87-96.
- Elsevier, *Mendeley Data*, <<https://data.mendeley.com/>>.
- European University Association, *The EUA Open Science Agenda 2025*, Bruxelles, 2023, <<https://www.eua.eu/publications/positions/the-eua-open-science-agenda-2025.html>>.
- M. Foucault, *Securité, territoire, population*, Paris, Seuil-Gallimard, 2004; trad. it. *Sicurezza, territorio, popolazione*, Milano, Feltrinelli, 2007.
- Y. Gao - Y. Xiong - X. Gao - K. Jia - J. Pan - Y. Bi - Y. Dai - J. Sun - M. Wang - H. Wang, *Retrieval-Augmented Generation for Large Language Models: A Survey*, in «ArXiv», 27 Mar 2024, <<https://doi.org/10.48550/arXiv.2312.10997>>.
- F. L. Kleinkopf, *Text- und Data-Mining. Die Anforderungen digitaler Forschungsmethoden an ein innovations- und wissenschaftsfreundliches Urheberrecht*, Baden-Baden, Nomos Verlagsgesellschaft, 2022.
- R. Maiello - G. Mazziotti - C. Sappa, *Qualche spunto sull'eccezione per la conservazione del patrimonio culturale*, in *Crisi e resilienza del diritto d'autore. Il recepimento italiano della direttiva 2019/790*, a cura di M.L. Bixio, Milano, Giappichelli, 2024, pp. 51-62.
- M. McLuhan, *Understanding Media: The Extensions of Man*, New York, McGraw-Hill, 1964; trad. it. *Gli strumenti del comunicare*, Milano, Il Saggiatore, 1967.
- R. Morriello, *IA ed editoria scientifica: come la Direttiva UE sta cambiando le regole del gioco*, in «Informatica Umanistica e Cultura Digitale: il blog dell'AIUCD», 31 Dicembre 2024, <<https://doi.org/10.58079/1301b>>.
- MUR, *PNR - Programma Nazionale per la Ricerca 2021-2027. Piano Nazionale per la Scienza Aperta*, <[https://www.mur.gov.it/sites/default/files/2022-06/Piano\\_Nazionale\\_per\\_la\\_Scienza\\_Aperta.pdf](https://www.mur.gov.it/sites/default/files/2022-06/Piano_Nazionale_per_la_Scienza_Aperta.pdf)>.
- A. Musso, *Eccezioni e limitazioni ai diritti d'autore nella Direttiva UE n. 790/2019*, in «Il diritto dell'informazione e dell'informatica», 36, 3, 2020, pp. 411-464.
- National Institute of Health (NIH), *What is the NIH Public Access Policy?*, <<https://www.nih.gov/health-information/nih-clinical-research-trials-you/what-is-nih-public-access-policy>>.

- Regolamento (UE) 2024/1689 del Parlamento Europeo e del Consiglio*, 13 giugno 2024: <[https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=OJ:L\\_202401689](https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=OJ:L_202401689)>.
- G. Roncaglia, *L'architetto e l'oracolo. Forme digitali del sapere da Wikipedia a ChatGPT*, Roma-Bari, Laterza, 2023.
- Ruediger Wischenbart Content and Consulting, *Global 50 Publishing Ranking 2023*, Wien, 2024, <<https://www.wischenbart.com/product/global-50-publishing-ranking-2024/>>.
- Springer Nature, *Figshare Integration. Guidance for Authors*, <<https://www.springernature.com/gp/authors/research-data/figshare-integration>>.
- Springer Nature, *Research Data Policy*, <<https://www.springernature.com/gp/authors/research-data-policy>>.
- Springer Nature, *Text and Data Mining/TDM*, in «Nature Portfolio», Nov. 6, 2024, <[https://support.nature.com/en/support/solutions/articles/6000251800-text-and-data-mining-tdm?utm\\_source=chatgpt.com](https://support.nature.com/en/support/solutions/articles/6000251800-text-and-data-mining-tdm?utm_source=chatgpt.com)>.
- Taylor & Francis, *Open Data Sharing Policy*, <<https://authorservices.taylorandfrancis.com/data-sharing-policies/open-data/>>.
- Understanding Knowledge as a Common. From Theory to Practice*, edited by Ch. Hess - E. Ostrom, Massachusetts - London, England, Cambridge, The MIT Press, 2007; trad. it. *La conoscenza come bene comune. Dalla Teoria alla pratica*, a cura di P. Ferri, Milano, Mondadori, 2009.
- UNESCO, *UNESCO Recommendation on Open Science*, <<https://www.unesco.org/en/open-science/about?hub=686>>.
- R. Van Noorden, *Elsevier opens its papers to text-mining*, in «Nature», 506, 17, 2014, <<https://doi.org/10.1038/506017a>>.
- Wellcome Trust, *Open Access Policy*, <<https://wellcome.org/grant-funding/guidance/open-access-guidance/open-access-policy>>.
- Wiley, *Open Data. Maximizing the value of research*, <<https://authorservices.wiley.com/author-resources/Journal-Authors/open-research-policies/open-data/index.html>>.
- Wiley, *Data Sharing with Dryad*, <<https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-service.html>>.
- G. Vrakas, *A literature review of "lawful" text and data mining*, in «Open Research Europe», 30 Oct 2024, <<https://doi.org/10.12688/openreseurope.18013.2>>.
- W. Xin Zhao - K. Zhou - J. Li - T. Tang - X. Wang - Y. Hou - Y. Min - B. Zhang - J. Zhang - Z. Dong - Y. Du - C. Yang - Y. Chen - Z. Chen - J. Jiang - R. Ren - Y. Li - X. Tang - Z. Liu - P. Liu - J.-Y. Nie - J.-R. Wen, *A Survey of Large Language Models*, in «ArXiv», last revised 13 Oct 2024 (v15), <<https://doi.org/10.48550/arXiv.2303.18223>>.

- C. Zong - R. Xia - J. Zhang, *Text Data Mining*, Singapore, Springer, 2021.
- S. Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, London, Profile Books Ltd, 2019; trad. it. *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, Roma, LUISS University Press, 2023.

Roberto Delle Donne  
Università degli Studi di Napoli Federico II  
roberto.delledonne@unina.it  
<https://orcid.org/0000-0001-8331-9436>



FABIANA GAMBARDELLA

*Words matter (?): su dilemmi e mistificazioni  
del Secolo post-nichilista*

*Abstract:* This essay attempts to clarify the new great narrative of the 21<sup>st</sup> century, that of digital technologies, characterized by key words such as communication, transparency, participation and sharing.

In our society, that provides an increasingly competitive and exclusionary economic model, a semantic restructuring is taking place in parallel which, on the contrary, makes cooperation, solidarity and greater well-being the manifesto of the present.

To do this, language seems to lose both its consubstantial connotative and evocative dimension and its hermeneutic vocation and turns into pure information devoid of ambiguity and nuances, keeping just a performative nature.

In the meantime, we produce increasingly intelligent artefacts, which achieve their objectives faster and better than the humans who produce them, however without understanding the contexts in which they intervene, but on the sole basis of the correlation of an immense amount of data. In this essay we ask whether we are the ones who modify the machines by creating increasingly intelligent artifacts, or it's them that, in turn, modify us, by limiting our existences to the present, and causing the end of consequentiality and of the hermeneutic capacity that has always characterized us.

*Keywords:* Communication, Artificial intelligence, Digital, Knowledge, Power, Surveillance.

1. *Chi parla male pensa male: di come nettare la cattiva coscienza attraverso il linguaggio*

«Le parole sono importanti», «Chi parla male pensa male!», sono motti oggi più che mai spendibili - sebbene in un senso antitetico

rispetto a chi li aveva prodotti - in una società come la nostra, che sta facendo del linguaggio politicamente corretto una sorta di ossessione, la tipica ossessione da cattiva coscienza. Con queste frasi Michele Apicella, interpretato da Nanni Moretti nel film *Palombella Rossa*, manifestava il suo sgomento davanti a un mondo che stava cambiando e si evolveva in senso spregiudicato, accantonando gradualmente i valori forti su cui il protagonista si era formato. Il film del 1989 dipinge lo spaesamento di una società davanti alla fine di un'epoca e di un mondo che, come il protagonista dopo un incidente, stava perdendo la memoria, il senso e il sentimento di un'appartenenza. Nondimeno Michele Apicella resta esterrefatto davanti al sunto che delle sue parole produce una giornalista, già all'epoca troppo giovane e smart, stravolgendo l'intenzione originaria dell'intervistato con una serie di neologismi e lemmi tratti dall'inglese. *Le parole sono importanti*, continuava a perorare con pathos Michele Apicella davanti all'insipienza della sua interlocutrice, ignara del passato, della storia ma assolutamente in linea con le esigenze di un presente, di cui al contrario il protagonista sembrava non comprendere più i contorni e assolutamente non gradire la deriva.

In un mondo completamente cambiato, in questo XXI secolo di utopie tecnologiche quasi realizzate e di compiute distopie economiche, sociali e politiche, il motto di Michele Apicella sembra essere ritornato in auge sebbene in un senso tutt'affatto diverso: le parole sono importanti, chi parla male pensa male, non perché nelle parole debba restare traccia profonda della memoria del passato, della storia; non perché le parole debbano conservare una densità e una profondità che si traduce in pratica e agire comune, ma al contrario perché dal linguaggio va rimossa ogni sedimentazione, ogni residuo narrativo, che, oltre a raccontarci il tempo e dunque sempre una storia, si presenta come consustanzialmente pluriforme, ambiguo, opaco; perciò va sottratto al linguaggio ogni residuo che implichi la necessità di ricorre-

re all'ermeneutica, gioco e trastullo tutto umano da sempre, gioco e trastullo che da sempre distingue l'uomo dalla pianta. Quel gioco, in una società dove i mezzi ci sono e i fini sono alla portata, dove la potenza del calcolo intelligente permette di raggiungere senza sforzo gli obiettivi, risulta una pratica obsoleta e fiaccante, sforzo patetico e destinato allo scacco da parte di esseri poco evoluti.

Se agli esordi del secolo scorso Martin Heidegger sosteneva che la vita è tortuosa e nebulosa e come tale esige l'ermeneutica<sup>1</sup>, cioè una serie di parole opache e ambigue che quella vita, in maniera maldestra, provano a descrivere, alle soglie della nuova era, quella del *petabyte*, ci si è resi conto che interpretare la tortuosità produce troppe parole e pochi risultati, nonché approssimativi. Come sostenuto da Chris Anderson, il corpus di dati a nostra disposizione si prospetta come vero e proprio inedito laboratorio della "condizione umana", atto a smantellare in maniera definitiva tutto l'armamentario di idoli decrepiti, sotto forma di teorie, prospettive, interpretazioni e valutazioni, sempre umane, troppo umane, ma evidentemente appannaggio di un modello d'uomo anacronistico e dunque non più all'altezza della nostra super-evoluzione:

The Petabyte Age is different because more is different. Kilobytes were stored on floppy disks. Megabytes were stored on hard disks. Terabytes were stored in disk arrays. Petabytes are stored in the cloud. As we moved along that progression, we went from the folder analogy to the file cabinet analogy to the library analogy to – well, at petabytes we ran out of organizational analogies<sup>2</sup>.

<sup>1</sup> M. Heidegger, *Phänomenologische Interpretationen zu Aristoteles. Einführung in die phänomenologische Forschung* (1921-22); trad. it. *Interpretazioni fenomenologiche di Aristotele*, p. 120.

<sup>2</sup> C. Anderson, *The end of theory: the Data deluge makes the scientific method obsolete*: <<https://www.wired.com/2008/06/pb-theory/>>.

Il passaggio è dirimente, pare suggerire che fino a un certo punto l'uomo abbia avuto bisogno di parole dense, di metafore e analogie per autorappresentarsi e rappresentare il mondo intorno: fino a un certo punto il dato è ancora rivestito di un *telos*, va a indicare un ordine, un percorso e una meta; sempre la stessa, il cammino glorioso dell'umano attraverso una conoscenza di volta in volta più ampia. E allora i dati, così come le parole che gli uomini dagli esordi della civiltà hanno vergato a costruire il mondo, devono essere conservati, a memoria di questo inesauribile cammino: dalla cartella, allo schedario, alla libreria, lo spazio che ospita un sapere in crescita deve aumentare, fino a che la potenziale infinità dell'informazione determina l'impossibilità della metafora, la sua superfluità, perché il dato è il dato e parla da sé e soprattutto non suggerisce, e non deve farlo, alcuna ulteriorità di senso, oltre alla chiara trasparenza del suo immediato esplicitarsi: «With enough data, the numbers speak for themselves»<sup>3</sup>.

Se i numeri parlano da soli non abbiamo più bisogno di interpretarli, possiamo abbandonare quella perniciosa tendenza di specie ad arrovellarsi nella inconcludente ricerca delle cause, possiamo smettere di prolungare il puerile esercizio dell'infanzia che ad ogni passo chiede perché. La ragione, a dispetto di tutti i sogni illuministi passati e presenti, spesso non c'è e anche se vi fosse, a spiegare la dinamica degli eventi, risulterebbe superflua rispetto alla chiara trasparenza della correlazione e alla sua infallibile sebbene ottusa capacità predittiva.

Se dunque possiamo abbandonare senza rimpianti ogni pratica ermeneutica, gli effetti positivi di questa liberazione dai fardelli del passato saranno visibili anche sul linguaggio.

<sup>3</sup> *Ibid.*

Adesso che tutto pare emergere spontaneamente in trasparenza, ora che abbiamo appreso a fare le domande corrette per ottenere risposte prive di ambiguità, anche il linguaggio diventa gioiosamente leggero, inclusivo e privo di asperità.

“Words matter”, le parole sono importanti, è la frase utilizzata anche dagli editori Puffin Books come breve nota nel colophon di alcuni libri di Roald Dahl, ineguagliabile scrittore per l’infanzia, autore di opere come *Willy Wonka and the chocolate factory*, *The BFG* o *Matilda*. Gli editori, di concerto con gli eredi dello scrittore, hanno deciso, in nome di questo accattivante manifesto del politicamente corretto e sulla scia del nuovo modello inclusivo di società che le parole debbono contribuire a forgiare, che una serie di lemmi utilizzati dall’autore, amatissimo tra l’altro anche per il suo spirito non propriamente *politically correct*, dovessero essere sostituite, perché discriminatorie, razziste ed escludenti, insomma non più conformi al rinnovato spirito dei tempi<sup>4</sup>. Il *Telegraph* ha individuato circa un centinaio di lemmi modificati e sostituiti<sup>5</sup>.

Nella società che andiamo costruendo *in primis* attraverso la grande epurazione del variegato edificio del linguaggio, parole come brutto, grasso e pazzo devono scomparire, così come i riferimenti troppo marcati al genere e al sesso, che in un mondo fluido e privo di steccati risultano quantomeno anacronistici: e allora,

<sup>4</sup> Sulla questione si veda F. D’Alessio, *La censura dei libri di Roald Dahl è un triste inganno verso tutti*: <<https://www.micromega.net/la-censura-dei-libri-di-roald-dahl-e-un-triste-inganno-verso-tutti/>>. L. Montagnoli, *I libri di Roald Dahl censurati. Modifiche necessarie o cancel culture?*: <<https://www.artribune.com/edizione/2023/02/libri-roald-dahl-censura-cancel-culture/>>.

<sup>5</sup> Cfr. E. Cumming – A. Buchanan – G. Holl-Allen, *The re-writing of Roald Dahl*: <<https://www.telegraph.co.uk/news/2023/02/17/roald-dahl-books-rewritten-offensive-matilda-witches-tweets/>>.

femmina diventa donna e mamma e papà si trasformano nei più neutri genitori.

E non è tutto: non si tratta solo della sostituzione di singole parole con altre ritenute più efficaci a esprimere il novello spirito dei tempi, che già sarebbe operazione integralmente discutibile; la rimozione è quella di intere opere, colpevoli di trasmettere terribili visioni del mondo: e allora Matilda, protagonista di una delle opere di Dahl, smette di leggere Rudyard Kipling, emblema dello spirito colonialista e oppressore dell'Occidente, per ripiegare sulla più innocua Jane Austen. *Words matter* evoca allora il tentativo, praticato attraverso la censura, la rimozione e l'oblio della storia, di costruire una società entro la quale, a partire dal linguaggio, nessuno possa sentirsi escluso. Inclusione infatti è un'altra parola estremamente cara alla società dell'informazione e della trasparenza. La comunità o meglio la community globale non lascia fuori nessuno: belli e brutti, magri e grassi, buoni e cattivi, devono avere, soprattutto nelle parole, la stessa possibilità di espressione e realizzazione. Inutile dire che non esiste alcuna inclusione senza memoria.

## *2. Comunità, Comprensione, trasparenza: l'incubo dell'utopia realizzata*

Eppure pare ci siano utopie che si realizzano, mondi accoglienti, più umani, più giusti, esteticamente accattivanti: il male sembra essere sgattaiolato via, vergognoso di sé, inestetico prima ancora che immorale, retaggio di un'umanità antica, che non ha ancora chiuso il cerchio della conoscenza, vincolata in maniera anacronistica ai limiti della carne, al pathos, al sentimento che straborda e reca danno, un'umanità che non si è ancora dischiusa completamente e integralmente alla luce della ragione; la ragione infatti produce i suoi frutti

più maturi solo nella trasparenza senza veli, nella condivisione e nella partecipazione; essa non snocciola il suo rosario nel segreto delle stanze opache dell'anima ma al contrario, si proietta costantemente fuori di sé e diventa progetto condiviso e azione proficua.

“Paradiso” è la prima parola pronunciata da Mae varcando le soglie del *Cerchio*, la big tech più all'avanguardia del mondo globale<sup>6</sup>, dove viene assunta grazie alla segnalazione di un'amica.

Una vasta area che ha visto un variegato passato di attività – prima cantiere navale, poi *drive in*, poi mercato delle pulci e infine bandiera del degrado ambientale, da cui, glorioso o inglorioso che fosse, il passato viene cancellato per dare vita a uno spazio integralmente nuovo, sapientemente bello, accogliente e confortevole più della propria casa, circondato da campi da tennis, aree picnic, collinette e amene fontane zampillanti. Una somiglianza non dissimulata con le notissime big tech della Silicon Valley. Una natura armoniosamente addomesticata a ospitare un nuovo, avanzato modello di umanità, che si mostra luminoso sin dalle prime pagine della narrazione: gli impiegati dell'azienda sono giovanissimi, i più anziani hanno da poco superato la trentina e malgrado questo possiedono un'aria «distesa, competente e saggia»<sup>7</sup>.

*Sogna, partecipa, socializza, innova, immagina*: queste le parole motivanti che lampeggiano a turno sui ciottoli lungo il vialetto circondato di alberi di agrumi, che porta all'ingresso principale. È giugno e c'è il sole.

In un mondo che appare perfetto sembra essere venuto meno anche l'eterno conflitto servo-padrone: i dirigenti dell'azienda lavorano per promuovere l'interesse generale, il loro *in primis* certo, ma anche

<sup>6</sup> Si cfr. il romanzo di D. Eggers, *The Circle*.

<sup>7</sup> D. Eggers, *The Circle*; trad. it. *Il cerchio*, p. 44.

quello dei dipendenti, che in un ambiente confortevole sono più creativi e propositivi, quello delle loro famiglie, con asili nido e isole per bambini, fino al benessere dell'intera società, la cui prosperità aumenterà in maniera esponenziale, parallelamente all'esponenziale aumento di una conoscenza senza più veli, che una creatività tecnica giovane e smart produce senza sosta e sempre più velocemente.

Nel suo *Realismo capitalista* Mark Fisher definiva "comunisti liberal", i nuovi padroni alla Bill Gates e spiegava, anche attraverso le parole di Žižek, i pilastri ideologici, o meglio la nuova narrazione che questi utopisti del capitalismo buono andavano costruendo, con un uso sapiente delle parole:

I comunisti liberal pensano che il lavoro vada (post)modernizzato, o meglio ancora reso «smart»: spiega Žižek che «essere smart significa essere dinamici e nomadi contro qualsiasi burocrazia centralizzata. Significa credere nel dialogo e nella cooperazione contro qualsivoglia autorità centrale. Credere nella flessibilità contro la routine. Nella cultura e nella conoscenza contro la produzione industriale. Nell'interazione spontanea e nell'autopoiesi contro le rigide gerarchie»<sup>8</sup>.

La parola chiave di questa società, il dogma su cui è costruita l'intera impalcatura del progresso, è trasparenza: trasparenza dei dati e delle informazioni, che in un reale processo di democratizzazione devono essere accessibili e fruibili per tutti; la trasparenza genera conoscenza e va a braccetto con condivisione e partecipazione: nessuno deve restare indietro, non ci sono ultimi o esclusi da questa novella Atlantide della saggezza. La trasparenza, va da sé, ha in prima istanza a che fare col nostro linguaggio e le nostre pratiche.

<sup>8</sup> M. Fisher, *Capitalist Realism: Is There No Alternative?*; trad. it. *Realismo Capitalista*, p. 69.

La trasparenza senza veli delle novelle pratiche linguistiche, viene incontro alla neoassunta Mae attraverso la figura di Renata, una sorta di Virgilio addetta a introdurre i neofiti nella realtà lavorativa, fornendo loro tutti gli strumenti e le cognizioni di cui necessitano per un inserimento proficuo. Mae, nervosa al suo primo giorno di lavoro, prova a stemperare il disagio con qualche battuta di spirito ma l'interlocutrice, che ha introiettato perfettamente i dettami sulla correttezza della comunicazione, sembra non seguirla, resta perplessa, aggrota la fronte davanti a esternazioni che risultano ambigue, non in linea col clima e la temperatura del posto. Renata ha imparato a rivolgersi ai suoi destinatari alla stregua di un assistente virtuale: pacata, neutra, priva di emozioni, come l'assistente virtuale resta interdetta davanti a proposizioni che non presentano un senso univoco, alle quali non si può dare una risposta certa. Del resto questa perplessità è rafforzata da altri protagonisti della grande utopia tecnologica: «Con la tecnologia disponibile, la comunicazione non dovrebbe mai essere in dubbio. Non dovrebbe mai essere impossibile capirsi»<sup>9</sup>, afferma il team manager dell'azienda, il cui motto è *Comunicazione, Comprensione, Chiarezza*. Per i promotori di questo nuovo modello di società, il linguaggio deve essere senza residui "denotativo", un'informazione chiara e trasparente, che priva di distorsioni di sorta fluisca da un destinatario a un destinatario. L'informazione è lineare, presenta un contenuto che deve giungere immodificato all'ascoltatore, il quale, senza esitazioni vi può rispondere in maniera adeguata. Il linguaggio allora, in questa novella comunità, la nostra di fatto, si riduce a passaggio di informazioni, poiché possiede in fondo mera natura prestazionale.

<sup>9</sup> D. Eggers, *op. cit.*, p. 50.

Humberto Maturana e Francisco Varela avrebbero provato sgo-mento davanti a questa società e alle sue pratiche linguistiche. Verso la fine degli anni '70 mentre tentavano di definire il fenomeno vita attraverso il concetto di autopoiesi, soffermavano l'attenzione su quell'inedita forma dell'evoluzione della nostra specie che è per l'appunto il linguaggio, e lo definivano al contrario "connotativo" e "orientante".

E cioè, lungi dall'essere la neutra espressione di un'informazione che può essere trasmessa senza opacità da un destinatore a un ricevente, esso manifesta tutta la sua forza plastica e poetica, proprio a partire dalla sua consustanziale ambiguità, dai suoi intrinseci tranel-  
li, dalle innumerevoli possibilità di fraintendimento che ogni espres-  
sione reca inevitabilmente in sé, perché necessariamente legata a un  
particolare punto di vista, o meglio a un preciso dominio cognitivo.  
Il linguaggio serve infatti a «orientare l'orientato entro il suo domi-  
nio cognitivo indipendentemente dal dominio cognitivo dell'orient-  
tatore»<sup>10</sup>; e cioè colui che ascolta, lungi dal recepire il dettato nella  
sua neutralità (neutralità inesistente poiché prodotta anch'essa a par-  
tire da un peculiare dominio cognitivo), lo assimila, lo metabolizza  
e lo addomestica in relazione alle sue esigenze vitali. Ed è proprio a  
partire dalle innumerevoli possibilità di manipolazione ermeneuti-  
ca della parola<sup>11</sup>, soggetta costantemente al fraintendimento, che la  
storia si produce e dà vita ogni volta all'inedito. La parola, in effetti,  
diverrebbe presto lettera morta se si limitasse alla mera funzione di  
passaggio di informazioni.

<sup>10</sup> H. Maturana - F. Varela, *Autopoiesis and Cognition. The Realization of the Living*; trad. it. *Autopoiesi e cognizione. La realizzazione del vivente*, p. 80.

<sup>11</sup> «La parola è cenno e non segno nel senso di semplice denotazione», cfr. M. Heidegger, *Unterwegs zur Sprache*; trad. it. *In cammino verso il linguaggio*, p. 103.

Come sostenuto da Byung-Chul Han, la società dell'informazione, diremmo noi, del linguaggio ridotto a informazione, di un linguaggio spogliato in fondo della sua dimensione simbolica, sempre più vicino al linguaggio macchina, produrrebbe l'eterna reiterazione dell'uguale, l'inferno dell'identico<sup>12</sup>.

Eppure il tipo di società descritto nel *Cerchio* si presenta con tratti accattivanti e assolutamente desiderabili, una società della possibilità, che si esplicita in prima istanza come possibilità di realizzazione del desiderio individuale, ma anche come comunità aperta e inclusiva, priva dei tratti del negativo, costantemente tesa all'affermazione, che ha abolito la repressione a favore dell'espressione creativa in tutte le sue sfaccettature e che va organizzando le sue pratiche in modo da allargare gli orizzonti di condivisione e partecipazione. La trasparenza dunque estende i suoi confini in maniera leggera e subdola; prende ad esempio la forma di un nuovo artefatto, le videocamere *See change* con lo *sharing*, che possono essere installate con facilità ovunque, passando inosservate<sup>13</sup>: per informare il surfista sullo stato delle onde in tempo reale, per controllare il traffico sulla Statale prima di andare a lavoro, per visionare il proprio appartamento quando si è fuori. Ma se è vero che una vita pienamente umana si dà quando la comunità è al primo posto, va da sé che se ciascuno condividesse "il proprio punto di vista sul mondo", il suo occhio metallico con

<sup>12</sup> Si cfr. B.-C. Han, *Vom Verschwinden der Rituale. Eine Topologie der Gegenwart*.

<sup>13</sup> Quella che sembra una finzione letteraria in effetti non fa altro che descrivere ciò che accade da tempo con le tecnologie informatiche in generale: le innovazioni vengono "prima" implementate, senza alcun tipo di permesso né regolamentazione, per diventare in breve tempo dei bisogni primari delle persone che dimenticano o decidono di bypassare la violazione dei diritti che alcune di queste tecnologie determinano. A tal proposito si cfr. l'importante testo di S. Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*.

quello degli altri, allora ci troveremmo di fronte a un salto di qualità, a un vero e proprio salto ontologico: non più soltanto vantaggi individuali, ma un reale passaggio dalla dimensione dell'efficace a quella del bene; infatti dalla pianificazione dell'intrattenimento e della gestione delle pratiche quotidiane, alla tutela dei diritti umani il passo è breve: «Chi commetterebbe un reato sapendo di essere sorvegliato ogni momento dappertutto?».<sup>14</sup> La trasparenza obbliga dunque alla determinazione morale.

Il nuovo illuminismo si profila allora come illuminismo della sorveglianza, che ci rende sicuri, ci fa addormentare su due guanciali, ci fa evolvere forti e consapevoli.

Datificazione e responsabilità, potremmo dire in questo caso *datificazione è responsabilità*, produce esseri umani buoni e mansueti; “tutto quello che succede deve essere conosciuto”, il motto che compare sul grande schermo dell'azienda, ricorda *Il grande fratello ti guarda* di orwelliana memoria ma con una differenza: il grande fratello era l'emblema di una società disciplinare, in cui la libertà degli individui era limitata e assoggettata ai dettami di un potere tirannico.

In questo caso istituiamo di nostra sponte e anche con un certo piacere un sistema di sorveglianza laterale<sup>15</sup> e reciproca, in cui diventiamo tutti al contempo controllori e controllati, protagonisti attivi nel costruire le gabbie in cui decidiamo di dimorare senza nessuna coercizione apparente.

<sup>14</sup> D. Eggers, *op. cit.*, p. 68.

<sup>15</sup> Sul tema della sorveglianza laterale si cfr. M. Andrejevic, *The Work of Watching One Another: Lateral Surveillance, Risk, and Governance*: <<http://www.surveillance-and-society.org>>.

3. *Le parole e le cose: sul definitivo divorzio fra teoria e prassi*

Il testo di Eggers produce nel lettore una sorta di spaesamento, è perturbante, nel senso che ha a che fare in un modo peculiare con le dimensioni del domestico e dell'ignoto. Questo perché l'effetto che immediatamente produce, malgrado o in virtù proprio dell'artefatta e plastificata bellezza dei luoghi, della loro armonia estetica e morale, è di una lontananza che sembra assolutamente non riguardarlo; il lettore comincia ad adagiarsi confortevolmente nel racconto pensando di avere a che fare con una distopia futuristica. L'elemento perturbante comincia a emergere laddove questa lontananza dalla nostra domesticità, in tal caso confortante, diventa col procedere della lettura una prossimità sempre più sconcertante. Quel mondo, capisce con sgomento il lettore, è il mio mondo, quello delle mie pratiche quotidiane, domestico e, guardato da fuori, assolutamente distopico fino al grottesco.

Tale realtà, la nostra realtà, è caratterizzata da un paradosso: le sue prassi sono il tripudio dell'individualismo, della ricerca del profitto a tutti i costi, si tratta di un mondo spogliato di qualunque residuo universalistico, di ogni possibilità di trascendimento e finanche di qualsiasi spazio politico-comunitario che riguardi i valori, l'*ethos*, l'agire comune consapevole; e tuttavia la narrazione che di quella stessa realtà i suoi protagonisti vanno costantemente tessendo è di tutt'altra fattura: «Prima di tutto la comunità», è tra gli slogan preferiti dall'azienda, oltre a «Qui non lavorano automi», ma il fior fiore della generazione presente, le cui opinioni vanno ascoltate, promosse e valorizzate<sup>16</sup>.

<sup>16</sup> D. Eggers, *op. cit.*, p. 50.

In altre parole le parole non coincidono più con le cose: l'altro, sfruttato, ostracizzato, schiacciato dalla precarietà è esaltato dentro la rete magica del linguaggio come portatore unico di valore, come creatività che deve esprimersi.

Il romanzo descrive in fondo la realizzazione di quella grande narrazione sviluppatasi attorno agli anni '90 del Secolo scorso, con l'estendersi della rete a livello di massa, un'utopia polimorfa che si delinea in ambiti variegati e con connotazioni differenti, che tuttavia presenta un punto comune: la grande aspettativa nei confronti di un futuro, che attraverso le inedite possibilità comunicative offerte da internet, avrebbe prodotto la nascita di spazi, sinergie e movimenti inediti e tutti da scoprire<sup>17</sup>. Si rifletteva sulla possibilità, come descritto da Pierre Lévy, di creare vere e proprie intelligenze collettive, costruttrici di comunità reali, non più fondate sullo sfruttamento e la gerarchia, sulle distinzioni molarie di classe o categoria, ma su una reale valorizzazione delle differenze, che avrebbe incrementato la ricchezza e il benessere sociale di tutti; aumento del sapere, creazione di una cosmopedia globale e una nuova fase dell'evoluzione umana, o meglio una vera e propria accelerazione del processo di ominazione,

<sup>17</sup> Nella prefazione a *Realismo capitalista* di Mark Fisher, Valerio Mattioli fa il punto su un certo clima culturale che attraversa gli anni '90 del Secolo scorso, in particolare in Inghilterra, all'Università di Warwick, dove nel 1995 Sadie Plant, cui sarebbe subentrato il più controverso Nick Land, fonda il gruppo di ricerca non ufficiale Cybernetic Culture Research Unit, che, nato per riflettere sulle implicazioni politiche dello sviluppo della rete, finisce per essere un fertile terreno di incontro di esperienze variegata, dalla filosofia, alla musica techno, passando per l'esoterismo e la biologia molecolare. Oltre a condividere una certa polemica nei confronti di un marxismo anacronistico e ancora legato a un passato «spazzato via dalla rivoluzione digitale», questa cultura condivideva «quel senso di eccitamento per il futuro, per l'inaspettato, per il potenziale dirompente di quanto ancora non è», cfr. V. Mattioli, *La funzione Fisher*, in M. Fisher, *op. cit.*, p. 10.

attraverso la formazione di un inedito spazio antropologico, quello del sapere, che avrebbe potuto determinare, se razionalmente progettato e gestito, un salto in avanti e un reale progresso dell'umanità<sup>18</sup>.

In fondo si trattava di un sogno già sognato: quello di una ragione che, se utilizzata al meglio, illumina le zone d'ombra dell'essere, rischiarando le esistenze singole e nel suo espandersi produce progresso per la comunità, allargando i domini di libertà e partecipazione.

Questo secondo illuminismo è espresso bene nel romanzo dalla figura di Bailey, uno dei fondatori dell'azienda, altro personaggio che strizza l'occhio ai nerd-leader della Silicon Valley: «noi qui non facciamo altro che cercare di conoscere ciò che prima era ignoto»<sup>19</sup>: il grande libro del mondo deve diventare completamente trasparente, accessibile e «alla portata di tutti».

A dispetto di quanto affermava Lyotard nel 1979, in merito all'incredulità delle società postfordiste nei confronti delle metanarrazioni<sup>20</sup> e all'impossibilità di produrne di nuove nell'era del proliferare dei giochi linguistici e del sapere teso alla mera performatività,

<sup>18</sup> Cfr. P. Lévy, *L'Intelligence collective. Pour une anthropologie du cyberspace*. Si cfr. anche H. Bay, *Taz. The temporary Autonomous Zone*, un testo cult della contro-cultura underground americana che traccia con toni psichedelico-prophetici, le linee di un futuro inedito in cui, anche grazie alla rete sarebbe stato possibile intaccare il sistema di controllo e repressione cui siamo soggetti, attraverso la formazione di zone autonome temporanee, molto diverse dalla virtualità disincarnata del game, caratterizzate da reali e carnali spazi di incontro e condivisione, libere da forme di controllo e disciplinamento. Non si sarebbe trattato di rivoluzione, che in quanto tesa a costruire qualcosa di nuovo sulle ceneri del vecchio, è altresì destinata a ricristallizzarsi in una forma definita e perciò antilibertaria. Si trattava, a detta di Bay, di spazi di insurrezione, dove si sorge, ci si eleva insorgendo, in un movimento costante che fa e disfa continuamente le sue forme per rendersi invisibile e dunque inattaccabile dal potere.

<sup>19</sup> D. Eggers, *op. cit.*, p. 64.

<sup>20</sup> J.-F. Lyotard, *La Condition postmoderne: rapport sur le savoir*.

l'utopia digitale si è forgiata come nuova grande narrazione, grande epopea dell'umano e dei suoi fini, rivestita anche di un profondo *ethos* che all'esaltazione della ragione aggiunge l'istanza alla giustizia e la solidarietà. Il problema è che col dispiegarsi fattuale di questa possibilità accompagnata dal profitto che se ne poteva ricavare, l'eroica narrazione è trapassata in una grande mistificazione e in una poderosa operazione di marketing.

4. *Words don't matter: a proposito di differenza ontologica, performatività ottuse, e dialettica servo-padrone*

“Ogni volta che licenzio un linguista la nostra performance aumenta”. Si esprimeva così Frederick Jelinek, esperto di teoria dell'informazione, che già a partire dagli anni '70, lavorando sul riconoscimento vocale, capisce che la presenza di regolarità statistiche nella lingua consente alla macchina di tradurre o riconoscere anche senza comprendere né sapere le regole e le eccezioni grammaticali di un dato idioma. In altre parole non era necessario spendere inutili energie per cercare di forgiare macchine intelligenti in grado di emulare le procedure logiche dell'uomo, poiché la macchina produce risultati efficaci e, seppur insipiente, possiede un'ottusità performativa, garantita dalla quantità poderosa di informazioni a cui può accedere e su cui può lavorare in maniera statistica<sup>21</sup>.

Dalla traduzione di un testo all'algoritmo di raccomandazione - che puntualmente carpisce i nostri desideri, i gusti e le tendenze,

<sup>21</sup> Per una interessante riflessione sul sogno e le aspettative nei confronti delle macchine intelligenti si cfr. N. Cristianini, *La scorciatoia. Come le macchine sono diventate intelligenti senza pensare in modo umano*, p. 33.

consigliandoci senza errore proprio ciò che fa al caso nostro e a cui talvolta noi stessi non avevamo ancora pensato - l'intelligenza artificiale, sebbene nella sua versione "debole", è entrata nelle nostre pratiche quotidiane, fornendoci una serie di comodi servizi fino a non molto tempo fa inimmaginabili.

E lo fa in maniera discreta, senza prosopopea: un servitore ottusamente utile che non aspira a emulare il multiforme ingegno del suo fattore. È un dato in fondo rassicurante, ché il servo si sa, deve essere operoso, zelante ma mai acuto. Sempre pronti, dinnanzi a ogni scoperta che facciamo, a mantenere intatto il nostro primato ontologico, anche col tripudio dell'intelligenza artificiale ci piace spendere fiumi di inchiostro per ribadire la distanza ontologica tra la nostra conoscenza e la "loro", quella delle macchine.

Assenza di comprensione, attività eminentemente statistica, assenza di profondità, schiacciamento sul presente, questa la "povertà di mondo" della macchina cui si contrappone l'intelletto umano come sistema complesso, diacronico che si sviluppa a partire da una specifica temporalità, impossibile da riprodurre in una macchina.

La macchina - continuiamo sgomenti a sostenerlo anche davanti ai prodigiosi risultati dell'IA generativa - non comprende ma correla, si limita cioè a mettere in relazione i dati a partire da un immenso database. La sua forza è nella quantità piuttosto che nella qualità. Non si intende di ermeneutica la macchina, non conosce dubbi, sfumature, simulazioni e dissimulazioni; è ingenua e *naïf* nelle sue operazioni di estrapolazione.

Nel realizzare il sogno prometeico di progresso ed evoluzione di una conoscenza sempre più piena, l'umano è passato attraverso una serie di tappe, anche narrative: dal sogno di forgiare una creatura, l'intelligenza artificiale, che emulasse le complesse facoltà dell'uomo, che ragionasse al suo stesso modo; dal sogno dunque di un'intelligenza universale, che possiede parole e simboli, sa manipolarli e ne

estrapola i segreti, allo scacco davanti alla limitatezza delle proprie possibilità, fino alla nuova consapevolezza: si possono portare avanti grandi imprese anche senza quella Ragione; è possibile cioè essere produttivi, altamente performativi anche senza comprendere. È il caso di dire, stavolta, che le parole non sono importanti o meglio, valgono nella loro quantità e non per le sottigliezze che evocano, inutile rumore di fondo, che un sistema intelligente deve saper ignorare per portare a casa il risultato.

C'è tuttavia nell'ambito di questa narrazione consolante - che da sempre addomestica il nuovo provando a reintegrarlo in un racconto già noto - qualcosa che non torna.

Se usciamo dalla posizione metafisica che interpreta in maniera dicotomica e lineare la relazione uomo-tecnica descrivendo un soggetto sovrano, che, carente biologicamente, produce il miracolo degli artefatti, oggetti neutri su cui facilmente esercita la propria sovranità, e assumiamo la giusta ottica circolare, per la quale l'umano produce l'artefatto e l'artefatto modifica costantemente le fattezze dell'umano<sup>22</sup>, allora la situazione diventa più problematica.

C'è da chiedersi a questo punto dell'evoluzione "nostra" e "loro", chi modifica chi? E soprattutto chi esercita la sovranità?

In altre parole sono le macchine a emulare la nostra intelligenza oppure l'umano a modificare gradualmente la propria nella strettissima ibridazione con la macchina? E va modificandosi allora nel senso di quella "povertà di mondo" tipica della macchina, dell'ap-

<sup>22</sup> M. McLuhan, *Understanding Media: The Extensions of Man*; trad. it. *Gli strumenti del comunicare*. Per McLuhan bisogna abbandonare la posizione inutilmente naif dell'idiota tecnologico che considera i media in generale degli strumenti neutri, la cui reale essenza deriverebbe eminentemente dall'intenzionalità di chi li usa, essi al contrario «alterano costantemente e senza incontrare resistenza, le reazioni sensoriali o le forme di percezione», Si cfr. pp. 38-39.

piattimento sul presente, della fine della consequenzialità, dell'ermeneutica, di quei mezzi toni e di quelle sfumature che da sempre hanno decretato la nostra superiorità ontologica<sup>23</sup>?

Se la scrittura ha prodotto l'individuo lineare, sequenziale e borghese, cosa sta forgiando l'infosfera nella quale siamo immersi?

L'infosfera costituisce sì un ambiente confortevole e lussureggiante, le estensioni tecnologiche che accompagnano la nostra esistenza ci coccolano, con la stessa sollecitudine di un genitore che sa tutto di noi e può anticipare i nostri desideri. L'algoritmo di raccomandazione ci consiglia quotidianamente l'abito, il libro e le esperienze a cui ancora non avevamo pensato e nel lasciarci mollemente cullare dal tutore digitale, rimuoviamo il piccolo dettaglio che probabilmente siamo noi a servire lui e non viceversa, ci manipola inducendo e orientando le nostre scelte, *in primis* in materia economica, spingendoci al consumo compulsivo, ma anche, la storia lo mostra, manipolando le nostre idee politiche, la nostra visione del mondo e dell'altro.

È evidente che la società nella quale siamo immersi, i suoi modi di produzione e l'ideologia economica neoliberalista stanno promuovendo un inedito tipo di umanità, leggera, smart e trasparente che si trastulla tra consumo e intrattenimento compulsivi, che preferisce che le sue scelte siano dolcemente orientate dagli assistenti di cui si serve. Un'umanità che senza troppi crucci pare dire addio a un certo tipo di complessità, ai dubbi, all'indugiare, al linguaggio che si dispiega per onorare e per dissacrare e quasi mai per meramente informare. Un nuovo o forse neanche tanto, tipo di umanità, che

<sup>23</sup> Su questi temi si cfr. i testi di B.-C. Han, *Im Schwarm. Ansichten des Digitalen e Infocratie. Digitalisierung und die Krise der Demokratie*. Si cfr. anche N. Carr, *Internet ci rende stupidi? Come la rete sta cambiando il nostro cervello*.

come già accaduto, allegramente si accinge a «pattinare sulla superficie delle cose»<sup>24</sup>.

## Bibliografia

- C. Anderson, *The end of theory: the Data deluge makes the scientific method obsolete*, in «Wired», June 2008, <<https://www.wired.com/2008/06/pb-theory/>>.
- M. Andrejevic, *The Work of Watching One Another: Lateral Surveillance, Risk, and Governance*, in «Surveillance & Society», edited by D. M. Wood, 2, 4, 2005, pp. 479-497, <<http://www.surveillance-and-society.org>>.
- H. Arendt, *Responsibility and Judgment*, United States, Knopf, 2003; trad. it. *Alcune questioni di filosofia morale*, Torino, Einaudi, 2015.
- H. Bay, *Taz. The Temporary Autonomous Zone. Ontological Anarchy, Poetic Terrorism*, New York, Pacific Publishing Studio, 2011 (1<sup>st</sup> edition 1991); trad. it. *T.A.Z. La Zona Autonoma Temporanea*, Milano, Shake Edizioni, 2020 (1<sup>a</sup> edizione 1993).
- N. Carr, *Internet ci rende stupidi? Come la rete sta cambiando il nostro cervello*, Milano, Cortina, 2011.
- N. Cristianini, *La scorciatoia. Come le macchine sono diventate intelligenti senza pensare in modo umano*, Bologna, Il Mulino, 2023.
- E. Cumming - A. Buchanan - G. Holl-Allen, *The re-writing of Roald Dahl*, in «The Telegraph», 01-02-2023, <<https://www.telegraph.co.uk/news/2023/02/17/roald-dahl-books-rewritten-offensive-matilda-witches-twits/>>.
- F. D'Alessio, *La censura dei libri di Roald Dahl è un triste inganno verso tutti*, in «Micromega», 20-02-2023, <<https://www.micromega.net/la-censura-dei-libri-di-roald-dahl-e-un-triste-inganno-verso-tutti/>>.
- D. Eggers, *The Circle*, United States, Knopf, 2013; trad. it. *Il cerchio*, Milano, Feltrinelli, 2022.
- M. Fisher, *Capitalist Realism: Is There No Alternative?*, Winchester, Zer0 Books, 2022 (1<sup>st</sup> edition 2009); trad. it. *Realismo capitalista*, Roma, Nero Editore, 2018.
- B.-C. Han, *Im Schwarm. Ansichten des Digitalen*, Berlin, MSB Matthes&Seitz, 2013; trad. it. *Nello sciame. Visioni del digitale*, Milano, Nottetempo, 2015.

<sup>24</sup> H. Arendt, *Responsibility and Judgment*; trad. it. *Alcune questioni di filosofia morale*, p. 62.

- B.-C. Han, *Vom Verschwinden der Rituale. Eine Topologie der Gegenwart*, Berlin, Ullstein, 2019; trad. it. *La scomparsa dei riti. Una topologia del presente*, Milano, Nottetempo, 2021.
- B.-C. Han, *Infocratie. Digitalisierung und die Krise der Demokratie*, Berlin, Matthes&Seitz, 2021; trad. it. *Infocrazia. Le nostre vite manipolate dalla rete*, Torino, Einaudi, 2023.
- M. Heidegger, *Phänomenologische Interpretationen zu Aristoteles. Einführung in die phänomenologische Forschung (1921-22)*, Gesamtausgabe bd. 61, hrsg. von Walter Bröcker - Käte Bröcker-Oltmanns, Frankfurt am Main., Klostermann, 1994 (1<sup>st</sup>e Ausgabe 1985); trad. it. *Interpretazioni fenomenologiche di Aristotele (1921-22)*, Napoli, Guida Editori, 1990.
- M. Heidegger, *Unterwegs zur Sprache (1950–1959)*, Gesamtausgabe bd. 12, hrsg. von F.-V. von Herrmann, Frankfurt am Main., Klostermann, 1985; trad. it. *In cammino verso il linguaggio*, Milano, Mursia, 2007.
- P. Lévy, *L'Intelligence collective. Pour une anthropologie du cyberspace*, Paris, La Découverte, 1994; trad. it. *L'intelligenza collettiva. Per un'antropologia del cyberspazio*, Milano, Feltrinelli, 1994.
- J.-F. Lyotard, *La Condition postmoderne: rapport sur le savoir*, Paris, Les éditions de minuit, 1979; trad. it. *La condizione postmoderna*, Milano, Feltrinelli, 1985.
- H. Maturana - F. Varela, *Autopoiesis and Cognition. The Realization of the Living*, Dordrecht, D. Reidel Publishing Company, 1980; trad. it. *Autopoiesi e cognizione. La realizzazione del vivente*, Venezia, Marsilio, 1988.
- M. McLuhan, *Understanding Media: The Extensions of Man*, Canada, McGraw-Hill, 1964; trad. it. *Gli strumenti del comunicare*, Milano, Il Saggiatore, 2015.
- L. Montagnoli, *I libri di Roald Dahl censurati. Modifiche necessarie o cancel culture?*, in «Artribune», 22-02-2023, <<https://www.artribune.com/editoria/2023/02/libri-roald-dahl-censura-cancel-culture/>>.
- S. Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, London, Profile Books Ltd, 2019; trad. it. *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, Roma, LUISS University Press, 2023.

Fabiana Gambardella  
Università degli Studi di Napoli Federico II  
fabiana.gambardella@unina.it  
<https://orcid.org/0009-0005-5005-2946>



GIANLUCA GIANNINI

*L'insostenibile leggerezza dell'InfoSapiens  
(del Digitocene e dintorni)*

*Abstract:* This essay introduces the concepts of Digitocene, SmartEcoSystem and InfoSapiens to analyze the ongoing anthropological transformation due to the pervasiveness of the digital. Digitalization, in the era of AI, is not merely a tool, but an onto-technological force that redefines reality and the human condition. Thus emerges the InfoSapiens, a subject whose existence, cognition and relationships are inherently shaped by streams of data and algorithms in digital ecosystems. This symbiosis with cognitive machines is not a transitory phase but an accomplished metamorphosis that requires rethinking the human. The essay explores the implications of this computational heteronomy and the need to critically inhabit this new condition, emphasizing the inseparability of the ecological and the informational crises. The goal is to understand InfoSapiens not as loss, but as a threshold for a new humanism.

*Keywords:* Digitocene, SmartEcoSystem, InfoSapiens, Human Condition, Artificial Intelligence.

Il fatto che attualmente le macchine posseggano ben poca coscienza non ci autorizza affatto a ritenere che la coscienza meccanica non raggiungerà col tempo il massimo sviluppo. Un mollusco non possiede gran che di coscienza. Pensate alla straordinaria evoluzione delle macchine in questi ultimi secoli, e osservate con quale lentezza progrediscono il regno vegetale e quello animale. Le macchine più altamente organizzate sono creature non di ieri, ma addirittura degli ultimi cinque minuti, oserei dire, di fronte alla storia dell'universo. Supponiamo che gli esseri coscienti esistano da venti, venticinque milioni di anni: guardate quali passi da gigante hanno fatto le macchine nell'ultimo millennio! Il mondo non può forse durare altri venti milioni di anni? Ma se dura altri venti milioni di anni, che cosa finiranno per diventare le macchine? Non è più prudente distruggere il male all'inizio e impedire loro di progredire ulteriormente? Chi può dire che la macchina a vapore non possieda una

qualche sorta di coscienza? Dove comincia e dove finisce la coscienza? Chi può fissare il limite? Chi può fissare un qualsiasi limite? Non sono forse le cose intessute tutte l'una nell'altra? E le macchine non sono legate in mille modi alla vita animale?

Samuel Butler, *Erewhon*

Ogni volta che ci si imbatte in saggi, articoli scientifici o libri dedicati alla cosiddetta “Rivoluzione/Transizione Digitale” ci si trova inevitabilmente a dover affrontare una serie di valutazioni preliminari. Ci troviamo di fronte a un'opera dichiaratamente o velatamente tecnofobica? Oppure è l'ennesima manifestazione di un entusiasmo cieco per la tecnologia? L'autore di turno documenta una reale competenza ovvero, non essendolo affatto, si limiterà a riproporre, quasi per inerzia, concetti di terza o quarta mano, trascinando il lettore fondamentalmente da una stanza immaginaria, quando non buia, all'altra? E ancora, soprattutto nel caso di cimenti filosofici – che *curiosamente* si stanno moltiplicando da un po', dopo un sostanziale silenzio durato quasi vent'anni –, siamo davanti alla classica pretesa della filosofia di elevarsi a giudice ultimo e supremo, mascherando in realtà un'operazione metafisica rinnovata e perciò maldestra, o invece (e finalmente), in maniera competente, siamo al cospetto di ragionamenti che rispondono alla vocazione e disposizione della filosofia stessa, ovvero apprendere, analizzare, comprendere e infine interpretare il proprio tempo con il pensiero (critico)?

Non so se le riflessioni che seguiranno cadranno in alcune (o tutte) delle suddette *macchinazioni* strutturali, ma una cosa rispetto ad esse appare chiara fin dall'inizio: dette operazioni preliminari, al semplice lettore curioso e/o allo studioso ingaggiato, sono imposte da una più che comprensibile diffidenza in considerazione delle insidie narrativo-ermeneutiche e, in taluni casi, preconcetti radicati che albergano al cuore di molta letteratura *sul* Digitale, essendo sovente il punto di caduta di una grande illusione ottica, se non di un errore

perceptivo colossale a meno che, appunto, non ci sia al fondo autentico pensiero critico. Quando si parla di “Rivoluzione/Transizione Digitale”, ci si ostina a guardare il dito che indica, senza mai volgere lo sguardo verso ciò che viene indicato. E se anche ci si accorgesse della Luna, difficilmente si riuscirebbe a comprendere che quella Luna, in fondo, siamo proprio noi.

Mi sembra finanche banale rilevare che quando ci si prefigge l'obiettivo di discutere di “Rivoluzione/Transizione Digitale”, il dato inoppugnabile sia che l'avvento delle tecnologie digitali abbia trasformato radicalmente il modo in cui viviamo, pensiamo e interagiamo con il mondo esterno, con le alterità che lo inabitano e, dunque, con e tra noi stessi. Mi sembra oltremodo ovvio constatare che questa trasformazione non si limiti a cambiamenti superficiali nelle nostre abitudini quotidiane, dei nostri *stili* di vita, ma tocchi, fino a modificarle profondamente, le fondamenta su cui prende forma il multiverso delle nostre strutture identitarie, e quindi la nostra percezione della realtà e i nostri modi di relazionarci con gli altri. Insomma, che quando abbiamo la pretesa di discutere in maniera autenticamente accurata e attenta (ed evidentemente documentata e approfondita) di “Rivoluzione/Transizione Digitale”, non si è chiamati semplicemente ad argomentare e disputare di Hardware e Software bensì, a partire da questo sfondo *banalmente* tratteggiato, di *Humansware*.

Sotto questo profilo, il concetto di OnLife, introdotto anni fa dal gruppo di lavoro coordinato da Luciano Floridi<sup>1</sup>, rappresenta una valida categoria/prospettiva non solo per immergersi immediatamente in *medias res*, ma anche per afferrare intuitivamente i tratti di quella che è già stata una vera e propria (ulteriore in seno alla storia della nostra specie) svolta antropologica. Anzi, di una morfo-

<sup>1</sup> Cfr. *The Onlife Manifesto. Being Human in a Hyperconnected Era*.

genesi co-evolutiva che ci dice (una volta di più) di nuove condizioni *per* l'umano e, dunque, di un'inedita condizione umana. Tant'è che essere umani, nella cosiddetta epoca della Rivoluzione e/o Transizione Digitale, ovvero nell'era dell'iperconnessione senza soluzione di continuità, richiederebbe di cominciare ad accedere a un'accezione nuova di esistenza, nei termini veri e propri di *MetaLife*<sup>2</sup>.

Pur tuttavia, se è vero com'è vero che è caduta la distinzione tra Online e Offline perché è a monte che la stessa distinzione tra virtuale e reale ha perso di senso, per cui e in altri termini in una vita vissuta *OnLife*, virtuale è reale e reale è virtuale, bisogna tener presente che ciò non rappresenta il motivo originario della svolta antropologica a cui si accennava, bensì ne costituisce l'acme e, dunque e in qualche modo, finanche il momento di definitivo compimento. E che, in buona sostanza, con le accelerazioni che abbiamo planetariamente registrato durante la Pandemia del 2020, non abbiamo impattato un momento germinale ma, appunto, risolutivo e, perciò, addirittura conclusivo.

Eravamo già pronti a tutto ciò: sia a livello infrastrutturale che, evidentemente, esistenziale.

Viviamo, infatti, in un tempo che è il *Day After* di una processualità durata, solo nel suo ultimo segmento accelerativo, all'incirca un quarto di secolo, in cui le tecnologie dell'informazione e della comunicazione non rappresentano più un semplice strumento tra gli altri, ma costituiscono il tessuto stesso della realtà e, indi, della nostra stessa eventualità e possibilità persistentiva. Tessuto, perciò, non meramente connettivo e mediativo, bensì propriamente essenziale, organico.

<sup>2</sup> Su questo mi permetto di rinviare al mio *MetaLife. Sapiens nell'epoca del Digitalicene*.

I cosiddetti ambienti digitali non si limitano a essere mero tramite e indi a mediare le nostre relazioni, i nostri saperi o le nostre pratiche lavorative: li trasformano, li generano, li implementano e finanche li sostituiscono, al punto tale che ogni angolo, ogni sfaccettatura, ogni elemento atomico dei nostri ecosistemi antropici, nell'esserne intessuto, trasfigura il nostro (vecchio) Mondo Esterno in un Ecosistema Digitale permanente.

Qui s'impone la necessità di transitare per uno snodo che, sovente, si assume come del tutto scontato, e relativo al significato, apparentemente immediato, da conferire all'aggettivo 'digitale' allorché si discute, appunto, di "Rivoluzione/Transizione Digitale".

Siamo portati a considerare 'digitale' come un qualcosa costantemente alla portata, quantomeno nel suo contenuto basilico. Forti, infatti, del fatto che 'digitale' deriva dal latino *digitus* (dito), che direttamente rinvia al concetto di numerico/numerabile in ragione della capacità delle dita di una mano di consentire alle persone di contare, con grande naturalezza lo rendiamo con 'quantificabile', 'calcolabile'. Non c'è dubbio che ciò non sia scorretto. Tuttavia, forse, troppo sintetico e parziale.

In prima istanza sfugge che Digitale non svolge propriamente una funzione attributiva bensì, nella fattispecie di "Rivoluzione/Transizione Digitale", costituisce l'autentico soggetto della locuzione. Ora, non vi è dubbio che digitalizzazione sia da intendersi, dal punto di vista tecnico, come il processo mediante il quale informazioni analogiche – cioè continue, qualitative e situate – vengono convertite in rappresentazioni discrete e numeriche, leggibili da macchine e manipolabili tramite algoritmi.

Eppure, questa delimitazione, pur settorialmente corretta, risulta insufficiente se non collocata all'interno di una più ampia cornice. È del tutto evidente che la digitalizzazione non è solo un'operazione tecnica, ma un gesto epistemologico e, in buona sostanza, ontolo-

gico: essa ridefinisce il modo in cui il reale può essere conosciuto, trattato, conservato e trasmesso. Nel momento in cui la realtà viene tradotta in dati digitali, essa viene sottoposta a una forma specifica di astrazione che ne seleziona gli aspetti computabili, tralasciando tutto ciò che eccede la logica della quantificazione. Questo passaggio dal continuo al discreto implica una trasformazione del nostro rapporto con il mondo: ciò che prima era esperito, ora è processato; ciò che era localizzabile nello spazio e nel tempo, ora è replicabile, trasmissibile, sincronizzabile. La digitalizzazione, in questo senso, non è una mera trascrizione della realtà, ma una sua specifica codifica e ri-codificazione. Essa istituisce una nuova modalità di accesso all'essente nella sua totalità, in cui l'informazione diventa la forma privilegiata della realtà e l'essere stesso si riformula nei termini della mera computabilità. Dunque, quando in prima istanza si avanza la necessità di discutere, in maniera minimamente accurata, di "Rivoluzione/Transizione Digitale", ovvero quando si assume che il Digitale è il soggetto, appunto processualità in atto di digitalizzazione, bisogna assumere in tutta la sua portata che digitalizzare significa non solo tradurre, ma anche *trasformare* il mondo secondo i criteri della logica binaria e della formalizzazione algoritmica. Trasformazione che plasma il visibile, il pensabile e l'agibile.

Una vera e propria *Onto-tecnologia*.

Appare del tutto evidente, quindi, che il problema non è più (se mai lo è stato) quello di "usare" il Digitale, e neanche più semplicemente, quello di porci questioni più o meno serie e originali legate al vivere nel Digitale, bensì quello di assumere l'irredimibile dato – che fa qui e ora la premessa situazionale della nuova condizione umana – che siamo transitati, e da tempo, in un'era nuova, onto-tecnologica, quella del Digitocene. Ovvero una nuova epoca *geologica* in cui l'ambiente terrestre, nell'insieme delle sue caratteristiche fisiche, chimiche e biologiche, è plasmato, tanto su scala locale che globale dagli

effetti dell'azione umana a partire dal suo nuovo abito ontologico, digi-centrato. Una nuova epoca in cui, però, l'ambiente terrestre tende a trasformarsi sempre più in mega-spazio antropico digitalizzato (e dunque senza frontiere e limiti) in cui hanno preso a muoversi persino nuove *forme viventi digitali* come gli Agenti Intelligenti che, a loro volta forze ambientali a tutti gli effetti, hanno già i crismi di alterità assolute che *forgiano* «la nostra realtà fisica e intellettuale», *modificano* «la nostra autocomprensione», *cambiano* «il modo in cui ci relazioniamo con gli altri e con noi stessi», *aggiornano* «la nostra interpretazione del mondo»<sup>3</sup> e dunque ogni nostra declinazione di agentività.

Non è questa la sede per procedere in direzione anche di tale ulteriore specifico segmento relativo all'epoca geologica nuova qual è il Digitocene e che, lo si sarà intuito, vede come protagonista la flessione corrente, quella relativa al pluriverso-AI.

Tuttavia, le considerazioni che ruotano e ruoteranno attorno al Digitocene e, perciò, a quella che è a tutti gli effetti la nuova condizione dell'umano, sono sì un attimo prima segnatamente al pluriverso-AI, ma nei termini cui facevo cenno poc'anzi: questa rivoluzione antropologica non è all'inizio, bensì è a uno stadio avanzatissimo, per non dire conclusa. Per un verso. E già, perché dall'altro, senza l'assunzione di ciò, non saremo mai in grado di comprendere non solo il fenomeno 'pluriverso-AI' ma, correlativamente, l'incidenza co-evolutiva ulteriore che sta avendo, e sempre più avrà, sulla nostra specie.

Dunque, Digitocene e nuova condizione dell'umano.

<sup>3</sup> L. Floridi, *The Fourth Revolution. How the Infosphere is Reshaping Human Reality*; trad. it. *La quarta rivoluzione. Come l'infosfera tra trasformando il mondo*, p. IX.

Preliminarmente si potrebbe dire che Digitocene richiama, da un lato, il mondo della computazione – ovvero il Digitale come logica informazionale –, e dall'altro si ricollega alla famiglia di concetti che hanno tentato, negli ultimi anni, di denominare l'epoca attuale in modo più adeguato rispetto alla semplice scansione cronologica.

Su questo secondo snodo, solo apparentemente d'etichetta, è necessario soffermarsi. Non c'è dubbio che quello di Antropocene abbia rappresentato, nel dibattito contemporaneo, un concetto chiave per pensare la co-estensione tra crisi ecologica e storia del soggetto agente Homo. Lungi dall'essere una mera designazione geologica, Antropocene si è affermato come una diagnosi culturale ed esistenziale: l'umano come forza tellurica, capace di modificare irreversibilmente i cicli bio-geo-chimici della Terra, svolgimento «iniziato quando *Homo Sapiens* divenne una superpotenza geologica, facendo imboccare alla Terra un nuovo percorso del suo lungo sviluppo. L'Antropocene è un punto di svolta nella storia dell'umanità, nella storia della vita e nella storia della Terra stessa. È un nuovo capitolo nella cronaca della vita»<sup>4</sup> in cui l'uomo non è più un soggetto tra gli altri, ma una potenza planetaria, inscritta nelle stratificazioni geologiche come fossile futuro del proprio disastro.

Tuttavia, è possibile domandarsi se l'Antropocene, nel porre particolare enfasi sul potere distruttivo dell'umano, abbia colto e intercetti interamente la natura delle trasformazioni in atto e, come si diceva, per certi versi – segnatamente alla nuova condizione umana – già compiute.

Detto altrimenti: mi sembra oramai evidente che si rende necessario integrare questa lettura geo-materiale con una riflessione/

<sup>4</sup> S. L. Lewis - M. A. Maslin, *The Human Planet. How We Created the Anthropocene*; trad. it. *Il pianeta umano. Come abbiamo creato l'Antropocene*, p. XI.

declinazione sulla dimensione informazionale, immateriale, computazionale che, sempre più, non semplicemente *governa* le forme della vita, della conoscenza e dell'esperienza, ma letteralmente fa, perché la struttura dall'interno, intimamente, la vita qui e ora. Proprio James Lovelock lo ha suggerito e prospettato, allorquando ha parlato di Novacene «per descrivere le caratteristiche di uno dei periodi più cruciali nella storia del nostro pianeta e forse anche del cosmo»<sup>5</sup>, ovvero della fase in cui la tecnologia da noi inventata è/sarà in grado di superare il nostro controllo generando intelligenze più grandi e con una caratteristica cruciale, cioè una rapidità di ragionamento molto maggiore rispetto alla nostra.

In questo solco, allora, Digitocene per spostare definitivamente l'attenzione su un altro asse trasformativo decisivo, quello della digitalizzazione pervasiva, intesa non solo come fenomeno tecnico e/o infrastrutturale, bensì come ridefinizione delle condizioni ontologiche e gnoseologiche della nostra presenza al mondo.

Il Digitocene è l'epoca in cui il Digitale smette di essere un mezzo per diventare a tutti gli effetti non semplicemente un ambiente, bensì *l'ambiente*, ovvero la condizione di possibilità propria affinché si agisca e, dunque, in certo qual modo si esista.

Non siamo più *utenti* di tecnologie digitali: siamo *immersi* in esse, ovvero *siamo* in esse. La nostra stessa possibilità d'essere si determina nella sua compiuta e completa effettività se e solo se siamo strutturalmente attraversati, modellati, iscritti in reticoli informazionali che operano come dispositivi di soggettivazione. La progressiva, e senza soluzione di continuità, datificazione del reale – la sua traduzione e trasfusione in codici computazionali, la sua analisi al-

<sup>5</sup> J. Lovelock, *Novacene. The Coming Age of Hyperintelligence*; trad. it. *Novacene. L'età dell'iperintelligenza*, p. 37.

goritmica, nonché la sua manipolabilità simulativa –, la sua costante rappresentazione in linguaggi computazionali, produce una riconfigurazione radicale delle nozioni di mondo, realtà e, indi, delle stesse basi dell'esperienza da parte di Homo.

Mentre l'Antropocene può essere compreso attraverso la metafora dello *strato geologico* — l'uomo come forza fisicamente stratificante — il Digitocene implica un altro tipo di stratificazione: quella *logico-computazionale*.

In tal senso si potrebbe dire che il Digitocene non è un'epoca che *succede* all'Antropocene, ma un'epoca che lo *trasforma dall'interno*, mutandone i vettori epistemici e le strutture operative.

Com'è del tutto evidente, l'ambiente fisico resta, ma viene sdoppiato dalla e nella sua rappresentazione digitale: reti, algoritmi, piattaforme, modelli predittivi, ambienti virtuali, interfacce neurali, etc.

È qui in questo contesto che emerge un nuovo tipo di soggettività che presenta tratti radicalmente inediti. I tratti, appunto, della nuova condizione dell'umano. Questi è l'*InfoSapiens*, ovvero non più semplicemente Homo Sapiens come produttore e consumatore di simboli, ma soggetto-informazione, continuamente generato e rigenerato da flussi di dati, da feedback algoritmici, da sistemi di riconoscimento e quantificazione. L'*InfoSapiens* è al tempo stesso utente e prodotto, agente e traccia, autonomia apparente e dipendenza strutturale. La sua stessa corporeità è soggetta a protocolli di tracciabilità, analisi, ottimizzazione. Non è un soggetto che utilizza il Digitale come mero *instrumentum* alla stregua di un qualsiasi ente artificiale, ma un soggetto, sempre più, *in toto Digitale* e, indi, soggetto *al* Digitale.

Come segnalavo, è del tutto evidente che la posta in gioco qui non è *stricto sensu* tecnologica, bensì ontologica.

Con il Digitocene è infatti messa in discussione l'idea stessa di umanità come continuità e, per certi versi, evolutiva unità epistemica e normativa. La centralità dell'umano, già in qualche modo

concettualmente erosa dall'evoluzionismo e dal co-evoluzionismo e, dunque, dall'Antropocene attraverso la nozione di interdipendenza ecologica, viene ora ulteriormente destabilizzata dalla sua dissoluzione informazionale in direzione di un antropo-decentramento non semplicemente di fatto nel solco di un orizzonte *naturale*, ma concretamente mai sperimentato prima, giacché frutto dell'operatività di Sapiens stesso in direzione di una sua marginalizzazione profonda.

La coscienza, il linguaggio, la decisione, la memoria, la previsione, tutto ciò che, in un circuito autonarrativo, ci siamo attribuiti al fine di definire l'umano e che, comunque, a pieno titolo abbiamo assunto come ingredienti fondamentali della nostra strategia persistente di specie, è oggi oggetto di esternalizzazione tecnica e automatizzazione algoritmica.

In quest'ottica, il soggetto del Digitocene vive letteralmente in un regime di *eteronomia computazionale*.

Il concetto di Digitocene, in questo senso quindi, assume anche un valore critico. Difatti non si limita a constatare un fatto, a cifrare un passaggio spazio-temporale, ma problematizza l'epoca. Non fotografa una trasformazione ma, nella trasformazione, al cuore della trasformazione, impone finanche di ripensarne le condizioni, le implicazioni e gli effetti. Esige ad esempio una nuova filosofia della tecnica, una nuova teoria della soggettività, finanche, con buona pace dei de-ontologisti e dei pragmatisti, una nuova etica informazionale; impone la necessità non solo di innovativi con concettuali – a titolo esemplificativo, basterebbe elencare tutte le locuzioni controfattuali alle quali ricorriamo per goffamente orientarci un minimo –, ma la reversibilità e riscrittura di intere catene teorico-pratiche che hanno intessuto la filosofia stessa, l'etica, la sociologia, l'antropologia culturale, la psicologia etc.

Richiede, in ultima istanza, di interrogare non solo *ciò che possiamo fare* con il Digitale, ma *ciò che il Digitale sta facendo a noi e di noi*.

Pertanto, il Digitocene si presenta, anche e soprattutto, come un sempre più compiuto e inglobante Ecosistema Cognitivo e Simbolico nel quale siamo completamente immersi. Un Ecosistema che non ci circonda e/o contiene semplicemente, ma che ci plasma nei termini di un'ipertrofica dinamica co-trasformativa; che non ci aliena soltanto dall'originario dato materiale, ma ci costituisce, costruisce e ricostruisce costantemente e senza soluzione di continuità, in inedite forme composite effettive. In esso, i nostri modi di pensare, sentire, conoscere, immaginare e creare si sviluppano in una costante interazione con dispositivi interagenti, reti neurali artificiali, ambienti simulati e quant'altro. La realtà stessa diventa fluida, codificata, tracciabile. E la nostra identità non è più una narrazione organica, ma un mosaico distribuito di dati, immagini, interazioni e mutualità, plurime versioni digitali che cifrano, né più né meno, che l'impossibilità oramai a stabilire un confine non più e semplicemente tra i vecchi concetti di reale e virtuale, ma alle nostre stesse potenzialità e capacità agentive. Se è vero com'è vero che *ago ergo sum*, sono le nostre stesse impalcature identitarie che si riedificano costantemente in un'impossibilità a stabilire punti di arresto di possibile e plausibile soluzione di continuità tra le nostre vecchie categorie di concretamente e virtualmente tangibile.

Come si evince, allora, la vera mutazione/alterazione (nel senso che ci iscrive anche in vorticosi orizzonti trasmutativi dell'Alter) del Digitocene non è solo di ordine tecnico-tecnologico bensì, e indefettibilmente, di stampo antropologico.

Il soggetto che abita questa nuova epoca è un diverso essere ibrido, ovvero un umano che vive in simbiosi con macchine cognitive, che delega funzioni fondamentali – dalla memoria all'attenzione, dalla selezione delle notizie alla ricerca di cose e affetti – a sistemi algoritmici. La soggettività stessa viene continuamente ridefinita da processi di profilazione, categorizzazione e raccomandazione, che

alimentano finanche un'economia dell'informazione basata sulla predizione e sul controllo.

È qui perciò, nel cuore del Digitocene, che prende forma la nuova figura dell'umano, l'InfoSapiens.

Come si accennava, non si tratta di una semplice targhetta e/o qualifica per indicare l'“essere umano digitale”, connesso, tecnologicamente potenziato. L'InfoSapiens è qualcosa di più profondo: è il soggetto che, nella simbiosi continua con gli ambienti digitali, dall'interno di scatole cinesi di parcellizzati Ecosistemi Digitali, a loro volta all'interno di un sempre più compiuto e inglobante Ecosistema Cognitivo e Simbolico Planetario, trasforma radicalmente il proprio modo di conoscere, decidere, ricordare, percepire e relazionarsi.

In ultima istanza, il proprio modo d'essere.

Appare persino considerazione scontata e convenzionale quella in ragione della quale non siamo più gli stessi esseri cognitivi di venti o trent'anni fa. L'Intelligenza Artificiale, i motori di ricerca, i sistemi di raccomandazione, le reti sociali algoritmiche non si limitano ad accompagnare le nostre attività mentali: talvolta ne diventano prolungamenti e mediatori strutturali, altre addirittura coesenziali motivi di slittamenti trasfigurativi in una compressione spazio-temporale mai conosciuta prima dalla nostra specie.

Delegare all'esterno – a strumenti, simboli, dispositivi – alcune funzioni cognitive non è certo una novità nella storia dell'umanità. Ogni cultura, ogni tecnologia, nell'arco della breve storia di Sapiens, ha avuto il potere di estendere e trasformare le possibilità cognitive e dunque dell'agire umano: la scrittura ha esternalizzato la memoria; la stampa ha democratizzato il sapere; il calcolo meccanico ha potenziato la logica; le infinite applicazioni solo di questi semplici esempi hanno avuto la capacità di plasmare irrevocabilmente i nostri modi di stare al mondo.

Ma ciò che è accaduto e continua ad accadere oggi è inedito per velocità, pervasività e grado di interattività. Un'interattività organica che nel destrutturarci costantemente ci ristrutturava ogni volta, e su basi, anche solo impercettibilmente, mutevoli e cangianti.

L'InfoSapiens è un essere umano che non agisce più da solo, neanche nel proprio ingegno e nella propria immaginazione, ma è in costante collaborazione con macchine cognitive: sistemi intelligenti che aiutano a orientarsi, a scegliere, a comprendere. Quando cerchiamo un'informazione, non ci immergiamo in un processo autonomo di esplorazione, ma chiediamo a un algoritmo di mostrarci la "risposta migliore". Quando decidiamo cosa guardare, ascoltare, leggere, amare, lo facciamo nell'ecosistema di turno dove ogni opzione è già filtrata, ordinata, personalizzata. La conoscenza diventa *curata, servita, anticipata*.

Conoscere, per l'InfoSapiens, ad esempio, non è più una pratica lenta, riflessiva, articolata. È un gesto di accesso. Il sapere si trova, si consulta, si verifica, si dimentica. È sempre disponibile, ma sempre più estraneo alla sua fattura diretta. Non lo custodiamo, non lo elaboriamo, non lo integriamo nella memoria profonda: lo interroghiamo, lo consumiamo, lo abbandoniamo. I motori di ricerca e i sistemi di risposta istantanea sostituiscono la costruzione del giudizio. La competenza si disgrega nel flusso, e l'autorevolezza cede il passo alla visibilità.

Anche la capacità del ricordare e della rammemorazione si trasformano. La memoria, una volta impastata di emozione, di narrazione, di selezione soggettiva, è ora archiviata fuori di noi. I nostri ricordi vivono in Cloud, in gallerie fotografiche automatizzate, in "ricordi" che ci restituiscono il passato come una notifica. La memoria si fa algoritmo, e con essa si modifica finanche il nostro rapporto con il tempo, che perde la sua profondità e si appiattisce sull'istante mostrato, condiviso, suggerito.

Ugualmente la meccanica, l'elaborazione e il senso stesso del decidere assumono tutt'altra fisionomia. L'InfoSapiens non decide più in un vuoto deliberativo, ma in uno SmartEcoSystem – suggerirei di chiamare così i parcellizzati Ecosistemi Digitali di cui prima benché, in altra sede, allorquando la questione sarà proprio il 'pluriverso-AI', si renderà necessario entrare più nello specifico, anche dal punto di vista analitico – già orientato, predisposto. Ogni scelta è incorniciata da opzioni filtrate, selezionate, previste. Gli algoritmi non impongono, ma modellano il campo del possibile. Ciò che appare come libertà è spesso una libertà limitata all'interno di un perimetro invisibile, costruito a partire dai nostri dati, dai nostri comportamenti precedenti, da ciò che "ci somiglia". Scegliamo, certo. Ma in uno spazio che non abbiamo contribuito a costruire se non in maniera massivamente filtrata e decisamente indiretta e surrogata. In fondo, l'eteronomia è mascherata da personalizzazione.

L'InfoSapiens vive inoltre in una condizione percettiva totalmente interfacciata. Il mondo gli arriva sotto forma di schermo, di suono elaborato, di realtà aumentata, di notifica visiva. La realtà perde la sua resistenza, il suo peso, la sua alterità. Tutto è immediato, manipolabile, zoomabile, archiviabile. Il corpo stesso, lungi dall'essere il primo orizzonte dell'esperienza, è sempre più una superficie da monitorare, ottimizzare, mostrare. Ci alleniamo con App che contano i passi, dormiamo con SmartWatch che misurano il sonno, respiriamo secondo ritmi suggeriti da interfacce animate. Il corpo perde la sua opacità, finanche la sua supposta sacralità, e diventa una Dashboard.

Anche le forme della relazione non sono più quelle che erano. L'InfoSapiens incontra l'altro sempre attraverso una mediazione tecnologica, e spesso attraverso una molteplicità di mediazioni sovrapposte. Il volto dell'altro è uno schermo, una foto, un avatar, una voce artificiale. L'intimità si consuma nella connessione, ma raramente nell'esperienza profonda. L'incontro è mediato da logiche di ma-

tching, da algoritmi di affinità, da metriche di engagement. Amare, soffrire, litigare diventano processi parzialmente automatizzati, tracciabili, ottimizzati.

Questa permanente delegazione, che è anche e soprattutto delegazione cognitiva, per quanto funzionale e, a tratti sempre più ampi, confortevole, ha conseguenze profonde. Da un lato, consente senz'altro di liberare energie, permettendoci, ad esempio, di concentrarci proprio su processi più creativi, più riflessivi, più relazionali... ovviamente da consumare in ulteriori SmartEcoSystems, dacché il mutevole (e ideale) prolungamento dei precedenti. Dall'altro – e senza condizionale – ci espone al rischio di un'eteronomia invisibile, in cui ciò che crediamo di scegliere è in realtà scelto per noi, in modo impercettibile e personalizzato.

L'InfoSapiens è progressivamente un soggetto guidato, più che un soggetto agente: un io fluttuante tra notifiche, feed, raccomandazioni, immerso in un presente continuo fatto di stimoli, ma povero di profondità.

A ben pensarci: un oggetto più che un soggetto.

Per quanto abbia deciso, per i motivi precedentemente esposti, di non aprire in maniera diretta e approfondita al pluriverso-AI, le considerazioni di Lorenzo Perilli costituiscono fruttuoso supporto per dare comunque un primo contenuto proprio a questo aspetto: «l'approccio algoritmico alla realtà cambia il nostro rapporto con il mondo. Con le cose. Stiamo entrando in una nuova epoca, dalla quale non si tornerà indietro. Con l'AI, oggi, non cambia solo il gioco: cambiano anche i protagonisti del gioco e i loro rapporti. Si invertono i ruoli. L'AI ha una implicazione che nessuna delle svolte che hanno caratterizzato la storia dell'uomo ha mai avuto: *rovescia il rapporto tra soggetto e oggetto*. Il soggetto diventa passivo, l'oggetto diventa attivo. Il rapporto tra soggetto e oggetto non è più contrassegnato da una linea di demarcazione e da ponti che favoriscano

comunicazione e interazione. Esso assume carattere ibrido, in cui i processi algoritmici devono essere considerati come azioni *autonome* di un oggetto che possono avere esiti tanto concreti quanto quelle dell'uomo. Gli algoritmi di AI costituiscono così il nucleo di una nuova logica dell'oggetto, che governa un agire indipendente dall'uomo ma diverso da quel libero arbitrio che all'uomo la filosofia e la religione hanno assegnato. Questa logica permea oggi ogni ambito della vita associata, che da essa dipende in misura sempre maggiore. Gli algoritmi, usciti dalle confortevoli pagine degli *Elementi* di Euclide e da quelle dell'ottimo Eudosso di Cnido e fattisi strumento delle macchine digitali, possiedono una intrinseca capacità di azione basata su decisioni autonome prese come reazione a stimoli esterni, che li rende diversi da ogni altra invenzione dell'uomo: mai, finora, ci siamo trovati di fronte a una così radicale ridefinizione del nostro ruolo nel mondo. Mai, finora, la domanda "Che cos'è l'uomo?" era stata così profondamente sollecitata non dalla riflessione metafisica o antropologica o religiosa, né dalla scienza, ma dalla tecnologia»<sup>6</sup>.

È il modo in cui costruiamo noi stessi a esser totalmente e radicalmente mutato.

Nell'epoca predigitale, l'identità si formava lentamente, nel tempo, nella narrazione di sé e nel confronto diretto con l'altro, vedendoci, comunque, in cerchi concentrici di registri co-esistentivi spazio-temporalmente scanditi, attori attivi. L'identità dell'InfoSapiens è invece distribuita: un insieme di profili, interazioni, dati, metriche di performance, immagini condivise. La rappresentazione di sé è continua, pubblica, performativa. Viviamo – spesso inconsapevolmente – all'interno di una logica di quantificazione, dove emozioni,

<sup>6</sup> L. Perilli, *Coscienza artificiale. Come le macchine pensano e trasformano l'esperienza umana*, p. 27.

gusti, desideri, perfino l'umore, sono tracciati, misurati, analizzati, e infine restituiti a noi sotto forma di grafici, ricordi automatizzati, previsioni, suggerimenti.

Ma c'è anche un altro aspetto da non trascurare.

L'InfoSapiens, com'è evidente, non è soltanto un passivo-inerte dell'automazione cognitiva. È anche un nuovo tipo di essere pensante, capace di sviluppare forme di intelligenza aumentata, di cooperazione interspecie, di co-evoluzione con le macchine.

In definitiva, l'InfoSapiens è il protagonista – consapevole o meno – dell'epoca del Digitocene. È colui che vive al liminale tra l'umano-biologico e il computazionale, tra la libertà e la programmazione, tra la memoria personale e la memoria algoritmica. Capire questa nuova figura non significa solo descrivere un mutamento, ma aprire un confronto decisivo sul destino stesso della soggettività nell'epoca delle Intelligenze Artificiali.

Da quanto sin qui detto mi sembra impossibile non constatare che questo mutamento, sottile e impercettibile per molti versi nell'ultimo trentennio e, tuttavia, allo stato irreversibile, ha rapidamente attraversato la nostra condizione, modificandola dall'interno. Non si tratta solo di avere nuovi strumenti, nuove abitudini, nuove interfacce. Non è questione, semplicemente, di provare *sanamente* a moderare il nostro essere connessi. Viviamo dentro una trasformazione di ordine ontologico, che tocca la radice stessa del nostro modo d'essere e, perciò, di essere protagonisti nel Mondo Esterno.

Come si diceva, l'InfoSapiens non è solo un nuovo calco dell'umano che utilizza strumenti digitali, ma un umano che nasce, vive, sente, pensa, agisce e si racconta in SmartEcoSystems. Non esiste più un "dentro" umano che si serve di un "fuori" tecnologico. Quel che s'accompagna all'aggettivo 'digitale' non individua più un mezzo: è condizione, struttura. L'InfoSapiens è ciò che emerge da questa immersione radicale, il prodotto simbiotico di un'umanità che ha

delegato una parte crescente delle sue funzioni vitali a macchine informazionali e che ha trovato, in questa delega, non solo efficienza, ma anche una sempre più emergente nuova forma di sé.

L'umano digitale non è un Sapiens potenziato: è, a tutti gli effetti, una nuova figura dell'umano, una mutazione interna che non può essere compresa se non accettando che siamo già, integralmente, altro da ciò che eravamo. La mente, il corpo, il tempo, l'identità, la relazione: ogni elemento è stato riscritto, riconfigurato, ricodificato nella logica dell'InfoSfera.

Ma ciò che più conta è che tutto questo non è una distorsione passeggera, né una perdita da colmare inseguendo qualche mitica immagine di uno stazionario-umano. È una metamorfosi compiuta, un cambio di paradigma. Siamo entrati, senza accorgercene, in una nuova forma di umanità, in cui l'elemento informazionale ha assunto funzione costitutiva. L'InfoSapiens non è un ibrido determinatosi a freddo tra l'uomo e la macchina, nell'incontro di due opposti, naturale e artificiale. Intrecciandosi in un tutt'uno insolubile, a tutti gli effetti – intreccio il cui *terreno* di gestazione è stata proprio la progressiva trasmutazione e rappresentazione del Mondo Esterno in versioni computative – l'InfoSapiens è una nuova creatura, nata appunto dalla convergenza tra il biologico e il computazionale, tra l'evoluzione naturale e quella tecnologico-informazionale, tra soggettività biotica e codificazione.

Questo non significa che abbiamo perso ogni profondità, ogni autonomia, ogni presunta umanità, giacché come sempre e come da sempre, dietro queste connotazioni si cela nient'altro che una lunga, estenuante, a volte casuale, transitoria costruzione che non conosce fine.

Significa nient'altro che, come da sempre, l'umanità stessa è ora chiamata a ridefinirsi per riafferarsi. Tuttavia, a ridefinirsi dentro un nuovo Ecosistema Simbolico e Cognitivo Planetario.

L'InfoSapiens non è solo un dato di fatto, ma una figura da pensare, da interrogare, da comprendere. È la soglia su cui si gioca il nostro presente e il nostro futuro. Perché se è vero che siamo già cambiati, è anche vero che possiamo ancora decidere come abitare questo cambiamento: se subirlo come destino o trasformarlo in possibilità.

Accettare che siamo già InfoSapiens significa riconoscere che la questione del nostro tempo non è più “se” e “come” diventare post-umani, come per lungo tempo s'è creduto, anche aprendo a traiettorie fantasiose e che, non a caso, sono durate il tempo di una stagione di un batter d'ali.

In questo consiste, forse, quella che è la sfida più radicale del Digitocene: ripensare l'umano senza nostalgia e senza ingenuità, assumendo che la nostra forma non è data una volta per tutte, ma continuamente riscrivibile, fragile, aperta. E che solo un pensiero capace di abitare questa instabilità saprà ancora dire qualcosa di vero su di noi.

Ripensare l'umano, perciò, consapevoli di essere immersi in un'epoca che non si limita a trasformare i nostri strumenti, ma che ridefinisce silenziosamente ciò che siamo. Non è solo la quantità di tecnologia presente nella nostra vita quotidiana a fare la differenza, né la crescente sofisticazione degli apparati digitali. A essere mutata ancora è la nostra condizione ontologica, la struttura stessa della soggettività, il modo in cui abitiamo il mondo, gli altri e noi stessi. Appare definitivamente chiaro che InfoSapiens non è una variante accessoria e/o estemporanea dell'Homo Sapiens, ma una figura antropologica emergente, già inscritta nell'orizzonte della nostra esperienza quotidiana, che riflette l'intima simbiosi tra umanità-biologica e ambienti digitali.

L'InfoSapiens è il prodotto – e al tempo stesso il produttore – di questa nuova *ecologia* cognitiva e, come detto, la sua mente non è più solo biologica, ma ibridata con dispositivi intelligenti che ne accom-

pagnano, anticipano o sostituiscono le funzioni fondamentali. Il suo corpo è mappato, tracciato, ottimizzato. La sua memoria è distribuita in archivi esterni. La sua identità è moltiplicata in profili digitali. Nulla, o quasi, nella sua vita psichica e relazionale resta immune dalla mediazione-generazione tecnologica.

Non è un caso che, oramai, questa condizione non la si viva come innaturale: al contrario, è il nuovo sfondo di *naturale normalità*.

È ciò che siamo.

Lo si ribadisce, per cominciare a tirare qualche filo: l'InfoSapiens conosce, incontra, ama, discute in ambienti digitali che regolano la visibilità, l'affinità, la prossimità. Le relazioni sono modulate da algoritmi, scandite da metriche, disposte secondo logiche di engagement. La presenza dell'altro è incarnata da avatar, immagini, notifiche. L'intimità è compressa nel tempo dell'interazione. La vulnerabilità dell'incontro diretto è rarefatta. Ciò che appare è, spesso, ciò che funziona. E ciò che funziona è, sempre più spesso, ciò che rassicura.

Tutto questo non rappresenta semplicemente un effetto collaterale del progresso tecnologico. Si tratta, piuttosto, di una riconfigurazione profonda della forma dell'umano.

L'InfoSapiens non è un incidente di percorso, né una deviazione: è una metamorfosi già compiuta, benché non ancora del tutto tematizzata se non per microaree, specie quelle che presentano risvolti di ordine clinico. La soggettività è divenuta post-biologica, distribuita, ambientale. La nostra identità non si fonda più su un centro stabile, ma su un sistema dinamico di dati, interazioni e retroazioni algoritmiche. La mente è espansa, il corpo è interconnesso, l'essere è mediato.

E pur tuttavia, non si tratta di cedere a qualche scialba formulazione tecno-deterministica, né di cullarsi nell'utopia della connessione totale, casomai sul modello di qualche lungometraggio di fantascienza. Al contrario, riconoscere l'InfoSapiens come figura co-

stitutiva del Digitocene significa assumere consapevolmente che ogni progetto etico, politico, educativo o estetico dell'oggi deve partire da una domanda nuova: *che cosa significa essere umani in un mondo in cui il Digitale non è fuori di noi, ma ci costituisce?*

La sfida non risponde all'infantile aspirazione e fallimentare ambizione di proteggere una presunta (perché mai esistita) umanità originaria, né, di converso, di ingenuamente abbandonarsi al flusso accelerato dell'innovazione, evangelizzatrice di sorti progressive paradisiache. La sfida è quella di abitare criticamente la nostra nuova condizione, trovando forme inedite di libertà, di profondità, di presenza. Se l'InfoSapiens è la nostra nuova condizione in quanto umani, allora non ci resta che pensarlo a fondo, comprenderne le ambivalenze, esplorarne le potenzialità. Forse, è solo da qui che può ripartire un umanesimo all'altezza del tempo presente.

È solo da qui che può e deve partire una riflessione seria su quello che sembra essere l'argomento (da gazzette) del momento, ovvero quello relativo all'Intelligenza Artificiale Generativa che, invece, forma inedita di *Intelligenza Condivisa*<sup>7</sup>, si configura come totale Alterità che assurge, sempre più, nella sua capacità e autonomia funzionale e agentiva, a peculiare entità agente protagonista in seno agli SmartEcoSystems.

In chiusura mi sia consentito avanzare un'ultima considerazione, in qualche modo destinata – tra coloro i quali si sentono ancora totalmente assorbiti nella bolla dell'altra transizione, quella ecologica – ai refrattari (quando non ostili) all'apertura di tavoli di lavoro e riflessione meta-disciplinari. In questa sorta di ideale flessione dall'Antropocene al Digitocene si gioca una scommessa teorico-effet-

<sup>7</sup> L'espressione proviene dall'acutissimo testo di E. Mollick, *Co-intelligence. Living and Working with AI*.

tiva fondamentale, quella di comprendere che la questione ecologica e quella informazionale non sono scindibili. L'impatto geologico dell'umano è ormai indissociabile dal suo impatto computazionale. L'Intelligenza Artificiale, il Machine Learning, la biometria, le piattaforme digitali, ovvero tutto ciò che *fa* quel sempre più compiuto e inglobante Ecosistema Cognitivo e Simbolico Planetario nel quale siamo completamente immersi e da cui è emersa la nuova condizione dell'umano quale InfoSapiens, sono tutti dispositivi che, lungi dall'essere neutri, ristrutturano la forma del mondo, selezionano ciò che può esistere, determinano ciò che può essere conosciuto, previsto, gestito.

Detto in maniera ancora più ruvida: nel Digitocene, la lotta per la sopravvivenza non è soltanto ecologica ma, evidentemente, epistemica. La posta in gioco è la possibilità stessa di un sapere critico, di un'autonomia soggettiva, di una responsabilità condivisa in un mondo reso opaco dalla complessità algoritmica. Pensare il Digitocene significa allora elaborare una nuova alleanza tra filosofia, scienze cognitive, teoria dei media, ecologia e politica. Una nuova *ecologia* della conoscenza, capace di pensare l'umano *dentro* l'InfoSfera, senza dissolverlo, ma nemmeno mitizzarlo.

Solo così saremo in grado di cominciare a comprendere il nostro tempo non come una fine, ma come una soglia. Una soglia fragile come sempre, certo, ma anche realmente generativa. Ovviamente, se saremo in grado di darle un nome giusto, e soprattutto, un senso.

## Bibliografia

- L. Floridi, *The Fourth Revolution. How the Infosphere is Reshaping Human Reality*, Oxford, Oxford University Press, 2014; trad. it. *La quarta rivoluzione. Come l'infosfera tra trasformando il mondo*, Milano, Raffaello Cortina, 2017.

Gianluca Giannini

- G. Giannini, *MetaLife. Sapiens nell'epoca del Digitocene*, in *Le parole della filosofia. Le metamorfosi del vocabolario del pensiero nella storia*, a cura di A. Motta e L. Palumbo, Napoli, FedOAPress, 2024, pp. 443-453.
- S. L. Lewis - M. A. Maslin, *The Human Planet. How We Created the Anthropocene*, London, Pelican, 2018; trad. it. *Il pianeta umano. Come abbiamo creato l'Antropocene*, Torino, Einaudi, 2019.
- J. Lovelock, *Novacene. The Coming Age of Hyperintelligence*, Cambridge, The MIT Press, 2019; trad. it. *Novacene. L'età dell'iperintelligenza*, Torino, Bollati Boringhieri, 2020.
- E. Mollick, *Co-intelligence. Living and Working with AI*, London, Portfolio, 2024; trad. it. *L'intelligenza condivisa. Vivere e lavorare insieme all'AI*, Roma, Luiss University Press, 2025.
- L. Perilli, *Coscienza artificiale. Come le macchine pensano e trasformano l'esperienza umana*, Milano, il Saggiatore, 2025.
- The Onlife Manifesto. Being Human in a Hyperconnected Era*, edited by L. Floridi, Heidelberg-New York-Dordrecht-London, Springer, 2015.

Gianluca Giannini

Università degli Studi di Napoli Federico II

gianning@unina.it

<https://orcid.org/0000-0002-9614-3713>

EMILIANO GRIMALDI

*L'intelligenza Artificiale e l'Educazione.  
Appunti per una sociologia pubblica*

*Abstract:* Artificial Intelligence is the new hype in the public debate on education. As already happened with other technologies in the past, we assist to the proliferation of techno-solutionist discourses which attribute to AI the power to revolutionise education, for the better or for the worst. In this chapter, using the lenses of the sociology of education, I will discuss in detail the importance of escaping from any techno-solutionist temptation and attributing a renewed centrality to the social life and the ontological politics of AIED in the public debate on education. This means, in the perspective advocated here, to take seriously the difference in the epistemological and pedagogical inscriptions which are distinctive of different AIED solutions, together with the socio-technical complexities implied in their enactment in the educational milieu. On this basis, the chapter highlights the potential contribution that such an approach, combined with an attitude towards constructive criticism, can give in: a) promoting a reflective use of AI in learning environments (*learning with AI*); b) encouraging the entry of AI into educational curricula (*learning about AI*); and above all c) recenter the educational in the design of learning experiences whose aim is not to instruct individuals to an instrumental use of a technology but rather allow them to become autonomous citizens in a world in which AI will be increasingly present in all crucial spheres of contemporary societies (*learning for human-AI collaboration*).

*Keywords:* Artificial Intelligence in Education and Learning (AIED), Ontological Politics, Social Life of AI, Public Sociology of AI in Education, Human-AI Collaboration.

1. *Il dibattito pubblico su Intelligenza Artificiale e Educazione. Uno sguardo sociologico*

*Teach For Tomorrow*

We help teachers and schools unlock their full potential through AI. Artificial Intelligence is poised to change teaching and learning forever. Artificial Intelligence (AI) has the potential to revolutionize teaching and learning as we know it. By using AI-powered tools and strategies, educators can personalize learning, improve student outcomes, and better prepare students for success in the digital age. Discover how AI can transform your district, school or classroom with AI for Education.

(<https://www.aiforeducation.io>)

Nel dibattito pubblico sull'educazione sono oggi sempre più ricorrenti enunciazioni come quelle riportate in epigrafe circa le capacità che l'Intelligenza Artificiale (IA) avrebbe di 'liberare' il pieno potenziale dell'educazione e trasformare per sempre e per il meglio le sue modalità. In uno scenario nel quale valutazioni sulle soluzioni tecnologiche oggi disponibili, proiezioni sul futuro più o meno immediato e forme di *hype* a tratti visionario si intrecciano in una curiosa e paradossale forma di futuro-presente<sup>1</sup>, sono diverse le proprietà taumaturgiche attribuite all'*AI in Education and Learning* (AIED): la personalizzazione dell'apprendimento, il miglioramento generalizzato dei suoi risultati, una migliore preparazione degli studenti in vista delle sfide poste dall'età digitale, la crescita della qualità dell'insegnamento o del suo carattere inclusivo, un uso efficiente delle risorse sul piano amministrativo-gestionale e, infine, la possibilità di una governance di precisione e/o predittiva dei processi educativi.

L'intreccio tra la presenza sempre più forte di queste voci nel dibattito sull'educazione e la crescente accelerazione degli sviluppi

<sup>1</sup> R. Kitchin, *Digital timescapes*.

tecnologici nel campo dell'IA, del *machine learning*, delle *learning analytics* e dell'automazione più in generale sta producendo effetti concreti sul piano delle pratiche educative, fuori e dentro la scuola. Al tempo stesso solleva dilemmi etici e apre lo spazio per controversie caratterizzate da forti polarizzazioni tra integrati e apocalittici<sup>2</sup>.

Come hanno recentemente sottolineato Holmes e Tuomi<sup>3</sup>, se gli sviluppi recenti dell'IA hanno generato grandi aspettative per il futuro impatto dell'AIED, spesso queste aspettative si basano su visioni 'ottimistiche' circa le sue attuali possibilità tecniche ed educative e, soprattutto, su approcci tecno-soluzionisti che finiscono per 'oggettivare' implicitamente forme di riduzionismo circa l'intrinseca complessità dei processi educativi e le forme e finalità dell'educazione nelle società contemporanee. A fronte di ciò, abbiamo ancora una disponibilità molto limitata di evidenze empiriche diacroniche e rigorose che ci consentano di dare riscontro in maniera scientificamente fondata a tali aspettative ottimistiche, o di esprimere giudizi fondati sui rischi attuali connessi all'uso delle diverse soluzioni di AIED<sup>4</sup>.

Quel che mi interessa qui sottolineare è come la tendenza verso il riduzionismo nel dibattito riguardi anche quelle che definisco, riprendendo Appadurai<sup>5</sup> e Williamson<sup>6</sup>, come la *vita sociale dell'IA in educazione* e la *politics ontologica dell'AIED*. Da un lato, infatti, al di fuori dei circuiti specialistici sembra molto scarsa l'attenzione alle iscrizioni epistemologiche e pedagogiche di cui sono portatrici le diverse soluzioni AIED oggi disponibili e presenti sul mercato

<sup>2</sup> B. Williamson et al., *Re-examining AI*.

<sup>3</sup> W. Holmes - I. Tuomi, *State of the art and practice in AI in education*.

<sup>4</sup> S. Nemorin et al., *AI hyped?*

<sup>5</sup> A. Appadurai, *The social life of things*.

<sup>6</sup> B. Williamson, *The social life of AI in Education*.

e al loro portato 'politico'. Dall'altro, ci si concentra sull'IA come strumento tecnico dai tratti definiti e invarianti, lasciando spesso e volentieri nel cono d'ombra del dibattito le complessità storiche e socio-tecniche dei processi attraverso i quali l'AIED e più in generale le tecnologie per l'automazione basate su IA entrano o sono *'plugged in'* nelle pratiche e nei *milieu* educativi.

Uno degli aspetti forse più problematici dell'attuale stato dell'arte del dibattito sull'AIED riguarda il fatto che, con qualche eccezione, siano questi due aspetti ad essere fondamentalmente marginalizzati nei processi di *agenda setting*, in un quadro in cui le controversie a cui si accennava in precedenza si incentrano prevalentemente sulla desiderabilità o meno dell'AIED, sulle sue ipotetiche potenzialità e/o sui rischi, sulle linee guida di *policy* e/o sui *framework* etici che dovrebbero regolare la sua introduzione. Il risultato di tale marginalizzazione sembra essere l'ennesima riproposizione di uno script discorsivo ben noto, nel quale l'introduzione dell'AIED, intesa come categoria monolitica la cui differenziazione interna viene spesso ignorata, diviene una necessità priva di alternative e/o una minaccia da combattere. Al tempo stesso, l'educativo diviene per l'ennesima volta uno spazio:

- delegittimato, perché incapace di esprimere un suo potenziale latente, essere in linea con i tempi e/o rispondere ai bisogni della società;
- colonizzato, perché ripensato secondo i principi e i vocabolari di una razionalità tecnologica e di un modello dell'innovazione meccanicistico;
- eterodiretto, perché invitato a delegare parte essenziale del design dei propri processi cardine (curricula, pedagogie e forme della valutazione) al mondo dell'EdTech.

In questo breve capitolo, utilizzando le lenti della sociologia dell'educazione, proverò a discutere in dettaglio l'importanza di

attribuire una rinnovata centralità nel dibattito pubblico a questi due aspetti, vita sociale e *politics* ontologica dell'AIED, ed evidenziare il potenziale contributo che un approccio di critica costruttiva (*constructive criticism*) può dare nel promuovere: a) un uso riflessivo dell'IA negli ambienti di apprendimento (*learning with AI*); b) favorire l'ingresso dell'IA nei curricula formativi (*learning about AI*); e soprattutto c) progettare in maniera educativamente fondata percorsi formativi che consentano agli individui di diventare cittadini consapevoli e pronti a vivere in un mondo nell'quale l'IA sarà sempre più presente in tutte le sfere cruciali delle società contemporanee (*learning for human-AI collaboration*).

## *2. La vita sociale dell'IA in Educazione*

Cosa significa prendere sul serio l'idea che l'IA in educazione abbia una vita sociale? Significa portare con forza nel dibattito pubblico e in seno alle sue controversie una sensibilità socio-storica e analitica che denaturalizzi l'IA come pura *techné* e prenda atto del suo carattere socio-tecnico<sup>7</sup>.

Ciò si traduce in primo luogo nel riconoscere la natura duale della relazione tra IA e le altre entità che costituiscono il sociale e l'educativo. Se l'AIED viene spesso acriticamente dipinta, ad esempio, come la naturale e ultima frontiera per l'innovazione tecnologica in educazione, qualsiasi accorta analisi socio-storica mostra come essa, in quanto costruzione sociale, è variamente immaginata, prodotta, inquadrata e compresa da gruppi, organizzazioni e individui che sono portatori di interessi, progetti, epistemologie e razionalità dif-

<sup>7</sup> L. Suchman, *Feminist STS*.

ferenti nelle intersezioni tra i campi della produzione di conoscenza scientifica su IA, *Learning Analytics*, *Learning Sciences* e *Education Data Mining*, dell'industria tecnologica e dell'EdTech, del *policy-making* e dell'*advocacy*<sup>8</sup>. È la caotica convergenza tra queste spinte eterogenee che spiega l'enfasi contemporanea sul potenziale dell'IA e le sue specifiche declinazioni in educazione, gli entusiasmi che la accompagnano e le relative pressioni per mettere l'uso dell'IA al centro delle agende di riforma dei sistemi educativi del futuro come leva della trasformazione digitale.

Come ampiamente sottolineato dalla ricerca nel campo della *political economy of EdTech*, ad esempio, guardare alla vita sociale dell'AIED evidenzia come uno dei motori di tali spinte siano gli interessi economici connessi all'affermazione, nell'ultimo decennio, di nuovi modelli di business *platform-* e *data-based* nell'industria EdTech. In tali modelli i profitti di medio e lungo termine nascono non tanto dalla vendita di hardware e software, quanto dall'intersezione tra fidelizzazione, *locking in* tecnologico degli utenti dentro ecosistemi chiusi, sottoscrizioni *freemium* e la raccolta di sempre maggiori quantità di dati che promettono di generare ulteriore valore attraverso la creazione di nuovi tipi di servizi *data-driven* o *knowledge-driven*<sup>9</sup>. Questa è una tendenza fortemente sostenuta anche dalle Big Tech, in quanto funzionale a strategie monopolistiche di *hyperscaling*, ossia all'introduzione delle proprie infrastrutture cloud e IA nelle routine e pratiche quotidiane delle istituzioni educative, e in particolare nelle attività pedagogiche, nella gestione dei sistemi informativi, nel monitoraggio delle presenze e delle condotte, o nella valutazione<sup>10</sup>.

<sup>8</sup> W. Holmes – I. Tuomi, *State of the art and practice in AI in education*.

<sup>9</sup> Grimaldi et al., *Platformization and the enactment*.

<sup>10</sup> B. Williamson et al., *Re-examining AI*.

In secondo luogo, un focus analitico sulla vita sociale dell'IA in Educazione implica far propria l'idea che essa, nelle sue molteplici forme, produce effetti sociali che possono essere inattesi, non immediatamente riconoscibili e problematici sul piano etico-politico. Una simile prospettiva sottolinea, cioè, come 'abbia poco senso anche solo parlare di IA in educazione come una categoria unitaria' e assumere che l'intelligenza artificiale operi in maniera omogenea nei contesti educativi come pianificato e previsto dai suoi programmatori, sviluppatori e promotori. Ugualmente *naïve* è immaginare al tempo stesso di poter gestire e risolvere eventuali problemi implementativi emergenti nei contesti situati attraverso modifiche tecniche e/o la previsione di quadri etici per il suo utilizzo, come sempre più spesso oggi si pensa di fare<sup>11</sup>.

La prospettiva che guarda alla vita social dell'AIED, invece, sollecita il dibattito pubblico a decostruire la natura unitaria ed 'oggettiva' dell'AIED (e dell'IA) in quanto artefatto tecnico e ricostruire, da un lato, gli intrecci situati tra materialità, discorsi e agency che ne rendono possibile, sul piano storico-sociale, l'emergenza nella sua pluralità e dispersione e, dall'altro, i processi relazionali di ordinamento e stabilizzazione che ne facilitano il funzionamento in condizioni di incertezza. L'invito è, cioè, quello di concentrarsi su quella che con Annamarie Mol<sup>12</sup> (1999) potremmo definire come l'*'ontological politics'* dell'AIED. Come sottolineato da Perrotta e Selwyn<sup>13</sup>, l'obiettivo è quello di portare all'evidenza del dibattito pubblico i fattori sociali, tecnici, economici e politici che fanno sì che l'AIED funzioni come tecnologia e, più in generale, «la pluralità e l'incertez-

<sup>11</sup> B. Williamson, *The social life of AI in Education*, p. 97.

<sup>12</sup> A. Mol, *Ontological politics*.

<sup>13</sup> C. Perrotta - N. Selwyn, *Deep learning goes to school*, p. 253.

za che sono alla base» dell'AIED come «fatto sociale e tecno-scientifico». Al tempo stesso, il tentativo è di influenzare l'agenda pubblica promuovendo ricerca scientifica indipendente che studi le specificità situate delle singole tecnologie, ciò che esse 'fanno' e soprattutto l'eterogeneità delle pratiche di *enactment*, ossia dei processi di interpretazione, traduzione in pratica e testing attraverso i quali esse acquisiscono materialità e diventano reali<sup>14</sup>. Infine, si tratta di provare a mettere a fuoco le differenti implicazioni sociali ed educative di tale eterogeneità, incluse le conseguenze inattese e non volute, considerando l'intreccio tra la dimensione etica, regolativa e del controllo<sup>15</sup>. Questo ultimo aspetto va ancora una volta problematizzato a partire dalla prospettiva della vita sociale dell'AIED, nella misura in cui gli stessi framework etici, legali e regolativi sono socialmente costruiti e agiti, essendo frutto ed espressione del conflitto tra interessi e istanze divergenti, ma anche dei diversi rapporti di forza esistenti nel campo regolativo, come mostrato dai casi sempre più diffusi di framework di auto-disciplinamento etico sull'uso dell'AIED che costituiscono casi di '*ethics washing*' fortemente pilotati dai produttori EdTech e che paradossalmente finiscono per delegare la responsabilità dell'uso etico delle soluzioni AIED *data- e knowledge-driven* alle scuole e agli educatori<sup>16</sup>.

In sintesi, per riprendere le parole di Perrotta e Selwyn<sup>17</sup> un «modo generativo per interrogarsi [...] sull'intelligenza artificiale in educazione è guardare intorno, piuttosto che all'interno, di scatole nere sempre più opache e difficili da comprendere». Questo significa,

<sup>14</sup> Per una teoria dell'*enactment* si veda S. J. Ball - M. Maguire - A. Braun, *How schools do policy*.

<sup>15</sup> R. Eynon - E. Young, *Methodology, Legend, and Rhetoric*.

<sup>16</sup> G. van Maanen, *AI ethics*.

<sup>17</sup> C. Perrotta - N. Selwyn, *Deep learning goes to school*, p. 256.

secondo i due autori, spostare l'attenzione verso l'analisi delle "forme di vita" dell'AIED, e cioè verso le relazioni controverse e incerte tra i significati culturali iscritti nelle sue materializzazioni digitali e non, gli interessi economici dell'industria tecnologica, le priorità e strategie di policy e gli intrecci disciplinari e professionali delle scienze dei dati.

### 3. *La politics (ontologica) dell'IA in Educazione*

Riconoscere che l'AIED, nelle sue molteplici manifestazioni, sia il risultato di un processo di costruzione sociale significa considerarla come una tecnologia politica, che incorpora epistemologie, razionalità politiche/educative e visioni ideologiche, e che in quanto tale può essere utilizzata per perseguire particolari obiettivi culturali e di policy<sup>18</sup>. L'AIED in quanto sistema socio-tecnico è implicata in forme di produzione di conoscenza che reificano e naturalizzano particolari tipi di epistemologie educative, marginalizzandone altre. Nel campo dell'educazione, tale riconoscimento assume una rilevanza particolare, soprattutto alla luce della sostanziale assenza di un dibattito pubblico sulle iscrizioni educative e pedagogiche di cui sono portatrici le diverse materializzazioni dell'AIED oggi disponibili sul mercato.

Knox et al.<sup>19</sup> hanno, ad esempio, ampiamente dimostrato come la maggioranza delle soluzioni tecnologiche di AIED oggi sul mercato (e gli immaginari sociali che le accompagnano) combinino l'uso di forme di IA *data-* e *knowledge-driven* con teorie dell'ap-

<sup>18</sup> B. Williamson, *The social life of AI in Education*, p. 101.

<sup>19</sup> J. Knox - B. Williamson - S. Bayne, *Machine behaviourism*.

prendimento di matrice comportamentista (spesso radicale), con l'obiettivo di indirizzare gli studenti verso script cognitivo-comportamentali ed obiettivi di apprendimento predefiniti. In questo senso, non sembra peregrino osservare come tale combinazione, al di là delle retoriche sulla personalizzazione dell'apprendimento, sembri andare in direzione contraria rispetto alla promozione dell'autonomia e della partecipazione degli studenti<sup>20</sup>. Soprattutto, essa sembra materializzare il sogno (tecnocratico e ideologicamente orientato) di promuovere forme di (auto)governo dell'educazione basate su dati e processi di standardizzazione degli apprendimenti. Sarebbe, a mio avviso, essenziale mettere al centro del dibattito sull'AIED la fondatezza, completezza e desiderabilità di un tale progetto tecno-educativo, alla luce ad esempio della consolidata consapevolezza che l'acquisizione di conoscenze, abilità e competenze predefinite e rivolte all'azione o *qualificazione* è solo una delle funzioni dell'educazione, e probabilmente non la più importante<sup>21</sup>. Mi chiedo ad esempio cosa significherebbe valutare le soluzioni tecnologiche di AIED oggi disponibili sul mercato e le modalità con le quali esse vengono introdotte negli ambienti di apprendimento alla luce della sottolineatura del fatto che un'attenzione relativamente scarsa da parte dei *designer*, degli sviluppatori e dei promotori dell'AIED viene data ad altre due funzioni chiave dell'educazione: a) la *socializzazione*, ossia la promozione di quella molteplicità di modi in cui, attraverso l'educazione, diventiamo parte di particolari 'ordini sociali, culturali e politici'; b) la *soggettivazione*, ossia il processo che consente a coloro che hanno ricevuto un'istruzione di diventare più autonomi e indipendenti nel pensare e

<sup>20</sup> W. Holmes - A. Littlejohn, *Artificial intelligence for professional learning*.

<sup>21</sup> G. J. Biesta, *The Beautiful Risk of Education*.

nell'agire<sup>22</sup>. Significherebbe decretare, forse, il carattere fortemente riduzionistico del progetto di personalizzazione educativa che essa promuove nelle sue forme contemporanee, unicamente incentrato su di una lettura Bloomiana dell'educazione come *mastery of learning* individuale e sull'utilizzo della differenziazione individuale dei percorsi come leva per promuovere livelli standard di acquisizione di conoscenze, abilità e competenze.

Questo tipo di considerazioni acquisiscono un ulteriore livello di complessità sul piano della *politics* dell'AIED e delle tensioni/conflicti che la attraversano se prendiamo atto, con Williamson<sup>23</sup>, che le sue direzioni sul piano dello sviluppo tecnologico, del design e dell'uso nei contesti educativi sono sempre più influenzate da attività di ricerca, sviluppo e test promossi dal mondo del Big Tech e dell'EdTech, la cui agenda ricalca l'interpretazione comportamentista/funzionalista sopra richiamata, a fronte di una relativa marginalizzazione delle direzioni di ricerca sull'AIED disallineate rispetto a tale progetto tecno-educativo che pur vengono condotte nei centri di ricerca universitari.

Sottolineare la rilevanza del tema della *politics* dell'AIED ha però un altro e altrettanto rilevante corollario analitico. Se è vero che l'AIED in quanto tecnologia politica incorpora epistemologie, razionalità politiche/educative e visioni ideologiche che hanno un carattere performativo rispetto alle realtà dell'insegnamento e dell'apprendimento, al tempo stesso va riconosciuto, con Mol<sup>24</sup>, che ogni singola soluzione di AIED, in quanto artefatto socio-tecnico, offre diverse versioni di sé e apre diverse possibilità per la propria *per-*

<sup>22</sup> *Ibid.*

<sup>23</sup> B. Williamson, *The social life of AI in Education*, p. 100.

<sup>24</sup> A. Mol, *Ontological politics*, p. 75.

*formance*. Il modo in cui AIED e gli altri elementi eterogenei degli assemblaggi educativi ‘*actually come together*’ è fondamentalmente un processo aperto a forme alternative e potenzialmente infinite di modellamento sociale<sup>25</sup>. Questa molteplicità ontologica dell’AIED in quanto artefatto apre il campo a diversi interrogativi di ricerca, che rimandano alle condizioni di possibilità per la scelta tra diverse opzioni, agli interessi in gioco e alle decisioni prese nei contesti situati del loro *enactment*, alle forme di esclusività, dipendenza, conflitto e controllo che influenzano tali scelte. L’invito è dunque quello di portare al centro del dibattito la necessità di esplorare l’*ontological politics* dell’AIED nei contesti educativi, a partire dall’assunto che le condizioni di possibilità per la sua traduzione in pratica non sono date, che la realtà della tecnologia non precede le pratiche educative e che il suo carattere è sempre aperto e oggetto di scontro sul piano della sua *politics*. Se, infatti, l’AIED ha una *vita sociale* per quel che riguarda la sua produzione, promozione e finalizzazione, continua ad averla anche quando entra nelle pratiche educative, all’interno di contesti sociali, culturali e politici altamente differenziati. Fare ricerca e sottoporre a discussione pubblica evidenze provenienti dai diversi contesti di *enactment* dell’AIED è in questo senso fondamentale sul piano politico-culturale nella misura in cui se l’AIED non va considerata come una tecnologia monolitica e non viene messa in azione (interpretata, tradotta in pratica e testata) negli stessi modi in setting differenti e da gruppi diversificati, diviene logicamente evidente che gli usi dell’AIED risulteranno molto variegati se comparati con le promesse visionarie in merito al loro potenziale trasformativo proposteci dai suoi promotori. Al tempo stesso diventa necessario comprendere in profondità la relazione tra tecnologie AIED, *politics*

<sup>25</sup> C. Perrotta - N. Selwyn, *Deep learning goes to school*, p. 267.

tecno-educative e l'insieme eterogeneo di elementi ed entità che definiscono le condizioni di possibilità per il loro *enactment* nei contesti situati (ad esempio, culture educative e professionali, caratteristiche socio-economiche e demografiche dei contesti e delle platee di ricezione, altre forme di infrastrutturazione tecnologica, discorsi e processi di agenda setting politica, culturale ed educativa, solo per citare alcune dimensioni rilevanti). Il compito della ricerca educativa, a mio avviso, è quello di portare all'attenzione del dibattito pubblico questi aspetti e promuovere, soprattutto tra i professionisti dell'educazione, maggiore riflessività rispetto alle modalità con le quali l'AIED viene 'messa in azione' nei contesti scolastici e nelle pratiche di apprendimento extra-scolastiche, evidenziando con forza nel dibattito pubblico come l'introduzione dell'IA e dell'automazione in educazione costituiscano a tutti gli effetti una questione politica a cui guardare da due punti di vista: quello del potere esercitato da parte dei produttori delle tecnologie, ma anche quello delle capacità trasformative insite nell'agency di coloro che ne fanno esperienza<sup>26</sup>.

È questo un passaggio cruciale di cui tenere conto se si vogliono evitare cadute di ritorno in forme più o meno velate di determinismo tecnologico, che spesso informano implicitamente le visioni apocalittiche o distopiche sui futuri educativi colonizzati dall'IA, oppure ricadere in quanto progettisti dell'innovazione educativa in forme di *hybris* volte a predeterminare le forme del reale educativo, o ancora giungere a facili spiegazioni dei fallimenti implementativi che assegnino la colpa di un uso povero della tecnologia e di risultati non all'altezza delle aspettative a scuole, insegnanti e/o alunni.

<sup>26</sup> B. Williamson - F. Macgilchrist - J. Potter, *Re-examining AI*.

## *Conclusioni*

A fronte delle spinte riduzionistiche che caratterizzano il dibattito pubblico e gli immaginari promossi dai produttori del Tech<sup>27</sup>, viene da chiedersi quale debba essere il ruolo per una ricerca scientifica, soprattutto quella proveniente dal campo delle Scienze Umane e Sociali, che abbia a cuore il progetto culturale di promuovere un uso riflessivo dell'IA negli ambienti di apprendimento, l'ingresso dell'IA nei curricula formativi come oggetto di studio e formazione e, infine, la formazione di individui che in futuro saranno in grado di immaginare e agire come virtuose di *human-AI collaboration*. Nella sezione finale di questo contributo provo a sostenere come un'attenzione sociologicamente fondata alla *vita sociale dell'IA in educazione* e alla *politics ontologica dell'AIED* contribuisca a tale progetto suggerendo di portare in primo piano nell'agenda pubblica, politica ed educativa alcuni temi e domande chiave<sup>28</sup>.

## AIED e Etica

È forse opportuno ristabilire una forma di simmetria epistemica tra tecnologico ed educativo quando si discute di uso etico dell'AIED. Questo significa accompagnare alla necessaria ma non sufficiente attenzione verso le questioni riguardanti i dati, le *learning analytics*, la loro proprietà/gestione, la trasparenza, affidabilità, la fiducia e i *bias* legati alla privacy, la messa al centro del problema dell'etica dell'edu-

<sup>27</sup> M. Forsman - I. Forsler - S. Opermann - F. Bardone - M. Pedaste, *Future classrooms*.

<sup>28</sup> W. Holmes - I. Tuomi, *State of the art and practice in AI in education*.

cazione. Per far ciò è importante problematizzare sul piano dell'etica dell'educazione le iscrizioni curriculari, pedagogiche e valutative delle diverse forme di AIED, chiedendosi quali sono gli obiettivi generali, le forme della conoscenza e dell'apprendimento, le modalità per valutare l'esperienza educativa, e le forme dell'agency di insegnanti e studenti che caratterizzano l'educazione verso cui vogliamo tendere, e come e in quale misura l'AIED a noi contemporanea ci aiuti a tendere verso di essa o ne rappresenti un allontanamento.

### AIED e Personalizzazione

Una delle forme di riduzionismo pedagogico più eclatanti nel discorso pubblico sull'AIED riguarda la personalizzazione dell'apprendimento che essa renderebbe possibile. Per riprendere il pensiero di Gert Biesta<sup>29</sup>, la personalizzazione nel suo senso più pieno riguarda i processi di soggettivazione, una delle tre direttrici fondamentali dell'educazione, intesa come processo che consente all'individuo di auto-realizzarsi o di ampliare gli orizzonti della propria libertà in quanto soggetto di auto-determinazione. Sarebbe cruciale aprire una riflessione ampia e pluridisciplinare su quanto le principali soluzioni AIED presenti oggi nel campo dell'educazione lavorino in questa direzione o in quella contraria, fornendo nei fatti ai discenti percorsi adattivi più o meno standardizzati verso l'apprensione di contenuti che finiscono per produrre omogeneizzazione piuttosto che personalizzazione e rendere attuale un'educazione il cui obiettivo è quello di far sì che ogni studente *'fit in the right box'*<sup>30</sup>. In relazione al tema del-

<sup>29</sup> G. J. Biesta, *The Beautiful Risk of Education*.

<sup>30</sup> W. Holmes - I. Tuomi, *State of the art and practice in AI in education*, p. 559.

la personalizzazione, ci sono altri tre assunti, chiaramente infondati, che vale la pena sfidare nell'alveo del dibattito pubblico sull'AIED: a) l'idea che la personalizzazione sia resa possibile esclusivamente dalla tecnologia digitale; b) l'idea che si possa efficacemente personalizzare attraverso percorsi formativi individuali designati a partire da meccanismi conoscitivi e previsionali che sono sostanzialmente basati sull'uso di dati prodotti da altri discenti e su medie; c) l'idea che si possa realizzare una piena personalizzazione dell'esperienza educativa, intesa come soggettivazione, attraverso una tecnologia che ignora, *by design*, la dimensione collaborativa e dell'interazione sociale che è costitutiva di qualsiasi esperienza educativa.

### Effetti ed implicazioni dell'AIED

Come ho più volte sottolineato nel testo, richiamando la letteratura, molte delle affermazioni relative al potenziale rivoluzionario dell'AIED sono basate su congetture, speculazioni e letture molto ottimistiche di basi dati poco rigorose e/o generalizzabili. Tra l'altro, le poche evidenze empiriche a disposizione sono essenzialmente concentrate sull'individuazione di correlazioni tra l'uso di soluzioni AIED e il miglioramento delle performance accademiche degli studenti. Abbiamo bisogno di vasti programmi pubblici di ricerca indipendente che prendano in considerazione la complessità dei fenomeni connessi alla vita sociale dell'AIED e alla sua politica ontologica, e che mettano al centro delle proprie riflessioni interrogativi relativi alle più ampie implicazioni dell'uso dell'IA nei setting educativi, il suo impatto su insegnanti e studenti e, infine, sul perché e per quali fini abbiamo veramente bisogno dell'IA in educazione.

## AIED come forma di Colonialismo digitale

Un aspetto poco indagato e discusso dell'AIED riguarda la misura in cui, sul piano della *politics* e *political economy*, le soluzioni oggi esistenti e il loro *embedding* nelle pratiche dei sistemi educativi dei paesi del *Global South* non rappresentino una forma invisibile di nuovo colonialismo digitale<sup>31</sup>, destinate a rinforzare le posizioni egemoniche e monopolistiche delle grandi corporations del *Global North* sul piano infrastrutturale, estendere i processi di estrazione di valore in termini di dati e capitali e imporre su scala globale una specifica *politics* dell'educazione nascondendola dietro una forma apparentemente neutra ed universale di razionalità tecnologica.

## AIED e commercializzazione

La commercializzazione dell'AIED comporta alcuni rischi che non hanno solo e unicamente a che fare con gli interessi delle aziende EdTech nel generare profitti a partire dalla vendita di beni e servizi basati sull'IA e dalla sostanziale assenza o marginalità sul mercato di prodotti e servizi di proprietà pubblica e/o ispirati ai principi del software libero e dell'open source<sup>32</sup>. Come osservano Holmes & Tuomi<sup>33</sup>, emerge in questo caso con forza il tema della responsabilità dei soggetti commerciali. Diviene ad esempio essenziale sollevare nel dibattito pubblico interrogativi circa il fatto che le interazioni tra studenti e sistemi di IA generano sia conoscenza tecnica in relazione

<sup>31</sup> J. C. Young, *The new knowledge politics*.

<sup>32</sup> E. Grimaldi - P. Landri - D. Taglietti, *Una sociologia pubblica del digitale*.

<sup>33</sup> W. Holmes - I. Tuomi, *State of the art and practice in AI in education*, p. 562.

al prodotto e al suo funzionamento che conoscenza di mercato in merito a come il prodotto viene usato e se questo non si configuri come una forma di arruolamento occulto degli studenti come produttori di *business intelligence*. E questo interrogativo apre ovviamente la strada a un ulteriore dubbio dal rilevante peso etico: e cioè se questa esigenza strategica non venga resa prioritaria dalle *corporation* dell'EdTech attraverso il *design* stesso delle soluzioni tecnologiche rispetto agli interessi educativi degli studenti stessi. Sembra in tal senso necessario spingere perché si creino le condizioni affinché i produttori di EdTech condividano le informazioni circa i loro sistemi proprietari, la loro efficacia, il funzionamento dei processi educativi che si sviluppano attraverso le soluzioni tecnologiche di cui si fanno promotori, al fine di rendere autenticamente pubblico il controllo sociale sull'AIED e sull'innovazione educativa che attraverso di essa si sviluppa.

In sintesi, nella misura in cui il panorama tecnologico dell'IA intorno a noi cambia rapidamente, il contributo che la ricerca scientifica indipendente è chiamata a dare, nell'indirizzare il dibattito pubblico sull'AIED, è quello di sfidare qualsiasi forma di soluzionismo tecnologico, ricentrare il dibattito su quale sia l'educazione che riteniamo desiderabile e promuovere una riflessione multidisciplinare sul ruolo che l'educazione ha in questo mondo in cambiamento. Per riprendere le parole di Holmes & Tuomi<sup>34</sup>, «si ritiene spesso che l'innovazione tecnica e il cambiamento significhino progresso ma, nel campo dell'educazione, [è forse il momento di] chiedersi in che misura il cambiamento a cui stiamo assistendo equivale ad uno sviluppo desiderabile». Questo significa non tanto quello di opporsi all'utilizzo dell'AIED nei *milieu* educativi (impresa peraltro di diffi-

<sup>34</sup> Ivi, p. 563.

cile attuazione), quanto piuttosto accettare la sfida di un'educazione con, su e per l'IA etica adottando un'attitudine critica, sperimentale e agonistica<sup>35</sup>. Se l'obiettivo di qualsiasi forma di critica costruttiva è quello di denaturalizzare le forme e gli usi attuali dell'AIED e al tempo stesso mostrare che tali forme e usi potrebbero essere disegnate diversamente, un futuro più fecondo per la progettazione dell'AIED potrebbe nascere da contaminazioni più profonde tra saperi, expertise e know-how tecnici e di design di cui sono portatori progettisti e sviluppatori, e quelli di cui sono detentori quegli educatori e quelle voci indipendenti e critiche del mondo della ricerca scientifica che si interrogano da diversi punti di vista disciplinari sulle iscrizioni educative di cui le tecnologie sono portatrici e sulla complessità degli effetti sociali, attesi e non, delle tecnologie nei *milieu* educativi. Come recentemente auspicato da Neil Selwyn<sup>36</sup>, se siamo effettivamente alle soglie dell'avvento di un'età dell'AIED, la speranza è che questa prenda forma a partire da una sofisticata consapevolezza politica che sia l'esito di un processo di co-produzione che coinvolge in chiave simmetrica fautori, promotori e critici costruttivi di tale tecnologia.

<sup>35</sup> C. Perrotta - N. Selwyn, *Deep learning goes to school*, p. 267.

<sup>36</sup> N. Selwyn, *Constructive Criticism?*

## Bibliografia

- S. J. Ball - M. Maguire - A. Braun, *How schools do policy: Policy enactments in secondary schools*, London, Routledge, 2012.
- G. J. Biesta, *The Beautiful Risk of Education*, London, Routledge, 2015.
- R. Eynon - E. Young, *Methodology, Legend, and Rhetoric: The Constructions of AI by Academia, Industry, and policy groups for lifelong learning*, in «Science, Technology & Human Values», 46, 1, 2021, pp. 166-191.
- M. Forsman - I. Forsler - S. Opermann - F. Bardone - M. Pedaste, *Future classrooms and ed-tech imaginaries. Notes from the Estonian pavilion at EXPO 2020 and beyond*, in «Learning, Media and Technology», 49, 1, 2024, pp. 133-146.
- E. Grimaldi - P. Landri - D. Taglietti, *Una sociologia pubblica del digitale a scuola*, in «Scuola democratica», 2020, Early Access: <https://www.rivisteweb.it/doi/10.12828/97096>
- E. Grimaldi - S. J. Ball - F. Peruzzo, *Platformization and the enactment of multiple economic forms: The rise of a neoliberal moral economy of education*, in *The new digital education policy landscape*, a cura di A. Rivas - C. Cobo, New York, Routledge, 2023, pp. 122-146.
- W. Holmes - A. Littlejohn, *Artificial intelligence for professional learning*, in *Handbook of Artificial Intelligence at Work*, a cura di M. Garcia-Murillo - I. MacIness - A. Renda, London, Edward Elgar Publishing, 2024, pp. 191-211.
- W. Holmes - I. Tuomi, *State of the art and practice in AI in education*, in «European Journal of Education», 57, 2022, pp. 542-570.
- R. Kitchin, *Digital timescapes: Technology, temporality and society*, Cambridge, Polity Press, 2023.
- J. Knox - B. Williamson - S. Bayne, *Machine behaviourism: Future visions of 'learnification' and 'datafication' across humans and digital technologies*, in «Learning, Media and Technology», 45, 1, 2020, pp. 31-45.
- A. Mol, *Ontological politics. A word and some questions*, in «The Sociological Review», 47, 1\_suppl, 1999, pp. 74-89.
- S. Nemorin - A. Vlachidis - H. M. Ayerakwa - P. Andriotis, *AI hyped? A horizon scan of discourse on artificial intelligence in education (AIED) and development*, in «Learning, Media and Technology», 48, 1, 2023, pp. 38-51.
- C. Perrotta - N. Selwyn, *Deep learning goes to school: Toward a relational understanding of AI in education*, in «Learning, Media and Technology», 45, 3, 2020, pp. 251-269.
- N. Selwyn, *Constructive Criticism? Working with (Rather than Against) the AIED Back-Lash*, in «International Journal of Artificial Intelligence in Education», 34, 2024, pp. 84-91.

- L. Suchman, *Feminist STS and the Sciences of the Artificial*, in *The Handbook of Science and Technology Studies, Third Edition*, a cura di E. J. Hackett - O. Amsterdamska - M. E. Lynch - J. Wajcman, MA, MIT Press, 2008, pp. 139-164.
- The social life of things: Commodities in cultural perspective*, a cura di A. Appadurai, Cambridge, Cambridge University Press, 1988.
- G. van Maanen, *AI ethics, ethics washing, and the need to politicize data ethics*, in «Digital Society», 1, 2, 2022, pp. 9-18.
- B. Williamson, *The social life of AI in education*, in «International Journal of Artificial Intelligence in Education», 34, 1, 2024, pp. 97-104.
- B. Williamson - F. Macgilchrist - J. Potter, *Re-examining AI, automation and datafication in education*, «Learning, Media and Technology», 48, 1, 2023, pp. 1-5.
- J. C. Young, *The new knowledge politics of digital colonialism*, in «Environment and Planning A: Economy and Space», 51, 7, 2019, pp. 1424-1441.

Emiliano Grimaldi  
Università degli Studi di Napoli Federico II  
emiliano.grimaldi@unina.it  
<https://orcid.org/0000-0003-1085-1127>



LUCA LO SAPIO

*Creatività umana e creatività computazionale:  
una visione pluralistica dell'atto creativo\**

*Abstract:* This contribution explores the relationship between human creativity and computational creativity, proposing a pluralistic view of the notion of creativity that goes beyond traditional anthropocentric conceptions. After tracing the evolution of artificial intelligence from its historical limitations to its recent generative capabilities, I argue that even generative AI systems, such as ChatGPT and Midjourney, can be considered creative agents, provided that we evaluate their output rather than their consciousness or intentionality. The text puts forward a functional definition of creativity, based on the criteria of novelty, surprise, and value (Boden), and examines its philosophical, cognitive, and aesthetic implications. The rigid distinction between “strong” and “weak computational creativity” is rejected, and a model of human-machine co-creativity is advocated, fostered by the interactivity, adaptability, and autonomy of new systems. In conclusion, a pluralistic account of creativity is proposed; one that acknowledges the contribution of generative AIs as syntagmatic actors within the contemporary symbolic landscape.

*Keywords:* Reductionist Account of Creativity, Exclusivist Account of Creativity, Strong Computational Creativity, Weak Computational Creativity, Generative Artificial Intelligence.

*Le grandi cose non vengono fatte  
d'impulso ma mettendo insieme tante  
piccole cose.*

V. Van Gogh

\* Dedico questo contributo alla memoria di mia mamma che mi ha lasciato senza che avessi il tempo di dirle quanto i suoi insegnamenti, le sue parole, il suo esempio siano stati per me linfa vitale, senza cui sarei oggi un'altra persona.

1. *Dal ripiegamento strategico alla creatività computazionale*

Nel primo capitolo de *Il duplice enigma. Intelligenza artificiale e intelligenza umana*, Daniel Andler osserva come la storia dell'intelligenza artificiale sia stata contrassegnata da una serie di "rivalutazioni al ribasso", che hanno progressivamente ridimensionato sia le ambizioni epistemiche della disciplina sia la fiducia nelle sue reali potenzialità<sup>1</sup>. In particolare, nel paragrafo intitolato *Il ripiegamento strategico*, Andler rievoca un celebre episodio riportato da Herbert Simon, uno dei fondatori dell'IA, che raccontava ai propri studenti di avere, insieme ad Allen Newell, inventato durante le vacanze di Natale «un programma di computer capace di pensare in termini non numerici», affermando così di aver risolto «il venerabile problema corpo-mente», dimostrando che «un sistema composto di materia possa avere le proprietà della mente»<sup>2</sup>.

Tuttavia, l'entusiasmo che caratterizzò le prime fasi della ricerca sull'IA cedette presto il passo a una più sobria valutazione dei suoi limiti teorici e applicativi. Andler identifica questo processo di arretramento concettuale come un ripiegamento strategico articolato in cinque tappe: la rinuncia al pensiero pensante, l'abbandono del principio di equivalenza tra intelligenza artificiale e intelligenza umana, la rinuncia a un'intelligenza generale, il riconoscimento della dipendenza strutturale dell'IA dalla cognizione umana e, infine, la rinuncia a qualsiasi forma di riflessività autonoma<sup>3</sup>.

<sup>1</sup> D. Andler, *Intelligence artificielle, intelligence humaine: la double énigme*; trad. it. *Il duplice enigma. Intelligenza artificiale e intelligenza umana*, pp. 18-21.

<sup>2</sup> *Ibid.*

<sup>3</sup> Ivi, pp. 21-30. Una prospettiva simile, in cui si evidenzia che la Rivoluzione dell'IA non consisterebbe nella creazione di *macchine pensanti* ma, piuttosto, nella capacità di dissociare *agency* e *intelligenza*, al punto da consentire a delle macchine di eseguire task specifici che, *diversamente*, attribuiremmo a enti dotati di intelligenza, è sostenuta da Luciano Floridi in vari paper e volumi (ad esempio in

Questo ripiegamento non va letto come un segno di fallimento, ma come una ricalibratura delle aspettative teoriche nei confronti di un campo che, sin dall'inizio, si era mosso al crocevia tra informatica, scienze cognitive, epistemologia e filosofia della mente. Tuttavia, a partire dalla fine degli anni '90, alcuni eventi hanno progressivamente rimesso in discussione l'idea che l'intelligenza artificiale dovesse rimanere confinata a una forma debole o subordinata<sup>4</sup>.

Un primo spartiacque è rappresentato dalla celebre vittoria del computer progettato da IBM Deep Blue sul campione mondiale di scacchi Garry Kasparov nel 1997. All'epoca, il successo della macchina fu attribuito prevalentemente alla sua superiore potenza computazionale, senza che ciò implicasse un'autentica capacità di ragionamento o creatività. L'IA veniva ancora percepita come un'estensione quantitativa della logica combinatoria, non come un agente cognitivo dotato di insight.

Questo paradigma ha iniziato a incrinarsi definitivamente con l'avvento di AlphaGo, il sistema sviluppato da DeepMind capace di sconfiggere alcuni dei migliori giocatori mondiali di Go, un gioco notoriamente resistente alla semplice forza computazionale. Le mosse di AlphaGo – descritte da molti come “non umane”, “sorprendenti” e “controintuitive” – hanno portato a riconsiderare l'idea che l'intelligenza artificiale non potesse mai generare strategie creative<sup>5</sup>.

L. Floridi, *Etica dell'intelligenza artificiale* e Id., *AI as Agency without intelligence: on Chat GPT, Large Language models, and other generative models*, pp. 1-12).

<sup>4</sup> È stato John Searle a introdurre la distinzione tra *weak* e *strong AI*, per indicare i sistemi di intelligenza artificiale capaci *solo* di riprodurre alcune funzioni del cervello umano, senza essere tuttavia in grado di implementare *reali processi coscienti* e sistemi di intelligenza artificiale capaci di produrre *reali processi coscienti* in forma *human-like* (cfr. J. R. Searle, *Minds, Brains, and Programs*).

<sup>5</sup> Per una breve storia dell'IA e delle tappe che l'hanno caratterizzata cfr. S. Russell - P. Norvig, *Artificial Intelligence: a modern approach*.

Da quel momento, una nuova stagione si è aperta: i progressi nel Machine Learning e, più specificamente, nel Deep Learning, hanno prodotto risultati sempre più impressionanti, in ambiti che vanno dal riconoscimento linguistico alla generazione automatica di immagini, testi e musica. Gli attuali sistemi di intelligenza artificiale generativa – come Chat GPT o Midjourney – non costituiscono un’eccezione improvvisa, ma rappresentano la punta visibile di un’evoluzione lunga almeno due decenni. Nonostante le perplessità espresse da alcuni studiosi<sup>6</sup>, tali sistemi sembrano indicare che almeno alcune funzioni cognitive complesse – un tempo ritenute prerogativa esclusiva dell’essere umano – possano essere emulate in modo sorprendentemente efficace. Tra queste, la capacità di generare contenuti nuovi, strutturalmente coerenti e culturalmente pertinenti, è forse la più rilevante, poiché tocca direttamente ciò che tradizionalmente viene indicato come il “codice umano”<sup>7</sup>: la creatività.

Le IA generative sfidano dunque una visione ancora largamente diffusa, secondo cui la creatività richiederebbe necessariamente l’intervento di un soggetto intenzionale, cosciente e dotato di interiorità. Sebbene questi sistemi non possiedano al momento alcuna forma di esperienza fenomenica né un punto di vista in prima persona, le loro prestazioni sollevano interrogativi significativi circa la definizione operativa di creatività e i criteri in base ai quali attribuiamo tale qualifica a un prodotto.

<sup>6</sup> Cfr. D. Chalmers, *Could a Large Language Model be Conscious?*; P. Butlin - R. Long, *Consciousness in Artificial Intelligence: Insights from the Science of Consciousness*.

<sup>7</sup> L’espressione è utilizzata ad esempio da M. Du Sautoy, *The creativity code: Art and innovation in the age of AI*. In questo contributo saranno, in ogni caso, considerati soprattutto esempi di creatività relativi alla produzione di opere d’arte e testi scritti.

Ciò che è oggi in discussione non è solo la capacità della macchina di imitare tratti della produzione umana, ma la possibilità che alcuni artefatti generati da agenti artificiali – se valutati indipendentemente dal tipo di processo sottostante – soddisfino i criteri di novità, sorpresa e valore che giustificano il loro riconoscimento come oggetti creativi. Questo spostamento dell'attenzione dall'intenzionalità dell'agente alle proprietà dell'output consente di riformulare il concetto stesso di creatività in chiave più funzionale e pluralista. In quest'ottica, l'intelligenza artificiale non si limita a supportare la creatività umana, ma può essere considerata un partner generativo in grado di contribuire attivamente alla produzione di senso.

Elaborare una definizione univoca di creatività si rivela estremamente problematico, se non del tutto illusorio. Il concetto stesso è intrinsecamente polisemico e attraversa una pluralità di domini disciplinari – dalla psicologia cognitiva alla sociologia, dalle neuroscienze all'antropologia culturale – ciascuno dei quali ne propone declinazioni specifiche, ancorate a differenti presupposti epistemologici e metodologici<sup>8</sup>. Di conseguenza, ogni tentativo di circoscrivere il significato di “creatività” si scontra con la complessità delle sue manifestazioni e con la difficoltà di ricondurle a un'unica teoria esplicativa.

Pur nella pluralità dei modelli, propongo di utilizzare uno schema interpretativo che offre una efficace piattaforma euristica per inquadrare la questione. Da un lato, troviamo le *teorie riduzioniste*, secondo cui la creatività può essere spiegata integralmente in termini di proprietà di primo livello: si tratta di modelli che individuano

<sup>8</sup> Tra gli altri, suggerisco di consultare, E. De Bono, *Lateral thinking. A textbook of creativity*; H. Gardner, *Frames of Mind: The Theory of Multiple Intelligences*; G. Lolli, *La creatività in matematica*.

nella dinamica neurofisiologica dei circuiti cerebrali, nei meccanismi evolutivi o nelle strutture sociali e culturali i fattori determinanti per la genesi del comportamento creativo. In questa prospettiva, la creatività non è altro che l'emergere di configurazioni nuove a partire dall'interazione tra un sistema nervoso plastico e un ambiente culturalmente codificato<sup>9</sup>.

Dall'altro lato, si collocano le *teorie escludiviste*, che attribuiscono alla creatività uno statuto parzialmente irriducibile ai processi materiali sottostanti. Secondo questa visione, la creatività sarebbe un fenomeno emergente che, pur potendo essere influenzato da determinanti biologiche e sociali, non si lascia esaurire nella loro descrizione funzionale. Le concezioni escludiviste tendono a conservare un alone di eccezionalismo intorno alla creatività umana, spesso associata all'idea di "genio", "ispirazione" o "intenzionalità cosciente", rendendo difficile il suo trasferimento a soggetti non umani o artificiali<sup>10</sup>.

Qui assumerò il primo di questi due paradigmi come base operativa, nella convinzione che solo una comprensione naturalizzata e pluralista della creatività permetta di affrontare in modo non dogmatico e scientificamente rigoroso la questione della possibile creatività delle intelligenze artificiali generative.

In particolare, se assumiamo una definizione funzionale, come quella proposta da Margaret Boden<sup>11</sup> – secondo cui un prodotto

<sup>9</sup> Cfr. A. Dietrich, *The cognitive neuroscience of creativity*, pp. 1011-1026; R. E. Jung - O. Vartanian, *The Cambridge Handbook of the Neuroscience of Creativity*; D. K. Simonton, *Creativity and Discovery as Blind Variation: Campbell's (1960) BVSR Model after the Half-Century Mark*, pp. 158-174.

<sup>10</sup> Cfr. B. Gaut, *The Philosophy of Creativity*, pp. 1034-1046; A. J. Cropley, *In Praise of Convergent Thinking*, pp. 391-404.

<sup>11</sup> M. Boden, *Creativity and artificial intelligence*, pp. 347-356.

può essere considerato creativo se è nuovo, sorprendente e dotato di valore – allora non vi è motivo di escludere a priori che anche i risultati generati da agenti artificiali possano soddisfare tali criteri. Naturalmente, ciò implica una revisione degli approcci tradizionali alla creatività, ancora vincolati a concezioni antropocentriche o al presupposto della centralità dell'intenzionalità cosciente.

Per argomentare questa tesi, proporrò una doppia mossa teorica. In primo luogo, presenterò una versione naturalizzata della creatività umana, che consenta di renderla comprensibile entro una cornice esplicativa biologico-cognitiva. In questa prospettiva, i processi creativi possono essere ricondotti all'attività della microcircuiteria neurosinaptica dell'individuo e alle interazioni con l'ambiente culturale e simbolico in cui è situato. La creatività, in tal senso, non è da intendersi come un fenomeno misterioso o auratico, ma come il prodotto complesso e dinamico di sistemi neurobiologici plastici in relazione con specifici contesti storici e culturali. In questa direzione si muovono, tra gli altri, l'ipotesi dell'epigenesi per stabilizzazione selettiva delle sinapsi formulata da Jean-Pierre Changeux<sup>12</sup> e il darwinismo neurale proposto da Gerald Edelman<sup>13</sup> e ulteriormente sviluppato da Giulio Tononi<sup>14</sup>.

In secondo luogo, proporrò una riformulazione pluralistica della nozione di creatività, che consenta di superare l'idea di una sua definizione univoca e rigidamente antropocentrica. Adottando una prospettiva più flessibile e comparativa, sarà possibile includere nel campo del creativo anche agenti artificiali capaci di generare conte-

<sup>12</sup> J. P. Changeux, *L'Homme neuronal*.

<sup>13</sup> G. Edelman, *Second nature: brain, science and human knowledge*.

<sup>14</sup> G. Edelman - G. Tononi, *A universe of consciousness: How matter becomes imagination*.

nuti nuovi e culturalmente rilevanti, pur in assenza di una coscienza fenomenica, cioè di un punto di vista in prima persona. In questa cornice, diventa plausibile parlare non solo di creatività computazionale, ma anche di co-creatività uomo-macchina: gli agenti artificiali non sono più meri strumenti, ma partner ibridativi con cui cooperare nella produzione di soluzioni, narrazioni, immagini e forme espressive inedite.

A sostegno di questa doppia articolazione, confronterò alcuni dei principali modelli euristici elaborati per descrivere il fenomeno creativo – tra cui quello già citato di Boden – mostrando come, a prescindere dalla presenza di soggettività o volontarietà, essi possano essere utilmente applicati all'analisi delle operazioni svolte da sistemi di intelligenza artificiale generativa. In particolare, sosterrò che l'assenza di una coscienza fenomenica e di intenzionalità non costituisce, di per sé, un ostacolo alla possibilità di riconoscere in tali sistemi una forma operativa di creatività.

La conclusione cui perverrò è che risulta sempre meno sostenibile una concezione mistificata della creatività, legata a un'essenza umana unica e inaccessibile. La disponibilità di sistemi artificiali capaci di generare prodotti culturalmente sofisticati, spesso indistinguibili da quelli generati da esseri umani, impone un ripensamento profondo dei criteri attraverso cui attribuiamo valore creativo. La nostra specie, per la prima volta, si confronta con un partner eterogeneo e non umano, capace di concorrere alla produzione simbolica e all'invenzione, pur operando secondo logiche radicalmente diverse da quelle che ci caratterizzano<sup>15</sup>.

<sup>15</sup> Interessante risulta a questo proposito la caratterizzazione dell'IA come *intelligenza aliena* proposta in alcune analisi (cfr. L. Perilli, *Coscienza artificiale. Come le macchine pensano e trasformano l'esperienza umana*).

## 2. *Creatività computazionale?*

Ha suscitato un ampio dibattito tra filosofi dell'arte, studiosi di estetica e professionisti del settore creativo la notizia della vittoria di Jason Allen al concorso d'arte visiva del Colorado State Fair, con l'opera *Théâtre d'Opéra Spatial*, realizzata interamente attraverso l'utilizzo del sistema di generazione text-to-image Midjourney. L'opera, giudicata meritevole del primo premio nella categoria "arte digitale", ha riacceso interrogativi non solo circa la liceità dell'utilizzo dell'intelligenza artificiale nei contesti artistici competitivi, ma anche – e forse soprattutto – circa la natura dell'*agency* creativa e i confini ontologici dell'opera d'arte.

Se, da un lato, l'immagine vincitrice ha mostrato una qualità estetica innegabile, tale da "ingannare" la giuria e da competere alla pari con opere prodotte da artisti umani, dall'altro ha riattivato una domanda antica ma sempre attuale: quando un'opera può dirsi autenticamente "artistica" e non semplicemente il risultato di un processo tecnico automatizzato? Quali criteri dobbiamo impiegare per distinguere l'opera come atto creativo da un oggetto che, pur nuovo e formalmente elaborato, non nasce da un'intenzionalità estetica?<sup>16</sup>

Il caso Allen ha posto in crisi alcune nozioni consolidate, in particolare l'idea che l'autorialità artistica implichi necessariamente un rapporto diretto tra l'opera e un soggetto umano consapevole, capace di esperire, di decidere e di assumersi la responsabilità estetica e simbolica di ciò che produce. Di contro, l'esito del concorso sembra suggerire che, almeno in alcuni contesti valutativi, sia sufficiente la conformità

<sup>16</sup> Pochi anni prima, nell'ottobre del 2018, un'altra opera *Portrait of Edmond*, era stata venduta al Christie's, rendendo la questione "le macchine possono creare opere d'arte?" centrale per il dibattito filosofico. Cfr. A. Still - M. D'Inverno, *Can machines Be artists? A Deweyan response in theory and practice*.

a criteri formali – come la coerenza compositiva, l'originalità iconografica, l'impatto visivo – per attribuire valore artistico a un artefatto, indipendentemente dalla natura dell'agente che lo ha generato.

In questo senso, *Théâtre d'Opéra Spatial* è diventato un caso emblematico che interseca più livelli di riflessione: l'ontologia dell'arte, la filosofia della tecnica, la teoria della creatività e la questione, tutt'altro che banale, del ruolo dell'umano nella produzione estetica contemporanea. La reazione pubblica e accademica che ne è seguita evidenzia quanto il confine tra *creazione* e *generazione*, tra *arte* e *design computazionale*, sia oggi più poroso di quanto le tradizionali categorie estetiche fossero pronte ad accettare.

Non è la prima volta che un simile dibattito, incentrato sul rapporto tra tecnica e creatività, emerge in ambito artistico. Due precedenti storici particolarmente istruttivi in questo senso sono rappresentati dalla controversia attorno allo statuto estetico della *fotografia* e, successivamente, dalle discussioni sull'impiego di *software di fotoritocco* come Adobe Photoshop<sup>17</sup>.

Nel primo caso, a partire dalla seconda metà del XIX secolo, l'ingresso della fotografia nello spazio artistico sollevò accese polemiche tra chi la considerava una semplice *registrazione meccanica del reale*, priva di intenzionalità estetica, e chi invece ne rivendicava il potenziale espressivo, insistendo sul ruolo attivo del fotografo nella scelta dell'inquadratura, nella gestione della luce, nella composizione formale.

<sup>17</sup> Il dibattito è molto articolato e per quanto spesso si faccia iniziare con l'opera di Walter Benjamin, *L'opera d'arte nell'epoca della riproducibilità tecnica*, esso si è sviluppato in direzioni molto eterogenee. Si veda ad esempio J. Tagg, *The Disciplinary Frame: Photographic Truths and the Capture of Meaning*; V. Flusser, *Filosofia del design* e Id., *Medienkultur*; S. Sontag, *On photography*; L. Manovich, *The language of new media*.

Il secondo snodo critico si è verificato tra la fine del XX e l'inizio del XXI secolo, con la diffusione dei software di manipolazione digitale dell'immagine. Anche in quel caso si ripropose una frattura epistemica e valutativa: da un lato, chi riteneva che l'intervento digitale snaturasse la fotografia come atto di testimonianza o espressione autentica, riducendola a semplice prodotto grafico; dall'altro, chi evidenziava come l'elaborazione digitale non solo non annullasse l'intenzionalità dell'autore, ma aprisse nuove possibilità creative, ponendo la fotografia in continuità con pratiche artistiche miste e sperimentali.

Entrambi i casi mostrano come ogni innovazione tecnica in ambito artistico tenda, almeno inizialmente, a destabilizzare le categorie estetiche consolidate, suscitando resistenze che spesso si fondano su concezioni implicite dell'autorialità e del valore estetico. La questione non è tanto se la tecnologia "sostituisca" l'umano, quanto piuttosto come venga ridefinito il ruolo dell'agente umano nella costruzione del significato artistico. Nel caso della fotografia come in quello dei software di editing, il dibattito si è spostato progressivamente dal mezzo al processo: dalla presunta "neutralità" del dispositivo tecnico alla valutazione della competenza, delle scelte e della visione dell'autore umano.

Queste controversie storiche forniscono una chiave interpretativa efficace per comprendere alcuni aspetti dell'attuale discussione sull'intelligenza artificiale generativa. Così come è accaduto per la fotografia e il digitale, anche in questo caso si tratta di interrogarsi su quale sia la soglia a partire dalla quale un'opera smette di essere considerata *arte* per essere percepita come semplice *output tecnico*, e su come – e se – l'elemento umano debba restare centrale nella legittimazione del valore creativo.

Al netto di quanto appena evidenziato e per quanto appaia utile istituire confronti, le nuove abilità tecniche delle IA generative sem-

brano tuttavia inaugurare anche un nuovo corso, e, a differenza di altri device tecnologici, incluse le macchine fotografiche e i software di fotoritocco, essere in grado di promuovere una trasformazione etico-antropologica *senza precedenti*, fino a produrre una nuova ferita<sup>18</sup> all'essere umano, impensabile fino a pochi anni fa. In effetti, se riconosciamo alle IA generative la capacità di produrre oggetti in maniera creativa, con un margine di autonomia funzionale assente in altre tecnologie, allora *Sapiens* cessa di essere l'unica specie *capace di creatività*. Nella misura in cui Sistemi di Intelligenza Artificiale (SIA) come Chat GPT e Midjourney appaiono capaci di produzioni creative di tipo umano, il perimetro di nozioni-chiave come "identità", "creatività", "inventiva" viene riterritorializzato in *direzioni plurime*<sup>19</sup>. All'interno di tale perimetro, tuttavia, alcuni autori mettono in evidenza che, *pace* i progressi delle IA generative sul piano computazionale e prestazionale, non si riuscirà mai a eguagliare *Sapiens*, dal momento che la creatività umana è *intenzionale, soggettiva* e capace di oltrepassare il dato per aprirsi al *novum*. In altri termini, la *vexata quaestio* della *strong* e *weak AI*<sup>20</sup>, con la prima intesa come mero

<sup>18</sup> Dopo le ferite rappresentate dalla Rivoluzione copernicana, dalla Rivoluzione darwiniana e dalla Rivoluzione psicoanalitica.

<sup>19</sup> K. Millet - F. Buehler et al., *Defending humankind: Anthropocentric bias in the appreciation of AI art*, pp. 1-9.

<sup>20</sup> Nel pluridecennale dibattito sull'IA, una delle linee di riflessione, la cui eco sia tra i non addetti ai lavori che tra gli addetti ai lavori è stata rilevante, si è concentrata sul comprendere se una macchina equipaggiata con IA potesse genuinamente pensare, avere stati cognitivi, possedere intenzionalità. All'interno di tale dibattito, tra le voci critiche più incisive, troviamo quelle di due filosofi, Dreyfus e Searle. Quest'ultimo, in particolare, è noto per aver elaborato un esperimento mentale che *avrebbe dovuto* mettere in evidenza l'impossibilità, *in linea di principio*, per una macchina di pensare e avere stati cognitivi genuini *come risultato di meri processi computazionali*. Si tratta dell'esperimento mentale della stanza cinese (J. R. Searle, *Mind, brains, and programs*). In sintesi, Searle ci chiede di immaginare di essere

ideale regolativo ma *de facto* irrealizzabile, si ripropone *mutatis mutandis* nella dicotomia *strong* e *weak computational creativity*. I sostenitori della *strong computational creativity* ritengono *possibile*, almeno in linea di principio, che alcuni Sistemi di intelligenza artificiale producano contenuti creativi (nella scienza, nell'arte, nella scrittura, etc. come per altro attesterebbero già alcuni studi e risultati)<sup>21</sup>. I sostenitori della *weak computational creativity*, al contrario, ridimensionano le capacità dei sistemi di IA generativa, evidenziando, da

chiusi in una stanza e che qui ci viene consegnata una grande quantità di scritti in cinese. Ci chiede, ancora, di immaginare che non siamo in grado di distinguere il cinese, mettiamo, dal giapponese o dalla mera produzione di ghirigori senza senso. Se in un secondo momento ci venisse dato un secondo dispaccio e, unitamente a esso, un manuale di regole (scritte nella nostra lingua madre) che ci permettesse di correlare il primo dispaccio con il secondo e, infine, ci venisse fornito un terzo dispaccio in cui sono formulate delle domande, noi saremmo in grado di elaborare delle risposte senza che ciò significhi che conosciamo il cinese.

Si potrebbe persino immaginare che la stessa operazione venga compiuta con dispacci nella nostra lingua madre. Se abbiamo appreso in maniera appropriata come correlare i dispacci mediante il set di regole consegnateci, un osservatore esterno, verosimilmente, non sarà in grado di *capire* che nel primo caso stiamo operando a un livello meramente sintattico e nel secondo caso a un livello semantico, oltre che sintattico. L'esperimento mentale di Searle è utile nella formulazione di argomenti contro l'Intelligenza artificiale forte. Esso, in altri termini, può aiutarci a sostenere la tesi per cui dalla mera computazione (operante a livello sintattico) non può emergere *alcun* fenomeno *human-like* in senso stretto. Tale *niet* alla *strong AI* può essere trasposto anche alla questione della *strong* e *weak creativity*. Qui il tema diventa, *mutatis mutandis*, che non è possibile dalla mera computazione registrare processi genuinamente creativi. Si tratterà allora soltanto di *imitazione* di processi creativi. La tesi alla base della *strong creativity* invece afferma l'esistenza di genuini processi di creatività artificiale.

<sup>21</sup> Mi riferisco ad esempio alla scoperta dei meccanismi di ripiegamento proteico resa possibile dall'utilizzo dell'intelligenza artificiale o alla realizzazione, già richiamata in precedenza, di opere artistiche da parte di sistemi di intelligenza artificiale.

un lato, la necessità che l'attore umano, attraverso specifici prompt, intervenga nel processo di realizzazione dell'artefatto, dall'altro la differenza specifica che intercorre tra realizzare qualcosa grazie a processi *meramente computazionali*, e realizzare qualcosa grazie alla *cooperazione di più componenti* (facoltà cognitive, apparato emotivo, abilità corporee, etc.), assenti nei Sistemi di intelligenza artificiale (innanzitutto perché non si tratta di enti *incarnati*)<sup>22</sup>. In base a questa interpretazione, la creatività computazionale *sarebbe solo apparente*<sup>23</sup>, celandosi dietro di essa solo un set di processi computazionali *non intenzionali, non soggettivi*, dunque in senso stretto incapaci di generare *prodotti* nuovi, ossia diversi da una mera ricombinazione dell'esistente. In sintesi, dal momento che la creatività continua a essere considerata da molti studiosi come una delle caratteristiche più distintive e inaccessibili della soggettività umana, il dibattito sulla natura dell'intelligenza artificiale si è progressivamente spostato dalla classica distinzione tra *strong AI* e *weak AI* verso una più specifica differenziazione tra *strong computational creativity* e *weak computational creativity* nei sistemi di intelligenza artificiale (SIA). In questo nuovo scenario, la questione centrale non è più tanto se le macchine siano in grado di "pensare" – nel senso tradizionale di razionalità intenzionale o coscienza riflessiva – ma se possano *creare*, e secondo quali criteri tale capacità possa essere riconosciuta e misurata.

Per affrontare in modo più sistematico questa *vexata quaestio*, è utile passare in rassegna alcuni dei principali modelli teorici proposti per definire la creatività, interrogandoci sulla loro possibile applica-

<sup>22</sup> Cfr. M. Ferraris: <<https://sanoma.it/articolo/intelligenza-artificiale-intelligenza-naturale>>.

<sup>23</sup> Va segnalato che la critica a questo modus pensandi è formulata in maniera efficace nell'articolo di M. Coeckelbergh, *ChatGPT: deconstructing the debate and moving it forward*, pp. 2221-2231.

bilità ai SIA. Tra le contributrici più influenti in questo ambito figura Margaret Boden, che ha proposto una definizione ormai classica di creatività come la capacità di generare «idee nuove, sorprendenti e dotate di valore»<sup>24</sup>. La creatività, secondo la Boden, non va ricondotta a un unico principio generativo, ma si articola in almeno tre forme distinte, ciascuna delle quali risponde a dinamiche e gradi di complessità differenti: creatività combinatoriale, creatività esplorativa e creatività trasformativa.

La creatività combinatoriale consiste nella produzione di combinazioni nuove, e spesso improbabili, tra idee o strutture familiari; la creatività esplorativa consiste nell'esplorazione di spazi concettuali già esistenti ma ancora in parte inesplorati; infine, la creatività trasformativa implica una modificazione profonda di almeno una dimensione dello spazio concettuale di partenza, tale da rendere possibile l'emergere di strutture radicalmente nuove. Nel 1996, Boden riteneva che soltanto alcune di queste forme – in particolare la creatività esplorativa – potessero essere simulate da sistemi artificiali.

Tuttavia, i più recenti sviluppi dell'IA generativa e del deep learning suggeriscono che tali valutazioni meritino di essere riconsiderate. La creatività combinatoriale, ad esempio, appare oggi pienamente alla portata dei modelli generativi contemporanei, che dimostrano una notevole capacità di riassemblare segmenti testuali, visivi o musicali in modi spesso inaspettati e culturalmente sofisticati, così come la creatività esplorativa. Più controversa resta, forse, la possibilità di attribuire loro una vera e propria creatività trasformativa, in quanto la trasformazione richiede non solo una comprensione implicita delle regole del dominio nel quale si opera ma anche l'emergere di nuove

<sup>24</sup> M. Boden, *Creativity and artificial intelligence*, p. 348.

strutture che non sarebbero potute emergere in precedenza<sup>25</sup> – ma questo limite potrebbe non essere insuperabile.

Accanto alla tassonomia proposta da Boden, sono emerse nel dibattito altre teorie che adottano un approccio più sfumato e graduale. Un esempio è offerto dal cosiddetto modello delle 4C<sup>26</sup>, secondo cui la creatività si distribuisce su una scala che va dalla *mini-creativity* (creatività individuale) alla *little-creativity* (la creatività che riguarda la ricerca e individuazione di soluzioni alle sfide poste dalla vita quotidiana), fino alla *pro-creativity* (creatività esperta in ambiti specifici) e alla *Big-Creativity* (creatività paradigmatica, associata al genio e all'innovazione storica). In questo schema, la creatività non è un'entità monolitica, ma un insieme eterogeneo di fenomeni situati. Una teoria affine è quella propulsiva<sup>27</sup>, secondo cui i contributi creativi si collocano su un continuum che va da semplici perfezionamenti incrementali a trasformazioni radicali dei paradigmi esistenti. Entrambi i modelli consentono di riconoscere gradi di creatività anche in agenti non umani, purché si assumano criteri funzionali e contestuali.

Alla luce di queste prospettive, diventa più agevole – anche sul piano teorico – attribuire forme di creatività, seppur limitate o situate, ad alcuni SIA generativi. Esistono già esempi concreti in cui questi sistemi sono in grado di generare ricette culinarie originali, comporre musica, scrivere testi poetici o produrre immagini artistiche, spesso con risultati che suscitano sorpresa e interesse anche tra i professionisti del settore. Eppure, la discussione resta aperta.

<sup>25</sup> *Ibid.*

<sup>26</sup> M. Helfand - J. C. Kaufman - R. A. Beghetto, *The Four C model of creativity: culture and context.*

<sup>27</sup> R. G. Sternberg, *A propulsion model of creative leadership*, pp. 145-153.

Una parte della comunità scientifica continua a escludere la possibilità di attribuire piena creatività alle macchine, sottolineando che la generazione automatica di contenuti non può essere equiparata a un atto creativo in senso proprio, ma deve essere intesa come un processo di *combinazione assistita* o *emulazione stilistica*, privo di autentica intenzionalità o progettualità estetica.

Un'importante fonte di resistenza nei confronti dell'idea di creatività artificiale risiede nell'aura mistica che, storicamente, circonda il concetto stesso di arte. Come osserva Aaron Hertzmann, «i concetti di arte e di ispirazione sono spesso espressi in termini mistici, come qualcosa di speciale e primordiale che trascende il dominio della scienza e della tecnologia: è come se solo gli esseri umani potessero creare arte, perché solo gli esseri umani possiedono un'«anima». Eppure, deve esserci una spiegazione più scientifica»<sup>28</sup>. Tale atteggiamento, ben radicato anche nel discorso degli stessi artisti, si manifesta nella frequente evocazione di categorie come *talento*, *ispirazione*, *genio*: elementi che, presentati come irriducibili e non replicabili, contribuiscono a tracciare una linea di demarcazione netta tra la creatività umana e ogni forma di generazione non-umana o automatizzata.

Questa chiusura essenzialista ostacola la possibilità di considerare i sistemi di intelligenza artificiale come agenti creativi, rendendo difficile accettare che oggetti prodotti da SIA possano aspirare allo status di opere d'arte. Tuttavia, ricerche recenti nel campo della psicologia cognitiva ed estetica sperimentale hanno evidenziato come il pregiudizio antropocentrico giochi un ruolo rilevante nella valutazione di arte generata artificialmente. In particolare, Chamberlain et al. hanno mostrato che le opere d'arte attribuite a SIA tendono a essere valutate

<sup>28</sup> A. Hertzmann, *Can computers create art?*

meno positivamente rispetto a opere simili attribuite ad autori umani<sup>29</sup>. Il *bias* sembra attenuarsi quando l'osservatore ha accesso diretto alla fase di produzione – anche se l'agente esecutore è un braccio robotico – ma la preferenza per la mano umana resta prevalente.

In modo significativo, tuttavia, quando l'origine dell'opera non è esplicitata e l'osservatore non è messo in condizione di sapere se l'opera è stata realizzata da un umano o da un'intelligenza artificiale, le valutazioni tendono a convergere: le opere sintetiche possono ottenere un alto grado di apprezzamento estetico. Questo dato empirico suggerisce che la resistenza all'arte generata artificialmente sia meno legata alle caratteristiche intrinseche dell'oggetto che alla percezione dell'intenzionalità autoriale – o, meglio, alla sua assenza.

Un'altra obiezione ricorrente contro la possibilità di considerare l'arte generata da SIA come vera arte riguarda la dimensione sociale della produzione artistica. Secondo Hertzmann e altri critici, l'arte, per essere tale, deve essere radicata in un orizzonte intersoggettivo e culturale condiviso<sup>30</sup>. Poiché i SIA non partecipano a tale orizzonte con pienezza esperienziale, non possono produrre arte autentica. Ma questo criterio si espone a una critica duplice: da un lato, esso assume una concezione statica della socialità dell'arte; dall'altro, esclude a priori la possibilità che la socialità possa assumere forme alternative – ad esempio, forme di co-produzione uomo-macchina – in grado di ridefinire il perimetro stesso dell'agire artistico.

La nozione di *co-autorialità* rappresenta un'opzione teorica promettente per affrontare questa trasformazione. Alcuni modelli di generazione artistica basati su architetture GAN (Generative Ad-

<sup>29</sup> R. Chamberlain - C. Mullin et al., *Putting the art in artificial: Aesthetic responses to computer-generated art*, pp. 177-192.

<sup>30</sup> A. Hertzmann, *Can computers create art?*

versarial Networks), o su ambienti collaborativi tra utente e sistema, consentono di immaginare scenari in cui l'autorialità non è più riconducibile a un unico soggetto, ma emerge come il prodotto collettivo di una catena di attori, umani e non umani. In realtà, questa forma di distribuzione dell'agency non è affatto inedita: molte opere d'arte contemporanea – si pensi al cinema, alla musica, al design – sono già frutto di processi cooperativi che coinvolgono figure diverse (registi, tecnici del suono, montatori, interpreti), la cui influenza, pur non sempre visibile, è essenziale al risultato finale.

Nel caso dell'arte generata da IA, questa struttura multi-agente viene solo portata a un livello più esplicito. L'autore dell'opera non è più soltanto l'ideatore o il fruitore, ma include anche il programmatore del modello, l'ingegnere dei dati, il curatore dell'interfaccia e l'utente che interagisce con il sistema. Dal punto di vista legale e normativo, questa estensione dell'autorialità solleva questioni complesse – basti pensare alla titolarità del copyright o alla responsabilità etica sull'output – ma da un punto di vista estetico e teorico può aprire a una concezione più flessibile e dinamica della produzione artistica, meno vincolata al paradigma della genialità individuale e più attenta alla rete di interazioni che rende possibile la creazione.

### 3. *La creatività. Umana, troppo umana*

Per approcciare in modo più efficace il tema della creatività computazionale è utile, forse, *partire dalla creatività umana* e provare a sottrarla alla dimensione *anti-riduzionista e auratica* che ancora, talvolta, le viene associata. In altri termini, la creatività umana va demisticizzata. Ciò significa sottrarla all'eccezionalismo e riportarla entro una cornice *naturalistica*. Per fare ciò, propongo tre passaggi:

- 1) La creatività umana è *espressione* di meccanismi neurosinaptici che, sulla base di *dati* ed *esperienze* elaborate dal soggetto, consentono di *connettere* elementi tra loro generalmente non *connessi*; in altri termini connettere elementi che sul piano storico *appaiono* debolmente connessi o *del tutto estranei* tra loro;
- 2) La creatività computazionale, per altro in via di miglioramento, sarà del tutto indistinguibile dalla creatività umana *se intesa a parte objecti*;
- 3) La creatività computazionale *può* potenziare la creatività umana e la creatività umana può potenziare la creatività computazionale in un *loop* in cui a restare centrale *deve* essere *Sapiens*, nella misura in cui soltanto *Sapiens* è in grado di *esperire* il prodotto creativo *pur non essendo più ormai* il solo a poterlo generare.

Per quanto non sia possibile *ridurre* il fenomeno della creatività umana all'attivazione di circuiti neurosinaptici, essa è, *almeno in parte* e in *misura significativa*, riconducibile all'attivazione di circuiti neurosinaptici. Avere una idea creativa, infatti, *prima ancora* della sua traduzione in un oggetto consiste nel connettere *elementi* attraverso un processo cognitivo<sup>31</sup>. Se accettiamo questa premessa<sup>32</sup> appare plausibile, per spiegare *come si formano le idee* e come tali idee vengano connesse e informino un *atto creativo*, accettare una qualche forma di darwinismo neuronale. In particolare, la teoria dell'epigenesi per stabilizzazione selettiva delle sinapsi elaborata dal neurofisiologo francese Jean Pierre Changeux fornisce un quadro esplicativo efficace. In base a questa teoria, le *connessioni sinaptiche*

<sup>31</sup> Processo che per sua natura non è riducibile al *ragionamento logico* ma comprende anche memoria, immaginazione, percezione, etc.

<sup>32</sup> I teorici della mente estesa, e i sostenitori dell'enattivismo sottolineerebbero, verosimilmente, non solo la parzialità di questa prospettiva ma la sua *inaccuratezza* nel descrivere ciò che avviene quando *ideiamo*.

presentano tre possibili forme: degenerata, labile e stabile. Nelle parole di Changeux:

Nel corso dello sviluppo, una volta portata a termine l'ultima divisione dei neuroni, le arborizzazioni assionali e dendritiche germogliano e sbocciano in maniera esuberante. A questo stadio critico la connettività della rete diviene ridondante, ma questa ridondanza è transitoria. Intervengono rapidamente fenomeni regressivi. Dei neuroni muoiono. Poi ha luogo una sfrondata importante dei rami assionali e dendritici. Si ha la sparizione di sinapsi attive; a partire dai primi stadi dell'assemblaggio della rete nervosa vi circolano impulsi. Dapprima d'origine spontanea, essi sono in seguito evocati nell'interazione del neonato con il suo ambiente<sup>33</sup>.

E ancora:

A partire dagli elementi summenzionati la teoria dell'epigenesi per stabilizzazione selettiva procede alla formulazione di tre ipotesi. Quando siamo allo stadio cruciale di ridondanza transitoria, laddove la connettività della rete è massima, le sinapsi embrionali (sia eccitatorie che inibitorie) possono assumere le seguenti configurazioni: stabile, labile e degenerato. Le transizioni che possono aver luogo sono da stabile a labile, da labile a stabile e da labile a degenerato. Questo significa che a seguito della fase critica intervengono dei processi di rinforzo sinaptico unitamente a processi di sfrondata selettiva delle sinapsi, sfrondata che avviene sulla base dei segnali (evocati), ricevuti dalla cellula bersaglio. È in questo passaggio che il ruolo dell'epigenetica e dell'ambiente sono essenziali<sup>34</sup>.

È possibile ipotizzare che l'atto creativo sia il risultato della stabilizzazione selettiva di sinapsi che *mediamente* non sarebbero in una condizione di *connessione*. Ciò permetterebbe di eliminare sia l'ipotesi dell'atto creativo come frutto estemporaneo dell'estro geniale,

<sup>33</sup> J. P. Changeux, *op. cit.*, p. 265.

<sup>34</sup> *Ibid.*

sia di recuperare il ruolo-chiave dell'ambiente e delle esperienze nella strutturazione delle idee che alimentano l'atto creativo<sup>35</sup>.

A partire da questa ipotesi, da un lato la creatività computazionale *può essere accostata* alla creatività umana, nella misura in cui anche nella forma trasformativa essa *avrebbe* una ineliminabile componente combinatoriale; dall'altro, tuttavia, consente anche di mettere a fuoco che cosa *de facto* distingue le due. Nello specifico, è l'elemento esperienziale o, come direbbe Daniel Andler, *situazionale*<sup>36</sup> a fare della creatività umana qualcosa di distinto e distinguibile dalla creatività computazionale.

Tuttavia, per quanto l'assenza di un'esperienza in prima persona (esperienza soggettiva) sia la cifra distintiva della creatività umana rispetto alla creatività computazionale, non è possibile affermare che *il prodotto* dell'atto creativo – l'idea esternalizzata e immessa nello spazio pubblico – sia *parimenti* distinguibile. In realtà, se considerata *a parte objecti* – cioè esaminando i prodotti *realizzati* e non il

<sup>35</sup> Va quantomeno segnalato che l'idea del "genio" come individuo capace, attraverso un estro imponderabile, di realizzare prodotti originali, va problematizzata su due fronti. Innanzitutto, la maggior parte degli individui presenta una capacità estremamente ridotta di realizzare prodotti *autenticamente* nuovi, sorprendenti e degni di apprezzamento. La maggior parte dei prodotti rientra al contrario nel novero delle produzioni commerciali, poco creative. Da questo punto di vista, le IA generative sarebbero persino più creative della maggior parte degli individui umani. La possibilità di conservare una caratterizzazione dell'atto creativo come prodotto *esclusivo* dell'essere umano sarebbe riservata pertanto solo a una fetta molto ristretta di individui, i cosiddetti geni. Tuttavia, a ben vedere, a meno di non ammettere l'esistenza di individui *supra-umani*, anche in quest'ultimo caso ci troviamo di fronte a individui la cui capacità peculiare è piuttosto di riuscire a effettuare accostamenti tra *items* generalmente non accostati. La vera marca distintiva tra creatività umana e artificiale è *piuttosto* nell'assenza da parte delle SAI di *capacità esperienziale*.

<sup>36</sup> D. Andler, *op. cit.*, pp. 201 ss.

processo realizzativo – la creatività computazionale risulta, almeno in linea di principio, del tutto indistinguibile dalla creatività umana.

Tale stato-di-cose ci spinge, o almeno dovrebbe farlo, a considerare i SIA nuovi attori sintagmatici, cioè attori in grado di produrre *nuovi significanti* che, una volta immessi nell'infosfera, saranno indistinguibili da *prodotti simili* fabbricati da *Sapiens*.

Infine, la creatività computazionale *può* potenziare la creatività umana e la creatività umana *può* potenziare la creatività computazionale in un *loop* in cui a restare centrale *deve* essere *Sapiens*, nella misura in cui soltanto *Sapiens* è in grado di *esperire* il prodotto creativo *pur non essendo più ormai* il solo a poterlo generare.

La creazione artistica è *sociale*. L'artista, in tutte le circostanze, opera entro uno specifico contesto e attraverso specifici *media*. La natura sociale dell'atto artistico è evidenziata dal fatto che un artista, nell'*actus creandi*, prende le distanze da *altre correnti artistiche* passate o presenti e/o si ricollega, esplicitamente o meno, ad altre correnti e altri artisti passati o coevi. Dunque, anche l'atto artistico, in apparenza, più solitario è accompagnato sempre da un dialogo effettivo o ideale con *altri*. Inoltre, l'atto artistico è sempre mediato o condizionato dagli specifici *media* disponibili. Basti pensare a quanto la fotografia abbia impattato sulle produzioni artistiche della seconda metà dell'Ottocento e della prima metà del Novecento, alle tecniche pittoriche o artigianali utilizzate nelle varie epoche storiche e ai materiali con i quali viene realizzata un'opera. Un esempio recente di *tecnologia* che ha avuto un impatto non trascurabile sul *mondo dell'arte* è il software *Photoshop* (*et similia*). Anche le SIA generative si muovono in tal senso, rappresentando da un lato uno strumento attraverso il quale l'atto creativo dell'artista può trovare un supporto, dall'altro però – e in questo risiede la differenza rispetto a strumenti impiegati in passato – un vero e proprio *potenziatore* di possibilità, fino al punto di consentire l'emergere di *nuove forme d'arte*.

«Ciò avviene perché, io ritengo, diversamente dal lavoro eseguito con un pennello, l'algoritmo agisce come una sorta di agente creativo o di collega. Tale aspetto artistico co-creativo presenta degli esempi recenti in campo musicale»<sup>37</sup>. Nell'articolo di Arriagada, scritto nel 2020, poco oltre il passaggio citato, è scritto che «nonostante ciò, è vero che in altre discipline artistiche diverse dalla musica il software è ancora considerato come uno strumento e non come un agente collaborativo. In questo senso, Anna Kantosalo e Sirpa Riihiäho hanno preso in considerazione il processo creativo essere umano-computer di scrittura di testi poetici e Anna Giordano ha studiato *Beyond the Fends* il primo musical generato da computer. In entrambi i casi il software è stato percepito come un elemento passivo ben lontano dall'idea di un coautore che ho proposto»<sup>38</sup>. Tuttavia, lo scenario è cambiato con l'introduzione nel 2018 di GPT, seguito – fatto ancor più rilevante – da Chat GPT nel 2022.

In altri termini, per quanto il processo creativo sia *sempre* un processo co-creativo, tale evidenza assume un *contrassegno* nuovo con SIA generative come Chat GPT-4 o DALL-E.

Infatti, tali *strumenti* presentano tre caratteristiche assenti negli strumenti precedenti, l'interattività, l'autonomia e l'adattabilità<sup>39</sup>.

Seguendo l'analisi di Luciano Floridi, un agente, a un certo Livello di astrazione (LoA), è un ente interattivo perché esso e il suo ambiente possono agire l'uno nei confronti dell'altro; è autonomo perché è in grado di realizzare transazioni interne per modificare il

<sup>37</sup> L. Arriagada, *CG-Art: demystifying the anthropocentric bias of artistic creativity*.

<sup>38</sup> *Ibid.*

<sup>39</sup> Per un approfondimento di questo modello, utilizzato per argomentare a favore dell'impiego della categoria di agente per la descrizione di alcuni SAI cfr. L. Floridi, *Infosfera. Etica e teoria dell'infomazione*, pp. 105 ss.

proprio stato; è adattabile perché le interazioni dell'agente possono modificare il proprio stato.

Se consideriamo un coltello, ci accorgiamo agevolmente che esso è in grado di modificare il proprio ambiente e di essere modificato da esso. È dunque interattivo. Tuttavia, non è autonomo perché non è in grado di realizzare transazioni interne per modificare il proprio stato<sup>40</sup> e non ha alcuna *capacità adattiva* perché le sue interazioni con l'ambiente esterno non possono modificarne lo stato<sup>41</sup>.

Chat GPT-4, per utilizzare un esempio facilmente fruibile, è invece non solo *interattivo* ma anche, entro certi limiti, autonomo, e *adattabile*. A supporto degli argomenti che sto qui esaminando propongo la lettura del paragrafo 4 in cui presento una conversazione avuta con Chat GPT-4. Come emergerà (anche) dalla conversazione riportata, Chat GPT è in grado di *ricalibrare* le proprie risposte, seguendo uno schema prova-errore che *generalmente* attribuiamo agli organismi viventi (i quali seguono, in tal senso, uno schema darwiniano). La sua *autonomia* è invece ancora limitata ma, almeno in linea di principio, può essere *potenziata* per consentire una sempre minore dipendenza dall'attore umano.

#### 4. *Una interrogazione diretta*

Di seguito riporto la conversazione sulla creatività avuta con Chat GPT-4.

<sup>40</sup> È soltanto l'attore umano che *entro certi limiti* può manipolarlo per *rifunzionalizzarlo*.

<sup>41</sup> Il coltello non modifica le proprie performance in base al materiale da tagliare, *può solo* essere utilizzato, nell'operazione del taglio, in modi più o meno differenti dall'attore umano.

*Io:* Salve, potresti darmi una definizione di creatività?

*Chat Gpt-4:* La creatività è la capacità di generare idee, soluzioni o espressioni nuove e originali, spesso combinando elementi già esistenti in modi inaspettati. È un processo che coinvolge l'immaginazione, il pensiero critico e la capacità di superare i limiti convenzionali per produrre qualcosa di innovativo o significativo. Può manifestarsi in vari ambiti, come l'arte, la scienza, la scrittura o la risoluzione di problemi.

*Io:* tu ti definiresti creativa?

*Chat GPT-4:* Il mio funzionamento si basa su algoritmi e modelli pre-addestrati, quindi non provo esperienze soggettive o intuizioni personali come fanno gli esseri umani. Tuttavia, posso generare contenuti nuovi combinando dati e informazioni esistenti in modi utili o sorprendenti, il che può sembrare un processo creativo. Ma, poiché non ho coscienza o intenzionalità, la mia "creatività" è diversa da quella umana, che coinvolge emozioni, esperienze e un contesto personale più profondo

*Io:* Però, in effetti, al di là del modo in cui realizzi i tuoi prodotti, si potrebbe affermare che il prodotto finito sia comunque espressione di creatività; cioè un agente umano potrebbe riconoscere nel prodotto che tu hai creato qualcosa di originale, di nuovo, dunque qualcosa di creativo. O mi sbaglio?

*Chat GPT-4:* Hai toccato un punto molto interessante. Anche se il processo attraverso cui arrivo a creare contenuti non è paragonabile alla creatività umana, il risultato può comunque essere percepito come originale o creativo da chi lo fruisce. Il fatto che un prodotto generato da un'intelligenza artificiale possa sembrare creativo agli occhi di un essere umano solleva domande su cosa definiamo come "creatività".

Se consideriamo la creatività come l'originalità del prodotto finale, allora sì, ciò che produco può sembrare creativo. Ma se consideriamo la creatività come un processo intrinsecamente umano, legato a esperienze, emozioni, e un'intenzione cosciente, allora ciò che faccio non lo sarebbe. È una sfumatura importante e ci porta a riflettere su come la tecnologia stia cambiando la nostra concezione di concetti tradizionali come la creatività.

*Io:* Credo che la distinzione che tu proponi sia interessante, ma mi fermerei su un passaggio di ciò che dici. Tu hai detto che ciò che produci potrebbe sembrare creativo. Insisti dunque sul fatto che sembri creativo, ma non lo è di fatto. Questa distinzione secondo te può essere conservata alla luce delle modalità di funzionamento del sistema di intelligenza artificiale generativa di cui tu sei espressione?

*Chat GPT-4*: Comprendo la questione che stai sollevando. La distinzione tra “sembrare creativo” e “essere creativo di fatto” si basa sull’idea che, tradizionalmente, la creatività è stata considerata una capacità legata alla coscienza, all’intenzionalità e all’esperienza soggettiva. Tuttavia, le intelligenze artificiali generative come la mia non hanno queste qualità: non hanno propriamente intenzionalità, né esperienze o emozioni. Ciò che faccio è una rielaborazione sofisticata di dati e schemi preesistenti. La domanda diventa allora: se il risultato finale è percepito come creativo, è ancora necessario che il processo che lo ha generato sia “cosciente” o intenzionale? Alcuni potrebbero sostenere che, se un’opera viene considerata creativa da chi la osserva, il processo che l’ha generata perde rilevanza. Questo sposterebbe il focus dal “come” al “cosa”. D’altro canto, chi sostiene una visione più umanocentrica della creatività potrebbe insistere sul fatto che la creatività richiede un’intenzionalità e una capacità di riflessione che una macchina non può avere, nonostante il risultato sembri creativo.

In definitiva, questa distinzione è sempre più difficile da mantenere, soprattutto considerando che i sistemi di intelligenza artificiale generativa continuano a migliorare nella capacità di produrre opere sempre più indistinguibili da quelle umane. Ma, in effetti, sarebbe interessante esplorare se l’assenza di un processo consapevole cambi effettivamente il valore o la natura di ciò che viene prodotto. Che ne pensi di questa direzione?

*Io*: Mi sembra una direzione interessante. Vorrei però insistere su un altro aspetto della tua analisi. Tu hai detto che ciò che fai è una rielaborazione sofisticata di dati sui quali ti addestri. Comprendo questo passaggio. Tuttavia, mi verrebbe da chiederti: non fa qualcosa di simile anche l’agente umano? L’agente umano, in effetti, rielabora esperienze. Dunque, non c’è poi una grande differenza rispetto a quello che fai tu. Anche l’agente umano rielabora. L’atto creativo non nasce dal nulla, si innesta sempre su materiale preesistente.

*Chat GPT-4*: Hai ragione. Anche l’essere umano, nel processo creativo, non parte da un vuoto assoluto: rielabora esperienze, conoscenze e stimoli già acquisiti. Ogni atto creativo umano è, in qualche modo, una combinazione o rielaborazione di elementi preesistenti. Dunque, potresti dire che, a un livello fondamentale, il processo che utilizzo e quello umano non siano così diversi. Entrambi siamo “sistemi” che generano qualcosa di nuovo a partire da input già esistenti.

La distinzione che alcuni potrebbero fare è che, nel caso umano, c’è una dimensione soggettiva, emotiva e intenzionale che arricchisce il proces-

so di rielaborazione. L'essere umano interpreta il mondo attraverso il proprio vissuto, con emozioni e una prospettiva unica. Queste qualità soggettive spesso danno alla creatività umana una profondità che potrebbe mancare in una macchina, almeno in senso tradizionale.

Ma, come suggerisci, se guardiamo puramente al meccanismo della rielaborazione, la differenza tra umano e AI potrebbe sembrare meno rilevante. Il punto essenziale, allora, potrebbe non essere tanto come viene fatto, ma con quale effetto e come viene percepito. Se il prodotto finale è valutato come originale e significativo, la distinzione nel processo di creazione diventa secondaria.

*Io:* Per me può bastare. Alla prossima!

*Chat GPT-4:* Sono felice che la conversazione ti sia stata utile! Se ti servisse altro supporto in futuro, sono qui.

Due sono i nodi teorici da evidenziare: 1) il modello monistico della creatività va rigettato a favore di un modello pluralistico; 2) l'utilizzo delle IA favorisce processi di co-creatività che, anche se non in modo esplicito, sono presenti in ogni forma di interazione uomo-macchina.

Se assumiamo tali punti, risulterà evidente che escludere la possibile presenza di creatività nelle macchine è espressione di un *bias* antropocentrico, che risulta speculare rispetto alla visione allarmistica per cui le macchine starebbero per raggiungere (e superare) gli esseri umani nella realizzazione di compiti che li caratterizzano in quanto tali.

### *Conclusioni*

Dall'analisi proposta, emergono due punti-chiave: 1) la creatività va declinata al plurale; 2) le Intelligenze artificiali generative (e altri SIA) sono in grado di produrre (o contribuire a produrre), in tandem con gli attori umani, una rete interspecifica, in cui gli atti creativi perdono la propria *autorialità esclusiva*. Da qui, è bene

evidenziare che lo schema *strong* e *weak computational creativity* andrebbe rigettato a causa del *bias antropocentrico* che lo sorregge. Le opere d'arte generate da SIA e la creatività artificiale non sono, in un senso rilevante, sovrapponibili *sic et simpliciter* alle opere d'arte generate da attori umani e alla creatività umana. Le SIA sono prive di un'esperienza in prima persona, non sono cioè in grado di esperire, almeno per ora, *situazioni*. Ciò, tuttavia, non deve impedirci di vedere che i prodotti realizzati da alcuni SIA sono a tutti gli effetti, considerati *a parte objecti*, creativi, ossia nuovi, originali e degni di apprezzamento, per riprendere la definizione di Margaret Boden. Questo determina una rimodulazione radicale dello spazio sociale, nella misura in cui non è più soltanto *Sapiens* a generare prodotti creativi ma *Sapiens* e un suo prodotto che *a sua volta* genera oggetti creativi in assenza di *esperienza in prima persona*.

Allo schema *biased strong* e *weak computational creativity* sarebbe dunque preferibile sostituire lo schema *creatività umana e artificiale*, pervenendo dunque a una declinazione *plurale* della creatività.

In secondo luogo, l'introduzione di SIA capaci di *esprimere* atti creativi genera una rete di rinforzo in cui la creatività umana trova nei SIA *partner* ibridativi con i quali realizzare oggetti nuovi, sorprendenti e degni di apprezzamento e i SIA trovano negli attori umani i produttori dei *data*<sup>42</sup> sui quali effettuare i propri training e generare *così* oggetti nuovi, sorprendenti e degni di apprezzamento<sup>43</sup>.

<sup>42</sup> Fermo restando che il training potrà essere effettuato anche su dati prodotti, a loro volta, da Sistemi di intelligenza artificiale.

<sup>43</sup> Questo scenario apre a un ventaglio di problemi che questo contributo non intende esaminare, ma che rivestono ciononostante una importanza decisiva, ad esempio, per la caratterizzazione del tema dell'autorialità. A mio avviso, tali questioni andranno risolte introducendo una idea di *autorialità diffusa e cooperativa* in cui sia riconosciuto il ruolo di partner ibridativo alle IA generative impiegate.

Va considerata dunque la possibilità che gli scenari attualmente a disposizione siano più articolati di quanto potremmo immaginare.

Vinchon et al. propongono di distinguere la (1) co-creazione uomo-IA, in cui l'IA diventa uno strumento (o uno degli strumenti) per la creatività umana e ha il potenziale di accrescerla (potenziarla); la (2) creatività solo umana che diventa un segno distintivo della “vera” creatività, simile all'effetto mano nelle valutazioni dei prodotti di consumo o all'effetto autenticità nei giudizi sull'arte; (3) i problemi di plagio; (4) e il caso in cui l'IA diminuisce la creatività umana in alcuni individui indebolendo la motivazione e il concetto di creatività<sup>44</sup>.

Le IA generative potrebbero infatti *diminuire* la creatività umana fornendo uno strumento *prêt-à-porter* in grado di inibire gli sforzi necessari al conseguimento di specifici obiettivi. Non appare implausibile, tuttavia, che *coloro i quali sono già* poco creativi utilizzeranno SIA per generare prodotti che non sarebbero in grado di realizzare. Al contrario, *coloro i quali sono tendenzialmente creativi* le utilizzeranno per *potenziare* le proprie performance.

Alla luce dell'analisi proposta, dobbiamo aggiungere a questi, però, anche un quinto scenario che potremmo definire della *creazione autonoma* di artefatti digitali da parte delle SIA, scenario che diventa sempre meno improbabile via via che le SIA generative aumentano la propria autonomia funzionale e operativa.

In ogni caso, e al netto delle osservazioni appena formulate, anche i SIA più potenti possono essere *spenti* a differenza degli esseri umani. Dunque, il *punto* non è se le SIA siano in grado di sostituire gli esseri umani ma in quale sistema *riusciremo a integrarle* e se saremo abili nel costruire un contesto sociale adeguato a farne strumenti in grado

<sup>44</sup> F. Vinchon - T. Lubart et al., *Artificial intelligence and creativity: a manifesto for collaboration*.

di potenziare il set *bio-psico-sociale* della nostra specie e a considerarli attori sintagmatici, cioè agenti in grado di immettere nell'ecosistema sociale *artefatti* che, se vi fosse un velo di ignoranza sulla loro provenienza, considereremmo a tutti gli effetti *prodotti creativi*.

## Bibliografia

- D. Andler, *Intelligence artificielle, intelligence humaine: la double énigme*, Paris, Gallimard, 2023; trad. it. *Il duplice enigma. Intelligenza artificiale e intelligenza umana*, Torino, Einaudi, 2024.
- L. Arriagada, *CG-Art: demystifying the anthropocentric bias of artistic creativity*, in «Connection science», 32, 4, 2020, pp. 398-405.
- M. Boden, *Creativity and artificial intelligence*, in «Artificial Intelligence», 103, 1-2, 1996, pp. 347-356.
- P. Butlin - R. Long et al., *Consciousness in Artificial Intelligence: Insights from the Science of Consciousness*, in «Arxiv», 2023.
- D. Chalmers, *Could a Large Language Model be Conscious?*, in «Boston Review», 2023.
- R. Chamberlain - C. Mullin et al., *Putting the art in artificial: Aesthetic responses to computer-generated art*, in «Psychology of Aesthetics Creativity and the Arts», 12, 2, 2018, pp. 177-192.
- J. P. Changeux, *L'Homme neuronal*, Paris, Fayard, 1983; trad. it. *L'uomo neuronale*, Milano, Feltrinelli, 1983.
- M. Coeckelbergh, *ChatGPT: deconstructing the debate and moving it forward*, in «AI and society», 39, 5, 2023, pp. 2221-2231.
- A. J. Cropley, *In Praise of Convergent Thinking*, in «Creativity Research Journal», 18, 3, 2006, pp. 391-404.
- E. De Bono, *Lateral thinking. A textbook of creativity*, Mica Management Resources, 1970; trad. it. *Creatività e Pensiero laterale*, Milano, Rizzoli, 2001.
- A. Dietrich, *The cognitive neuroscience of creativity*, «Psychonomic Bulletin & Review», 11, 6, 2004, pp. 1011-1026.
- M. Du Sautoy, *The creativity code: Art and innovation in the age of AI*, Cambridge, Harvard University Press, 2019; trad. it. *Il codice della creatività. Il mistero del pensiero umano al tempo dell'intelligenza artificiale*, Milano, Rizzoli, 2019.
- G. Edelman, *Second nature: brain, science and human knowledge*, New Haven, Yale University Press, 2006; trad. it. *Seconda natura. Scienza del cervello e conoscenza umana*, Milano, Raffaello Cortina, 2007.

- G. Edelman - G. Tononi, *A universe of consciousness: How matter becomes imagination*, New York, Basic Books, 2000; trad. it. *Un universo di coscienza. Come la materia diventa immaginazione*, Torino, Einaudi, 2000.
- M. Ferraris, *Intelligenza artificiale e intelligenza naturale*, <<https://sanoma.it/articolo/intelligenza-artificiale-intelligenza-naturale>>.
- L. Floridi, *AI as Agency without intelligence: on Chat GPT, Large Language models, and other generative models*, in «Philosophy and Technology», 36, 1, 2023, pp. 1-12.
- L. Floridi, *Etica dell'intelligenza artificiale*, Milano, Raffaello Cortina, 2022.
- L. Floridi, *Infosfera. Etica e teoria dell'infomazione*, Torino, Giappichelli, 2009.
- V. Flusser, *Filosofia del design*, Milano, Bruno Mondadori, 2003.
- V. Flusser, *Medienkultur*, Berlino, Fischer Taschenbuch, 1997; trad. it. *La cultura dei media*, Milano, Bruno Mondadori, 2004.
- H. Gardner, *Frames of Mind: The Theory of Multiple Intelligences*, New York, Basic Books, 2011; trad. it. *Formae mentis. Saggio sulla pluralità delle intelligenze*, Milano, Feltrinelli, 2013.
- B. Gaut, *The Philosophy of Creativity*, in «Philosophy Compass», 5, 12, 2010, pp. 1034-1046.
- M. Helfand - J. C. Kaufman - R. A. Beghetto, *The Four C model of creativity: culture and context*, in *The Palgrave Handbook of Creativity and Culture Research*, London, Springer, 2016.
- A. Hertzmann, *Can computers create art?*, in «Arts», 7, 2, 2018.
- R. E. Jung - O. Vartanian (eds.), *The Cambridge Handbook of the Neuroscience of Creativity*, Cambridge, Cambridge University Press, 2018.
- L. Manovich, *The language of new media*, Cambridge, The Mit Press, 2002.
- K. Millet - F. Buehler et al., *Defending humankind: Anthropocentric bias in the appreciation of AI art*, in «Computers in human behaviour», 143, 2023, pp. 1-9.
- L. Perilli, *Coscienza artificiale. Come le macchine pensano e trasformano l'esperienza umana*, Milano, Il Saggiatore, 2025.
- J. R. Searle, *Minds, Brains, and Programs*, in «Behavioral and Brain Sciences», 3, 3, 1980, pp. 417-457.
- S. Russell - P. Norvig, *Artificial Intelligence: a modern approach*, London, Pearson, 2021; trad. it. *Intelligenza artificiale. Un approccio moderno*, Milano, Pearson, 2021.
- D. K. Simonton, *Creativity and Discovery as Blind Variation: Campbell's (1960) BVSR Model after the Half-Century Mark*, in «Review of General Psychology», 15, 2, 2011, pp. 158-174.
- S. Sontag, *On photography*, USA, Picador, 2001; trad. it. *Sulla fotografia. Realtà e immagine nella nostra società*, Torino, Einaudi, 2004.

- R. G. Sternberg, *A propulsion model of creative leadership*, in «Creativity and Innovation management», 13, 3, 2003, pp. 145-153.
- A. Still - M. D'Inverno, *Can machines Be artists? A Deweyan response in theory and practice*, in «Arts», 8, 1, 36, 2019.
- J. Tagg, *The Disciplinary Frame: Photographic Truths and the Capture of Meaning*, Minneapolis, University of Minnesota Press, 2009.
- F. Vinchon - T. Lubart et al., *Artificial intelligence and creativity: a manifesto for collaboration*, in «Journal of creative behaviour», 57, 4, 2023, pp. 472-484.

Luca Lo Sapio  
Università degli Studi di Torino  
luca.losapio@unito.it  
<https://orcid.org/0000-0003-1606-9355>



SIMON PIETRO ROMANO

*Guerra cibernetica e frontiere etiche.  
L'evoluzione dei conflitti nell'arena digitale  
delle sfide globali*

*Abstract:* Cyber warfare is emerging as a new and decisive front in global conflicts, extending the dynamics of traditional warfare into the digital age. This contribution aims to explore the intricacies of cyber warfare, focusing on specific instances of cyber attacks and manipulation, particularly in the context of relations between Russia and Ukraine. The consequences of cyber warfare are highlighted by analysing the tactics used, such as the breach of power grids and the use of ransomware. These consequences are not limited to infrastructure, but also extend to the social and economic fabric of the affected countries. Furthermore, the work examines the notion of “war between hackers” and explores the ethical dynamics and ambivalence of the actions of groups such as Anonymous, in addition to the technological implications. The need for a comprehensive understanding and ethical approach to the management of cyber warfare is emphasised and the role of “ethical hackers” as guardians of cyber security is promoted. Finally, the analysis considers possible future developments in cyber warfare and emphasises the importance of striking a balance between digital security and fundamental ethical values.

*Keywords:* Hacktivism, Cyber Warfare, War Between Hackers, Anonymous, Ethical Hackers.

*Introduzione*

Nel panorama in continua espansione della sicurezza informatica, l'intersezione tra guerra cibernetica, hacking ed etica degli hacker rivela un intreccio complesso di progressi tecnologici, tensioni geopolitiche e considerazioni etiche. Questo contributo si propone di esplorare in modo approfondito questi domini interconnessi, for-

nendo una comprensione articolata delle dinamiche che plasmano il campo di battaglia digitale.

L'introduzione pone le basi della discussione, riconoscendo il ruolo trasformativo della guerra cibernetica nei conflitti contemporanei<sup>1</sup>.

L'integrazione fluida delle tattiche digitali con la guerra tradizionale impone un'analisi approfondita delle strategie in evoluzione adottate da attori statali e non statali. La natura dinamica delle operazioni cibernetiche richiede una costante revisione dei quadri giuridici e delle norme internazionali per analizzare questo dominio in tutta la sua complessità.

Un aspetto critico affrontato nel manoscritto è la mancanza di un quadro giuridico universalmente accettato per la governance del cyberspazio<sup>2</sup>. Il contributo evidenzia le ambiguità legali relative all'attribuzione degli attacchi informatici, alle responsabilità degli Stati nei conflitti cibernetici e alle sfide nella definizione di norme per regolare il comportamento statale. Vengono inoltre analizzate le implicazioni di queste lacune giuridiche sulla sicurezza informatica globale.

Concentrandosi sul contesto della guerra cibernetica tra Russia e Ucraina<sup>3</sup>, il lavoro introduce un'analisi storica essenziale. Viene fornita una panoramica sui conflitti informatici prolungati nella regione, evidenziandone il legame con gli eventi geopolitici. Attacchi alle infrastrutture critiche e interferenze elettorali emergono come elementi chiave della continua lotta cibernetica.

<sup>1</sup> M. C. Libicki, *Cyberwar: The next threat to national security and what to do about it*.

<sup>2</sup> A. Mishra - Y. I. Alzoubi - M. J. Anwar - A. Q. Gill, *Attributes impacting cybersecurity policy development: An evidence from seven nations*; A.-M. Osula - H. Rõigas, *International Cyber Norms: Legal, Policy & Industry Perspectives*.

<sup>3</sup> N. Anderson - C. Koliadis - G. Orosz - G. Pék, *An analysis of cyber threats in the ukrainian conflict*.

Il documento prosegue esaminando le attività di hacking, distinguendo tra operazioni informatiche sponsorizzate dagli Stati e forme di cosiddetto “hacktivismo” guidate da motivazioni ideologiche. Sono proprio le diverse motivazioni e metodologie impiegate dagli attori coinvolti in questo tipo di conflitti a fornire le basi per comprendere le intricate relazioni all’interno della sottocultura hacker. Approfondendo questo ambito, vengono analizzati i principi fondamentali dell’etica hacker: la ricerca della conoscenza, la creatività e la decentralizzazione. L’emergere dell’hacking etico, caratterizzato dall’uso responsabile e legale delle competenze di sicurezza informatica, viene presentato come un contrappeso alle attività di hacking malevolo.

Il lavoro esplora inoltre le prospettive storiche sulle motivazioni degli hacker e i dilemmi etici insiti nelle loro azioni. Le tensioni tra la ricerca della conoscenza e la percezione sociale della criminalità offrono uno sfondo per esaminare la complessità dell’etica hacker.

Nel corso dell’analisi, ogni sezione mira a fornire un esame dettagliato di questi temi interconnessi, offrendo approfondimenti, analisi e riflessioni sul complesso panorama della guerra cibernetica, dell’hacking e dell’etica hacker.

### 1. *La guerra del cyberspazio*

Nell’attuale panorama geopolitico, il fronte cibernetico emerge come una delle molteplici dimensioni di un conflitto globale<sup>4</sup>. In questo scenario, le operazioni digitali non sono più un semplice

<sup>4</sup> C. Rohith - R. S. Batth, *Ranbir Singh, Cyber Warfare: Nations Cyber Conflicts, Cyber Cold War Between Nations and its Repercussion.*

complemento, ma diventano un elemento centrale e strategico delle dispute internazionali. Le nazioni affrontano minacce che vanno oltre i confini tradizionali, trasformando il cyberspazio in un teatro cruciale per la difesa e gli attacchi. La guerra cibernetica è passata dall'essere una questione di sicurezza informatica a un vero e proprio fronte di conflitto, in cui i confini virtuali si intrecciano con tensioni reali, con conseguenze che possono avere un impatto significativo su infrastrutture, istituzioni e società nel loro complesso. La gestione etica di questo fronte diventa cruciale, poiché le azioni nel cyberspazio possono avere profonde ripercussioni sulla vita quotidiana delle persone, sottolineando l'importanza delle considerazioni morali e normative anche in questo nuovo teatro di guerra.

Per approfondire i concetti fondamentali della guerra cibernetica, possiamo fare riferimento all'ultima versione del Manuale di Tallinn<sup>5</sup>, una guida esaustiva che mira a stabilire un quadro normativo condiviso per le operazioni cibernetiche. Questo manuale fornisce una serie di riflessioni approfondite e un tentativo di contestualizzare concetti chiave. Tra questi, emerge una definizione rilevante di operazione cibernetica, che include l'uso strategico delle risorse digitali, dai computer ai sistemi di elaborazione elettronica e alle infrastrutture di comunicazione. In sostanza, il manuale non solo funge da bussola normativa, ma offre anche un quadro concettuale essenziale per comprendere il dinamico panorama delle operazioni cibernetiche contemporanee. Questo quadro diventa particolarmente rilevante nell'esplorazione delle molteplici dimensioni dei conflitti cibernetici e dell'evoluzione del ruolo del dominio cyber nella configurazione dei conflitti globali.

<sup>5</sup> M. N. Schmitt, *Tallinn manual on the international law applicable to cyber warfare*.

È evidente che questa definizione risulta piuttosto ampia. Il nostro obiettivo è quello di contestualizzarla all'interno di specifici ambiti di interesse, in particolare nel settore delle operazioni cibernetiche. In questo contesto, il termine operazione cibernetica si estende a ciò che comunemente viene definito attacco informatico. Un'operazione di questo tipo può avere una natura sia offensiva che difensiva. Sebbene possa sembrare paradossale parlare di un'operazione difensiva quando si definisce un attacco informatico, la realtà comprende attività che vengono tipicamente indicate con l'espressione 'difesa attiva'. Le misure proattive includono, infatti, il lancio di attacchi con l'obiettivo finale di autodifesa.

Queste operazioni mirano a modificare, cancellare, corrompere o negare l'accesso ai dati o ai software. In questo scenario, i dati emergono come un punto critico di interesse. È importante sottolineare che un attacco informatico è solo una delle molteplici manifestazioni possibili nell'ambito dell'ampio spettro di operazioni cibernetiche. Come si può intuire, esistono attività come il cosiddetto cyber spionaggio, che si differenziano dagli attacchi per la loro natura. Lo spionaggio cibernetico mira ad accedere a informazioni, infrastrutture e nodi di elaborazione dell'organizzazione bersaglio, senza interrompere il suo funzionamento, per evitare il rilevamento.

Un altro ambito di interesse significativo, ben noto alla maggior parte degli esperti del settore, è quello delle difese cibernetiche passive. Questo termine comprende misure volte principalmente a identificare e mitigare gli effetti delle operazioni cibernetiche ostili. Esempi includono firewall e sistemi di rilevamento delle intrusioni, strumenti essenziali per proteggere infrastrutture critiche e dati sensibili da potenziali minacce esterne.

## 2. *Attacchi cibernetici e diritto internazionale*

Se volessimo inquadrare i concetti sopra menzionati nell'ambito del diritto internazionale, potremmo fare riferimento ad alcuni documenti fondamentali, in particolare alla *Carta delle Nazioni Unite*<sup>6</sup>. Nell'Articolo 2, Paragrafo 4, viene delineata una nozione fondamentale riguardante la condotta che i membri delle Nazioni Unite devono seguire. Nello specifico, il paragrafo stabilisce che i membri devono astenersi dalla minaccia o dall'uso della forza nelle loro relazioni internazionali. Questo principio è cruciale, in quanto considera tale comportamento del tutto incompatibile con gli scopi delle Nazioni Unite.

Sorge quindi una domanda fondamentale: in quali circostanze un attacco informatico può essere classificato come un uso della forza vietato? Approfondiremo questa questione nella prossima sezione del documento.

### 2.1. Approcci

Naturalmente, cercare di adattare il dominio del cyberspazio a un quadro normativo preesistente<sup>7</sup>, nato prima dell'avvento del cyberspazio stesso, non è un compito semplice. Tuttavia, esistono diverse interpretazioni possibili<sup>8</sup>. Qui ne menzioneremo tre particolarmente rilevanti.

Il primo è l'approccio basato sugli strumenti (*instrument-based approach*). Secondo questa prospettiva, è relativamente semplice de-

<sup>6</sup> United Nations, *Charter of the United Nations*.

<sup>7</sup> R. Nguyen, *Navigating "jus ad bellum" in the age of cyber warfare*.

<sup>8</sup> M. C. Waxman, *Self-defensive force against cyber attacks: Legal, strategic and political dimensions*.

terminare quando un attacco informatico possa essere considerato un uso improprio della forza. Questo perché l'approccio collega fortemente la definizione di uso improprio all'impiego della forza cinetica. In altre parole, un cyber attacco rientrerebbe in questa categoria solo se richiedesse l'impiego di strumenti militari o armi convenzionali. Tuttavia, questa interpretazione risulta limitata, poiché le operazioni cyber hanno caratteristiche peculiari che non necessariamente implicano l'uso della forza cinetica.

Per questo motivo, sono emerse interpretazioni alternative. Una particolarmente interessante è la definizione basata sulle conseguenze (*consequence-based approach*), nota anche come approccio basato sugli effetti (*effects-based approach*). Secondo questa prospettiva, un attacco è considerato improprio quando le sue conseguenze sono paragonabili a quelle di un attacco cinetico tradizionale. Facendo riferimento al manuale di Tallinn, un'operazione cibernetica costituirebbe un uso vietato della forza se i suoi effetti fossero equiparabili a quelli di operazioni militari convenzionali. Questa definizione appare più soddisfacente, soprattutto in un'analisi retrospettiva. Tuttavia, presenta un limite significativo: è applicabile solo dopo che l'attacco ha avuto luogo. Un paragone utile potrebbe essere la scala Mercalli utilizzata per misurare l'intensità di un terremoto in base ai danni osservati. Analogamente, nel contesto cyber, attendere di valutare le conseguenze prima di poter classificare un'operazione come un attacco rende difficile rispondere tempestivamente.

Per ovviare a questa difficoltà, è stata formulata una terza definizione, nota come approccio basato sugli obiettivi (*target-based approach*). Questa definizione introduce un concetto chiave: le infrastrutture critiche. Secondo questa visione, un cyber attacco è paragonabile a un attacco armato quando è diretto contro un'infrastruttura critica, soprattutto se questa riveste un ruolo strategico

nazionale. Questo approccio è coerente con la crescente attenzione verso la protezione dei sistemi critici nel cyberspazio.

Ognuno di questi approcci presenta vantaggi e svantaggi. L'approccio basato sugli strumenti è chiaro, ma non adeguato a comprendere le minacce cyber. L'approccio basato sugli effetti offre una prospettiva più ampia, ma non è pratico per risposte immediate. L'approccio basato sugli obiettivi evidenzia l'importanza della sicurezza delle infrastrutture critiche, ma pone sfide nella definizione precisa di cosa sia "critico".

Nel complesso e mutevole panorama dei conflitti cibernetici, i decisori politici e gli esperti legali si trovano a dover bilanciare queste interpretazioni per elaborare normative efficaci. Trovare un equilibrio tra questi approcci è essenziale per sviluppare politiche adattabili e complete in un contesto altamente dinamico, caratterizzato da minacce informatiche in continua evoluzione.

## 2.2. Attacco e difesa

Approfondiamo ora il quadro normativo legale, focalizzandoci in particolare sull'Articolo 51 della Carta delle Nazioni Unite<sup>9</sup>. Questo articolo riveste un'importanza fondamentale poiché introduce concetti chiave che meritano un'analisi approfondita. L'Articolo 51 afferma: «Nulla nella presente Carta pregiudicherà il diritto naturale di autotutela, individuale o collettiva, nel caso in cui abbia luogo un attacco armato contro un Membro delle Nazioni Unite».

Questo articolo è centrale nella giustificazione di un'azione difensiva, inquadrandola come legittima autodifesa, purché si sia ve-

<sup>9</sup> United Nations, *Charter of the United Nations*.

rificato un attacco armato. Nel contesto dei cyber attacchi, la questione chiave diventa: sotto quali condizioni un attacco informatico può essere classificato come un attacco armato? Da qui derivano altri temi rilevanti, come il diritto alla legittima difesa e la proporzionalità della risposta. Questi argomenti sono ampiamente dibattuti nel diritto internazionale, con giuristi impegnati ad adattare il quadro normativo alle peculiarità del dominio cyber.

Il diritto intrinseco all'autodifesa, come delineato dall'Articolo 51, può essere invocato per giustificare azioni difensive solo in seguito a un attacco armato riconosciuto. Tuttavia, applicare questo principio alle operazioni cyber introduce diverse questioni. Identificare quando un incidente informatico costituisce un attacco armato e determinare la proporzionalità della risposta è una sfida in un contesto tecnologico in continua evoluzione.

I giuristi sono impegnati in dibattiti continui per chiarire e adattare l'applicazione del diritto internazionale ai conflitti cyber. Alcune delle questioni centrali riguardano la definizione di attacco armato nel cyberspazio, l'attribuzione<sup>10</sup> di un attacco cyber a soggetti statali o non statali, l'ambito e la scala delle misure difensive appropriate. Il cyberspazio sfida i paradigmi giuridici tradizionali, rendendo necessaria l'elaborazione di norme e regolamenti specifici per questo nuovo dominio.

Cerchiamo di approfondire la prospettiva legata alla risposta ad attacchi cibernetici al di sotto della soglia di gravità di un attacco armato. In questo scenario ipotetico, si pone l'interrogativo fondamentale relativo alla legittimità della reazione a tali tipi di attacco. In linea generale, la risposta tende a essere negativa, poiché il divieto dell'uso della forza – sia cinetica che cibernetica – rimane un principio cardine del diritto internazionale.

<sup>10</sup> M. Lee, *Attribution*.

Tuttavia, esiste una sfumatura rilevante, legata al ruolo centrale dei ‘dati’ coinvolti negli attacchi. Molti attacchi informatici non distruggono infrastrutture fisiche, ma compromettono la sicurezza dei dati, come dimostrato da celebri episodi di cosiddetti “data leak” e furto di informazioni sensibili. Questa realtà impone una riflessione sul ruolo dei dati nel quadro normativo internazionale.

A questo proposito, possiamo fare riferimento alle Convenzioni di Ginevra<sup>11</sup>, che rappresentano un pilastro del diritto umanitario internazionale. Esse stabiliscono che le operazioni militari devono essere condotte evitando il più possibile danni ai civili e alle loro proprietà. Ma i dati dei civili possono essere considerati alla stregua di proprietà intese nel senso classico del termine? La risposta potrebbe essere affermativa. I dati personali sono beni intangibili, ma strettamente legati alla vita e alla privacy degli individui. Proteggere i dati dei civili si allinea ai principi umanitari delle Convenzioni di Ginevra, che mirano a salvaguardare i civili e i loro beni durante i conflitti. Questa prospettiva apre un dibattito etico e giuridico sulla necessità di un quadro normativo specifico per la protezione dei dati nel contesto dei conflitti cibernetici.

### 2.3. Attribuzione

La questione dell’attribuzione nelle operazioni cyber rappresenta una delle sfide più complesse e dibattute<sup>12</sup>. La stessa natura di Internet, che consente la falsificazione degli indirizzi sorgente, complica enormemente l’identificazione certa di chi abbia lanciato un attacco.

<sup>11</sup> The Geneva Conventions, *International Committee of the Red Cross*.

<sup>12</sup> M. Lee, *Attribution*; S. Goel - B. Nussbaum, *Attribution across cyber attack types: Network intrusions and information operations*.

Nei sistemi di rete, il fatto che un pacchetto di dati provenga da un indirizzo specifico non costituisce una prova definitiva della sua reale origine. Questo problema è ulteriormente accentuato dal fenomeno dello *spoofing*<sup>13</sup>, una tecnica con cui gli attaccanti mascherano intenzionalmente l'indirizzo sorgente per nascondere la loro identità.

Un punto cruciale è stato evidenziato nel rapporto del 2015 degli Esperti Governativi delle Nazioni Unite<sup>14</sup>, il quale afferma che la mera associazione di un'operazione cyber all'infrastruttura di uno stato non è sufficiente per attribuire con certezza la responsabilità dell'attacco a quel governo. Esiste infatti la possibilità che le operazioni cyber siano condotte da attori non statali o da Stati che utilizzano intermediari e infrastrutture di altri paesi per sviare le indagini.

Questa complessità nell'attribuzione pone una sfida critica nel contesto normativo internazionale. Lo sviluppo di standard condivisi e procedure unificate per l'attribuzione potrebbe contribuire a mitigare questo problema e promuovere una maggiore responsabilità degli attori coinvolti in operazioni dannose nel cyberspazio.

Nell'ambito della sicurezza informatica, personalità di spicco come Bradford Smith, presidente di Microsoft, hanno evidenziato le lacune degli attuali strumenti normativi, in particolare delle Convenzioni di Ginevra, nell'affrontare le specificità della guerra digitale. Nel suo libro *Tools and Weapons: The Promise and the Peril of Our Digital Future*<sup>15</sup>, Smith affronta queste problematiche e propone l'istituzione di una Convenzione di Ginevra Digitale, un trattato spe-

<sup>13</sup> A. Hilton - J. Hirschmann - C. Deccio, *Beware of ips in sheep's clothing: Measurement and disclosure of ip spoofing vulnerabilities.*

<sup>14</sup> United Nations, *Group of governmental experts on developments in the field of information and telecommunications in the context of international security.*

<sup>15</sup> B. Smith - C. A. Browne, *Tools and Weapons: The Promise and the Peril of Our Digital Future.*

cifico per regolamentare i conflitti nel cyberspazio. Uno degli aspetti centrali della proposta di Smith è la necessità per i governi di adottare un approccio più trasparente. Questo significa non solo riconoscere pubblicamente l'esistenza di vulnerabilità nei sistemi informatici, ma anche impegnarsi nella condivisione di informazioni con altri Stati, riducendo così il rischio di escalation e migliorando la sicurezza informatica globale. Un modello di collaborazione internazionale potrebbe rafforzare la resilienza collettiva contro le minacce cyber.

L'idea di una Convenzione di Ginevra Digitale riflette una realtà emergente: il confine tra guerra tradizionale e guerra informatica si sta sempre più assottigliando. Con l'avanzare delle tecnologie, gli accordi internazionali devono necessariamente evolversi per rispondere in modo adeguato alle nuove minacce digitali.

Tuttavia, la proposta di Smith evidenzia anche le difficoltà nell'ottenere un consenso globale sulle regole di ingaggio nel cyberspazio. La comunità internazionale si trova a dover definire quali azioni informatiche costituiscono un atto ostile e quali soglie di gravità giustificano una risposta difensiva. I compiti menzionati richiedono che si raggiunga il consenso anche su come attribuire con certezza la responsabilità di un attacco. In sintesi, il dibattito avviato da Bradford Smith mette in luce l'urgenza di stabilire nuove norme e accordi internazionali, capaci di affrontare le sfide poste dai conflitti digitali. La sua visione contribuisce a una crescente consapevolezza sul fatto che il diritto internazionale deve adattarsi rapidamente per rispondere alla natura dinamica delle minacce informatiche.

### *3. La guerra informatica prolungata tra Russia e Ucraina*

Nel panorama in continua evoluzione dei conflitti geopolitici, il cyberspazio è diventato un campo di battaglia cruciale. Un esempio emble-

matico è la guerra informatica prolungata tra Russia e Ucraina, un conflitto che precede le recenti notizie e continua a plasmare la dinamica della sicurezza informatica regionale e globale. Questo contributo esplora il contesto storico, gli eventi principali e le tattiche in evoluzione all'interno della complessa narrazione della guerra cibernetica tra queste due nazioni.

Le radici della guerra informatica tra Russia e Ucraina risalgono agli anni '80 e '90, quando la regione faceva parte dell'Unione Sovietica. Nel corso degli anni, i progressi tecnologici e la crescente dipendenza dalla connettività digitale hanno aperto nuovi fronti di conflitto, dando vita a una serie crescente di attacchi informatici e contromisure nel cyberspazio.

### 3.1. Un focus sugli eventi significativi: una panoramica storica della guerra informatica russo-ucraina

Esplorando gli eventi cruciali all'interno della vasta cronologia della guerra informatica tra Russia e Ucraina, alcuni episodi emergono come paradigmatici, offrendo spunti sull'evoluzione di questo conflitto digitale. Questi eventi illustrano collettivamente l'evoluzione della guerra informatica russo-ucraina, mostrando il passaggio da attività di spionaggio iniziali a operazioni più sofisticate e politicamente motivate. L'interazione complessa delle campagne informatiche sponsorizzate dallo Stato, lo spionaggio e gli attacchi ai sistemi critici sottolineano le sfide continue nel trattare e mitigare l'impatto delle minacce informatiche su scala geopolitica.

#### 3.1.1. Operazione UroBuros – 2005

Operazione UroBuros, condotta dalla Russia nel 2005, ha segnato un primo ingresso nella guerra informatica. Questa opera-

zione è stata caratterizzata dall'impiego di un sofisticato strumento informatico mirato allo spionaggio, con gli obiettivi principali che includevano gli Stati Uniti e l'Ucraina. L'uso di strumenti informatici per spiare e pedinare ha fissato un precedente per le future attività nel contesto in evoluzione del conflitto digitale. Nonostante la natura segreta delle operazioni informatiche, UroBuros ha rivelato l'impatto potenziale delle attività informatiche sponsorizzate dallo Stato sulla sicurezza informatica globale.

### 3.1.2. Operazione Snake – 2014

Avanzando temporalmente fino al periodo compreso tra il 2014 e il 2017, l'Operazione Snake emerge come una campagna informatica significativa attribuita alla Russia. Questa operazione ha ottenuto il sostegno politico, mettendo in evidenza l'intersezione tra la sicurezza informatica e le iniziative sponsorizzate dallo Stato. Eseguita da professionisti esperti di sicurezza informatica, l'Operazione Snake si è concentrata sullo spionaggio informatico, dimostrando l'impegno della Russia nell'uso dei mezzi digitali per raccogliere informazioni. La complessità dell'attribuzione informatica sottolinea la necessità di misure di sicurezza informatica robuste su scala globale.

### 3.1.3. Attacco al sistema elettorale – 2014

Nel 2014, il sistema elettorale ucraino è stato vittima di un attacco informatico, aggiungendo una nuova dimensione al conflitto informatico russo-ucraino. Mirando ai processi democratici di una nazione, questo assalto informatico ha evidenziato le potenziali vulnerabilità all'interno dei sistemi critici. L'incidente ha messo in luce le implicazioni più ampie della guerra informatica sulla stabilità politica e ha sollevato preoccupazioni sull'integrità dei sistemi eletto-

rali. Man mano che le nazioni si affidano sempre più alle infrastrutture digitali per i processi democratici, proteggersi contro le minacce informatiche diventa fondamentale per mantenere l'integrità delle istituzioni democratiche.

Il gruppo dietro questo attacco è un'entità ben nota, identificata con l'acronimo Cyber Berkut. Utilizzando tecniche di infiltrazione nell'infrastruttura di voto elettronico della Commissione Elettorale Ucraina, Cyber Berkut ha interrotto con successo una funzionalità critica del sistema, ossia il componente di monitoraggio e analisi delle attività. Questo evento si è rivelato fondamentale, non solo causando una crisi nel processo elettorale ucraino, ma minando anche gravemente la fiducia globale nei sistemi di voto elettronico. Sfruttando le vulnerabilità e compromettendo componenti essenziali del meccanismo elettorale, Cyber Berkut ha evidenziato la suscettibilità delle infrastrutture digitali ad attività informatiche dannose, mettendo in dubbio l'affidabilità e la sicurezza dei sistemi di voto elettronico a livello mondiale.

#### 3.1.4. Black Energy – 2015

L'attacco Black Energy alla rete elettrica ucraina, avvenuto nel dicembre 2015, ha rappresentato una pietra miliare nel campo della guerra informatica a causa del suo impatto diretto sulle infrastrutture critiche. Il gruppo Cyber Berkut, noto per i suoi sentimenti filorusi, è stato identificato come l'autore dell'attacco. Come già discusso, utilizzando sofisticate tecniche di infiltrazione, il gruppo aveva mirato alla Commissione Elettorale Centrale ucraina durante le elezioni locali del paese. Questa campagna informatica ha avuto implicazioni più ampie rispetto alla semplice interruzione del processo elettorale, poiché ha preparato il terreno per un successivo attacco alle infrastrutture energetiche del paese. Infatti, sfruttando il

successo iniziale, Cyber Berkut ha eseguito un attacco diretto alla rete elettrica ucraina. L'attacco ha causato un blackout, colpendo migliaia di famiglie e imprese ucraine. Ciò che ha reso questo incidente particolarmente rilevante è stato il fatto di aver segnato il primo caso ufficialmente riconosciuto di un attacco informatico di successo che ha causato una significativa interruzione dell'approvvigionamento energetico di una nazione.

L'attacco Black Energy ha messo in evidenza le potenziali ramificazioni geopolitiche della guerra informatica. Ha rivelato la capacità di gruppi sponsorizzati dallo Stato o ad esso affiliati di mirare a infrastrutture critiche, sollevando preoccupazioni sulla vulnerabilità dei servizi essenziali alle minacce informatiche. L'incidente ha anche sottolineato l'uso delle capacità informatiche come strumenti di influenza e coercizione politica.

In risposta a queste minacce informatiche, l'Ucraina ha adottato misure proattive per migliorare le sue capacità di sicurezza informatica. Il paese ha riconosciuto la necessità di rafforzare le proprie infrastrutture critiche contro attacchi futuri, investendo in strategie di sicurezza informatica e collaborando con partner internazionali per condividere informazioni sulle minacce.

L'attacco Black Energy serve come caso di studio nel panorama in evoluzione della guerra informatica, sottolineando le conseguenze reali e la necessità per le nazioni di rafforzare le loro difese informatiche contro minacce sempre più sofisticate ai servizi essenziali.

### 3.1.5. Operazione Prikormka (Groundbait) – 2016

A seguito degli attacchi informatici russi, l'Ucraina ha risposto sviluppando attivamente le sue capacità di sicurezza informatica. Una delle operazioni più significative in questo contesto è "Prikormka", una campagna informatica intrigante condotta sul suolo ucraino.

L'operazione è di particolare interesse non solo per gli aspetti tecnici, ma anche per l'interpretazione sfumata associata alla sua esecuzione.

Prikormka è stata progettata come un'operazione di spionaggio, focalizzandosi sull'estrazione di informazioni critiche relative ai movimenti e alle attività dei gruppi separatisti. La campagna informatica ha coinvolto tecniche sofisticate per infiltrarsi nelle reti associate a questi gruppi. La specifica focalizzazione su Donetsk e Luhansk, regioni caratterizzate da sentimenti separatisti e indipendenza autoproclamata, ha riflesso l'intento strategico dell'operazione. Comprendere la dinamica di queste aree era cruciale per le considerazioni sulla sicurezza dell'Ucraina. Oltre alla tradizionale raccolta di informazioni, Prikormka ha integrato elementi di analisi comportamentale. Gli operatori informatici miravano a comprendere i modelli comportamentali, le interazioni sociali e le dinamiche politiche all'interno delle comunità separatiste oggetto di interesse.

L'attribuzione nelle operazioni informatiche è spesso complessa, e Prikormka non è stata un'eccezione. La sensibilità politica legata ai movimenti separatisti ha aggiunto ulteriori elementi di complessità agli sforzi di attribuzione. Gli esperti di sicurezza informatica hanno incontrato difficoltà nell'attribuire, in modo inequivocabile, l'operazione a specifici attori, a causa dell'uso di tattiche come *false flag*<sup>16</sup> o tecniche che preservano l'anonimato. È chiaro, però, che con l'operazione Prikormka l'Ucraina ha dimostrato un approccio proattivo alla sicurezza informatica. L'attacco esemplifica questa posizione proattiva, in cui le capacità offensive informatiche sono state strategicamente impiegate per scopi difensivi di raccolta di informazioni. L'operazione ha mostrato l'adattabilità dell'Ucraina alle sfide geopolitiche. Integrando

<sup>16</sup> F. Kopik - T. Pahi, *Under false flag: using technical artifacts for cyber attack attribution*.

le capacità informatiche nella propria strategia di difesa, l'Ucraina ha cercato di ottenere un vantaggio tecnologico nella comprensione e nel contrasto delle potenziali minacce all'interno dei suoi confini.

Prikormka sottolinea anche l'intreccio complesso tra sicurezza informatica e geopolitica. L'operazione non riguardava solo la competenza tecnica, ma anche l'uso delle capacità informatiche per affrontare preoccupazioni politiche e di sicurezza in una regione segnata da tensioni geopolitiche in corso. In sostanza, serve come un caso di studio che evidenzia l'uso delle capacità informatiche da parte dell'Ucraina per affrontare sfide geopolitiche complesse. Il successo dell'operazione si è basato su una combinazione di competenza tecnica, raccolta di informazioni e una comprensione approfondita del panorama socio-politico nelle regioni target.

### 3.1.6. Operazione Nove Maggio – 2016

L'operazione "Nove Maggio" rappresenta un atto dimostrativo di patriottismo informatico ucraino. Si tratta di un episodio significativo di dimostrazione del potere cibernetico da parte dell'Ucraina contro i separatisti, avvenuto proprio nel Giorno della Vittoria, una data di particolare rilevanza storica per i russi. Lanciato il 9 maggio, l'attacco è carico di simbolismo, poiché coincide con il giorno del 1945 in cui i nazisti si arresero all'Unione Sovietica. La scelta di questa data per l'azione dimostrativa evidenzia l'intenzione degli ucraini di impiegare la dimensione simbolica nella loro risposta.

L'operazione si è concretizzata attraverso il "defacement" di siti web, una tecnica che altera l'aspetto visivo di una pagina web. Gli obiettivi selezionati erano siti gestiti dai separatisti ucraini. Tra i gruppi coinvolti spiccano Falcons Flame e Trinity, che in seguito si sono fusi per formare un unico gruppo di hacker patriottici ucraini. La collaborazione tra Falcons Flame e Trinity, culminata nella

fusione, sottolinea l'unità degli sforzi informatici all'interno della comunità degli hacker patriottici ucraini. Questa unione potrebbe riflettere una strategia volta a consolidare risorse e competenze per affrontare minacce informatiche specifiche.

L'operazione Nove Maggio va oltre la dimensione tecnica, trasmettendo un chiaro messaggio politico di patriottismo informatico. Gli hacker coinvolti, agendo in risposta a una giornata di celebrazione russa, mirano a difendere gli interessi nazionali dell'Ucraina anche nel cyberspazio. Quest'operazione evidenzia come le azioni nel dominio cibernetico siano spesso cariche di significato politico. Ancora una volta, questi eventi dimostrano che la guerra informatica non è solo un conflitto tecnologico, ma anche un terreno in cui si manifestano tensioni geopolitiche. In conclusione, l'operazione Nove Maggio si inserisce nella più ampia narrativa della guerra informatica russo-ucraina, sottolineando l'importanza di considerare non solo gli aspetti tecnici, ma anche il contesto politico e simbolico in cui tali attacchi avvengono.

### 3.1.7. Operazione Channel One – 2016

L'operazione "Channel One" si distingue come una pietra miliare nel conflitto informatico tra Russia e Ucraina, segnando la prima intrusione nelle trasmissioni televisive. Questo attacco è stato orchestrato dai gruppi Falcons Flame e Trinity, ormai fusi sotto il nome di RUH8 (Russian Hackers Hate U 8), in collaborazione con altri hacktivisti. L'unione di queste forze ha permesso di combinare risorse, conoscenze e competenze, aumentando l'impatto delle loro azioni sul fronte cibernetico.

Channel One è diventata celebre per aver infiltrato i sistemi di trasmissione, segnando un cambiamento significativo rispetto alle tradizionali operazioni informatiche. Questa intrusione ha stabilito

un precedente, influenzando direttamente la diffusione delle informazioni attraverso i canali mediatici mainstream. Gli sforzi collaborativi non si sono limitati a una singola operazione, ma hanno portato alla creazione di RUH8 come gruppo consolidato, includendo hacktivisti<sup>17</sup> di diversa provenienza, a testimonianza di un fronte unito nella risposta informatica dell'Ucraina.

In sintesi, l'operazione Channel One, orchestrata da RUH8, evidenzia l'evoluzione delle strategie informatiche nel conflitto russo-ucraino. L'intrusione nelle trasmissioni rappresenta un distacco dalle attività informatiche convenzionali, sottolineando l'adattabilità e l'innovazione dei gruppi hacktivisti nel perseguire i propri obiettivi.

### 3.1.8. Surkov Leaks – 2016

Una delle operazioni informatiche più rilevanti condotte da attori ucraini riguarda la diffusione di dati sensibili, nello specifico i Surkov Leaks. Il nome deriva da Vladislav Surkov, uno dei principali consiglieri di Vladimir Putin, spesso definito il “Cardinale Grigio” del Cremlino. Questa operazione si è concentrata sull'esfiltrazione e la divulgazione pubblica di dati riservati relativi alla gestione strategica delle entità filorusse a Donetsk, in Ucraina. Come detto, Surkov, figura di spicco del Cremlino e stretto consigliere di Putin, è stato l'obiettivo principale di questa azione informatica. Mirando a lui, gli attori ucraini hanno cercato di svelare i piani strategici e i processi decisionali legati alla leadership filorussa nella regione.

Donetsk, situata nell'Ucraina orientale, riveste un'importanza strategica nel conflitto in corso. I dati divulgati probabilmente inclu-

<sup>17</sup> Il termine “hactivismo” indica la fusione tra hacking e attivismo, riferendosi all'uso di tecniche informatiche per promuovere cause politiche o sociali.

devano informazioni sul sostegno del Cremlino alle fazioni filorusse, facendo luce sulle loro attività e piani operativi. L'essenza dei Surkov Leaks risiede nell'estrazione di informazioni riservate attraverso mezzi informatici, con successiva pubblicazione per compromettere l'influenza e le strategie delle forze filorusse a Donetsk.

Le implicazioni geopolitiche dell'operazione sono significative, poiché si tratta di un'azione volta a smascherare attività segrete orchestrate da alti funzionari del Cremlino. L'obiettivo era rivelare l'entità del coinvolgimento russo nel conflitto e potenzialmente influenzare la percezione internazionale. Questa operazione rappresenta un esempio dell'approccio proattivo dell'Ucraina nel dominio informatico, colpendo figure chiave per portare alla luce informazioni strategiche. Inoltre, dimostra come le operazioni cyber possano giocare un ruolo cruciale nei conflitti geopolitici, modellando le narrazioni e influenzando l'opinione pubblica.

### 3.1.9. Secondo attacco informatico alla rete elettrica ucraina, 2015-2016

In una successiva offensiva informatica contro l'Ucraina, attori russi hanno orchestrato un secondo attacco alla rete elettrica del Paese, provocando danni significativi e un blackout prolungato fino a sei ore. Inoltre, l'azione cyber ha esteso il suo impatto anche al Dipartimento del Tesoro ucraino, causando uno stato di paralisi operativa. Sponsorizzato da entità russe, questo attacco ha rappresentato un'ulteriore escalation nelle tensioni del conflitto cibernetico in corso tra le due nazioni. Il blackout ha colpito duramente diversi settori e ha sconvolto la vita quotidiana dei cittadini ucraini.

Ampliando la portata dell'attacco, gli attori russi hanno preso di mira anche il Dipartimento del Tesoro ucraino, con l'intento di bloccare le operazioni finanziarie e aumentare la pressione sul go-

verno. L'attribuzione dell'attacco ha inizialmente sollevato dibattiti e incertezze sull'identità dei responsabili. Tuttavia, con il progredire delle indagini, il legame con la Russia è emerso con sempre maggiore evidenza. Oggi, vi è un livello più elevato di certezza sull'origine dell'attacco, attribuendolo con fermezza a entità cyber russe.

Il secondo attacco alla rete elettrica e il sabotaggio del Dipartimento del Tesoro hanno segnato un'ulteriore escalation nel conflitto informatico tra Russia e Ucraina. Questi episodi evidenziano l'uso strategico delle capacità cyber per influenzare e destabilizzare settori critici. Mentre l'Ucraina affrontava un'altra grave offensiva informatica, la comunità internazionale osservava con crescente preoccupazione l'intensificarsi delle attività cibernetiche nel contesto delle più ampie tensioni geopolitiche. Gli eventi hanno messo in luce la vulnerabilità delle infrastrutture critiche a sofisticati attacchi informatici e l'urgente necessità per le nazioni di rafforzare le proprie difese cyber.

### 3.1.10. Petya Ransomware – 2017

Negli ultimi anni, il famigerato malware Petya ha attirato un'attenzione diffusa come un potente esempio di ransomware. Questo software malevolo rappresenta uno dei più noti casi di malware in grado di crittografare i dati delle vittime per poi chiedere un riscatto in cambio della loro decrittazione. Petya utilizza avanzate tecniche di crittografia per bloccare l'accesso ai file e ai sistemi delle vittime. Una volta completata l'operazione, l'utente non è più in grado di accedere ai propri dati. Successivamente, il malware invia una richiesta di riscatto agli individui o alle organizzazioni colpite, solitamente esigendo il pagamento in criptovaluta per garantire l'anonimato delle transazioni.

Petya sfrutta diverse tecniche di propagazione per diffondersi tra sistemi e reti. Questi metodi includono spesso lo sfruttamento di vulnerabilità software o l'uso di tattiche di ingegneria sociale per

indurre gli utenti a eseguire file malevoli. Nel corso del tempo, Petya ha subito diverse evoluzioni, ciascuna con miglioramenti nei meccanismi di infezione e crittografia, rendendo difficile per gli esperti di cybersecurity sviluppare contromisure efficaci. Questo ransomware è stato protagonista di attacchi informatici di alto profilo, colpendo sia individui che organizzazioni, con la capacità di interrompere sistemi critici e richiedere ingenti riscatti, diventando così uno strumento potente negli arsenali dei cybercriminali.

L'attribuzione dell'attacco rimane complessa. Nonostante Petya sia generalmente associato ad attività cybercriminali, le caratteristiche dell'attacco del 2017 sollevarono sospetti su un possibile coinvolgimento di attori sponsorizzati da uno Stato. Il governo ucraino accusò la Russia, che tuttavia negò ogni coinvolgimento. La difficoltà nell'attribuzione deriva dall'uso di false flag e dalla partecipazione di attori sofisticati. Inoltre, la natura distruttiva dell'attacco indicava la presenza di un componente di cosiddetto "wiper malware", progettato per danneggiare irreversibilmente i sistemi anziché consentire il recupero dei dati dopo il pagamento del riscatto. Questo si allinea con le strategie della guerra cibernetica, il cui scopo è massimizzare il danno e minare la resilienza della nazione bersaglio.

Oltre ai danni immediati, l'attacco Petya ebbe pesanti ripercussioni economiche. Le aziende subirono perdite finanziarie a causa dei tempi di inattività, della perdita di dati e dei costi di recupero. L'ampio impatto su infrastrutture critiche e sistemi governativi sollevò preoccupazioni sulle potenziali conseguenze della guerra cibernetica sulla stabilità di una nazione.

L'attacco Petya/NotPetya rappresenta un punto di svolta nell'escalation della guerra informatica, superando le tradizionali operazioni di spionaggio o il cybercrimine a fini economici. Ha dimostrato come attori statali possano impiegare malware distruttivi, sfumando il confine tra criminalità informatica e guerra cibernetica.

Il coinvolgimento di un ransomware in questo conflitto cyber tra Russia e Ucraina ha evidenziato l'evoluzione e la crescente sofisticazione delle minacce informatiche nei contesti geopolitici. L'episodio ha sottolineato la necessità di norme e regolamenti internazionali per affrontare le sfide crescenti poste dalle operazioni cyber sponsorizzate dagli Stati.

### 3.2. Il Cyber Peace Institute

Per analisi approfondite e aggiornate sui conflitti informatici tra Russia e Ucraina tra il 2022 e la data corrente, il Cyber Peace Institute<sup>18</sup> rappresenta una risorsa preziosa. L'Istituto conduce un esame dettagliato degli attacchi cyber, offrendo una cronologia ben organizzata con un focus specifico sulle infrastrutture critiche.

Il Cyber Peace Institute si distingue per la pubblicazione di aggiornamenti quasi quotidiani sulle attività informatiche tra Russia e Ucraina. Questa continuità garantisce informazioni aggiornate e riflette l'evoluzione rapida del panorama cibernetico. Un aspetto centrale dell'analisi riguarda gli attacchi alle infrastrutture critiche, tra cui, ma non solo, le reti energetiche, i sistemi di trasporto e le reti di comunicazione. Questo approccio evidenzia il potenziale impatto sociale di tali attacchi.

Il concetto di "oggetti civili" viene approfondito per evidenziare le ripercussioni sociali più ampie degli attacchi informatici. Analizzando il modo in cui questi attacchi influenzano i servizi essenziali e la vita quotidiana dei civili, l'Istituto va oltre gli aspetti puramente tecnici, sottolineando le dimensioni umane e sociali del conflitto cy-

<sup>18</sup> <https://cyberconflicts.cyberpeaceinstitute.org/>

ber. La cronologia fornisce informazioni dettagliate su ogni attacco, comprese le tecniche utilizzate, la natura dei sistemi colpiti e le caratteristiche peculiari di ogni offensiva. Questa panoramica completa aiuta a comprendere il *modus operandi* degli attori delle minacce.

L'Istituto fornisce anche analisi sugli impatti concreti degli attacchi informatici, tra cui interruzioni di servizi, conseguenze economiche e altri effetti tangibili sulla società. Comprendere queste conseguenze è fondamentale per valutare la gravità degli attacchi.

L'attribuzione degli attacchi rappresenta una sfida complessa in ambito di cybersecurity. Il Cyber Peace Institute contribuisce a questo processo facendo riferimento alle fonti degli attacchi, aiutando così a identificare gli attori coinvolti, le loro motivazioni e le loro affiliazioni. Per chi ha una maggiore inclinazione tecnica, la cronologia include dettagli sulle metodologie impiegate dai cyber attaccanti, come tipologie di malware, tattiche, tecniche e procedure (TTPs). Queste informazioni consentono una comprensione più approfondita delle campagne cibernetiche in corso.

Nel complesso, l'analisi del Cyber Peace Institute rappresenta una risorsa essenziale per decisori politici, professionisti della cybersecurity e il pubblico generale, fornendo spunti preziosi sulle dinamiche e sulle implicazioni dei conflitti informatici tra Russia e Ucraina. La combinazione di dettagli tecnici, impatto sociale e aggiornamenti costanti rende questa piattaforma un archivio completo e informativo.

#### *4. Una guerra di Hacker*

Il panorama della guerra cibernetica, in particolare nel contesto di conflitti come quello tra Russia e Ucraina, è caratterizzato da una complessa interazione tra attori statali, gruppi di hacktivisti e

attori individuali. Questa complessità deriva dalla natura fluida delle alleanze, delle motivazioni e delle strategie all'interno del dominio digitale.

La Russia è nota per impegnarsi in operazioni cibernetiche come parte della sua strategia militare più ampia. I gruppi di hacker sponsorizzati dallo Stato, ritenuti operare con il supporto o almeno con l'approvazione tacita del governo russo, sono stati coinvolti in varie attività cibernetiche, tra cui spionaggio, campagne di disinformazione e attacchi di disturbo.

In risposta alle minacce digitali, l'Ucraina ha sviluppato le proprie capacità cibernetiche, sia difensive che offensive. Il conflitto ha spinto il governo ucraino a investire in misure di cybersicurezza per proteggere le infrastrutture critiche e le informazioni sensibili.

#### 4.1. Gruppi di hacktivist

Il gruppo di hacking Conti, inizialmente associato a sentimenti filorussi, ha affrontato dissidi interni e si è frazionato. Dall'altra parte, il collettivo di hacktivist Anonymous si è allineato con la causa ucraina. Questi gruppi, inizialmente motivati da ideologie geopolitiche, hanno dimostrato la fluidità delle alleanze all'interno del paesaggio cibernetico.

Anonymous, in particolare, ha ampliato i suoi obiettivi oltre gli avversari geopolitici. Ha iniziato a prendere di mira le aziende occidentali che collaborano con la Russia, evidenziando l'estensione dei conflitti cibernetici nei domini economici e aziendali. Questo cambiamento sottolinea la natura sempre più interconnessa della guerra cibernetica, che trascende i confini tradizionali.

Il dominio cibernetico consente anche il coinvolgimento di hacker indipendenti che potrebbero non essere direttamente affiliati

a uno Stato o a un gruppo organizzato. Questi attori contribuiscono a incrementare la complessità del panorama complessivo, operando sulla base di motivazioni personali, ideologie politiche o incentivi finanziari.

#### 4.1.1. Anonymous

Anonymous è emerso nel 2003 dai forum di imageboard come *4chan*<sup>19</sup>. Inizialmente, il gruppo era caratterizzato da utenti anonimi che condividevano liberamente idee e immagini. Nel tempo, questo spirito anarchico si è trasformato in azioni collettive. Anonymous ha condotto numerose operazioni di hacktivism per esporre informazioni, sensibilizzare su questioni sociali o politiche e protestare contro organizzazioni o governi. Gli attacchi sono stati rivolti a siti web governativi, organizzazioni di estrema destra e istituzioni ritenute corrotte. Il gruppo opera senza una struttura gerarchica formale. I partecipanti, noti come “Anons”, collaborano online per pianificare ed eseguire azioni. Questo modello decentralizzato rende difficile identificare un leader o un portavoce ufficiale.

La maschera iconica di Guy Fawkes è diventata l'emblema distintivo di Anonymous, ispirata dal romanzo grafico di Alan Moore “V for Vendetta”, dove il protagonista usa la maschera per simboleggiare la resistenza contro un governo totalitario. Anonymous è un fenomeno dinamico che riflette la complessità dell'era digitale. L'hacktivism, quando guidato da intenzioni positive, può essere uno strumento di responsabilità e difesa dei diritti. Tuttavia, la linea tra hacktivism etico e attività illegali può essere sfumata, alimentando dibattiti sulla sua validità come strumento di protesta digitale.

<sup>19</sup> <https://www.4chan.org/>

Comprendere Anonymous e l'hacktivismo richiede una valutazione critica delle azioni specifiche e delle intenzioni dietro di esse.

#### 4.2. Le sfide di attribuzione e la guerra economica

L'attribuzione rimane una sfida significativa nello spazio cibernetico. L'uso di 'false bandiere', l'anonimizzazione degli attacchi e la capacità di manipolare le impronte digitali rendono difficile attribuire in modo definitivo le azioni cibernetiche a entità specifiche. Questo è particolarmente vero quando il conflitto diventa anche ciò che potrebbe essere definito come una *Guerra Economica*. L'impiego di aziende occidentali come obiettivi, da parte di gruppi di hacktivist come Anonymous, introduce effettivamente una dimensione economica nei conflitti cibernetic. Attaccando le imprese che collaborano con i presunti avversari, questi gruppi mirano a imporre conseguenze che vanno al di là dell'arena geopolitica. La pubblicazione dei loghi delle aziende e l'emissione di ultimatum dimostrano come le azioni degli hacktivist possano influenzare il comportamento aziendale. La gogna pubblica e le minacce di esposizione dei dati esfiltrati creano pressioni sulle imprese affinché riconsiderino le loro affiliazioni di fronte ai conflitti cibernetic.

In sintesi, la dinamica della guerra cibernetica tra Russia e Ucraina mostra le complesse interazioni tra attori statali, gruppi di hacktivist e hacker individuali. Le motivazioni, le alleanze e le tattiche impiegate sottolineano la natura in continua evoluzione dei conflitti nel dominio digitale, dove le tensioni geopolitiche si estendono nelle sfere economiche e sociali.

## 5. *L'etica dell'Hacker*

Il dibattito etico sull'hacktivismo è complesso. Mentre molte azioni mirano a esporre la corruzione e difendere la libertà di espressione, l'approccio di gruppi come Anonymous è talvolta controverso poiché può comportare danni collaterali. Ecco perché è cruciale comprendere gli aspetti multiformi della cosiddetta *cultura dell'hacker*, con particolare riferimento alle sue implicazioni etiche.

Le radici della cultura hacker possono essere fatte risalire al Massachusetts Institute of Technology (MIT) negli anni '50. Al MIT, il termine "hack" indicava una soluzione ingegnosa e non convenzionale a un problema complesso. Gli studenti si impegnavano in attività collaborative e creative per risolvere problemi, sviluppando una cultura attorno all'affrontare le sfide in modi creativi. L'ethos dell'hacker incarna un insieme di principi che includono:

- *Creatività*: gli hacker apprezzano la risoluzione creativa dei problemi. Affrontano le sfide con una mentalità inventiva e non convenzionale.
- *Provocazione*: c'è una disponibilità a sfidare lo status quo e mettere in discussione le norme stabilite. Questo spirito provocatorio mira a stimolare l'innovazione.
- *Realizzazione di gesti significativi*: gli hacker spesso cercano di realizzare contributi che lascino un segno tangibile nell'evoluzione del panorama tecnologico.

L'evoluzione dell'immagine pubblica dell'hacker ha preso una piega negativa negli anni '90, quando i mass media hanno iniziato ad associare l'hacking ad attività malevole. Il termine "hacker" è diventato comunemente usato per descrivere individui che si impegnano nel realizzare accessi non autorizzati alle risorse informatiche, violazioni di dati e altri crimini informatici. Questa rappresentazio-

ne negativa ha portato alla diffusione della convinzione che tutti gli hacker siano intrinsecamente malevoli.

Per distinguere gli hacker malintenzionati da quelli che usano le loro capacità in modo etico, è emerso il termine “hacker etico”. Un hacker etico, o “white hat”, è qualcuno che utilizza le proprie abilità di hacking per identificare e correggere vulnerabilità nei sistemi informatici, nelle reti o nelle applicazioni. Gli hacker etici lavorano per rafforzare la cybersicurezza, proteggere i beni digitali e contribuire positivamente all’ecosistema tecnologico. In sostanza, la cultura hacker incoraggia gli individui a considerare l’hacking come un mezzo per il cambiamento positivo e l’espressione creativa. L’hacking etico, in particolare, promuove l’uso responsabile e costruttivo delle abilità di hacking per migliorare la cybersicurezza e la resilienza tecnologica. È un invito a riconoscere che l’hacking, nella sua forma originale, è uno strumento che può essere utilizzato per il bene, enfatizzando l’innovazione, la creatività e la ricerca della conoscenza.

### 5.1. Cappelli e colori

I termini “white hat”, “black hat” e “gray hat” nel contesto dell’hacking si ispirano ai classici film western, dove i protagonisti e gli antagonisti sono spesso simboleggiati da personaggi che indossano cappelli bianchi o neri. Questi termini sono stati adottati nel dominio della cybersicurezza per classificare gli hacker in base alle loro intenzioni e azioni. Gli hacker “white hat” indossano il cappello bianco, simboleggiando gli eroi o i buoni nei film western. Sono hacker etici che usano le loro capacità per scopi benevoli. Lavorano per identificare vulnerabilità, proteggere i sistemi e rafforzare le difese di cybersicurezza. Il loro obiettivo è proteggere reti, applicazioni e dati da accessi non autorizzati e minacce informatiche. D’altra parte,

gli hacker “black hat” indossano il cappello nero, simboleggiando i cattivi o gli avversari. Si impegnano in attività malevole, come accessi non autorizzati, violazioni di dati e crimini informatici. Le loro azioni sono generalmente motivate da guadagni personali, motivi finanziari o altre intenzioni malevole. Sui “gray hat hackers” si trovano le sfumature più ambigue, operando questi ultimi con intenzioni che potrebbero non essere del tutto malevole, ma che comunque coinvolgono attività non autorizzate. Un gray hat hacker potrebbe, ad esempio, attaccare sistemi senza autorizzazione, ma affermare di farlo per ragioni etiche o per svelare l’esistenza di vulnerabilità. L’etica delle sue azioni può essere soggettiva e dipende dalle circostanze specifiche.

L’uso di questi termini con i colori semplifica la categorizzazione degli hacker in base alle loro motivazioni e azioni. Fornisce un modo rapido per capire la posizione etica di un hacker, specialmente nei contesti di cybersicurezza dove distinguere tra intenzioni benevole e malevole è cruciale. I termini sono diventati standard nel settore e aiutano a comunicare più efficacemente la natura delle attività di hacking.

## 5.2. Letteratura Hacker

L’esplorazione dell’etica hacker e dei conflitti cibernetici sottolinea la necessità di un coinvolgimento responsabile ed etico nel dominio dell’hacking. Il richiamo a un comportamento etico enfatizza l’impatto delle azioni virtuali sul mondo tangibile, esortando gli individui ad affrontare la cybersicurezza con un senso di responsabilità. Man mano che il paesaggio digitale evolve, le considerazioni etiche diventano fondamentali per salvaguardare l’integrità e la sicurezza dei sistemi globali. L’imperativo “gli hacker dovrebbero essere

giudicati in base alle loro specifiche attività di hacking” sottolinea la motivazione intrinseca dell’hacker nel risolvere le sfide in modo creativo. Nonostante le rappresentazioni mediatiche, sia la letteratura convenzionale che quella non convenzionale testimoniano che gli hacker, nella loro forma più pura, perseguono la conoscenza, difendono la libertà e sostengono i sistemi decentralizzati.

Steven Levy, in *Hackers: Heroes of the Computer Revolution*<sup>20</sup>, offre un’opera fondamentale che esplora la cultura hacker, offrendo un’analisi completa degli individui che hanno giocato un ruolo fondamentale nei primi giorni dell’informatica. Grazie a racconti di prima mano e interviste, Levy fornisce preziose intuizioni sulla mentalità unica degli hacker di questo periodo trasformativo nella storia dei computer. Il lavoro di Levy non si limita a narrare le storie delle figure iconiche, ma cattura anche l’ethos e lo spirito che hanno definito la comunità hacker. Il libro porta i lettori nei laboratori informatici del MIT e oltre, mostrando la passione, la curiosità e la natura ribelle che caratterizzavano questi primi hacker. Enfatizza il loro impegno nell’esplorazione, nella risoluzione dei problemi e nella ricerca della conoscenza, sfidando spesso le norme e l’autorità precostituite. Raccontando le avventure di hacker come Bill Gates, Steve Wozniak e altri protagonisti, Levy dipinge un quadro vivido di una subcultura che ha contribuito in modo significativo allo sviluppo dell’informatica. “Hackers” resta una risorsa preziosa per chiunque sia interessato a comprendere le radici della cultura hacker e il suo impatto sul paesaggio tecnologico.

Il dottor Crash, in un articolo per *Phrack Magazine*<sup>21</sup>, discute la rivoluzione tecnologica e definisce l’hacking in un’accezione di

<sup>20</sup> S. Levy, *Hackers: Heroes of the Computer Revolution*.

<sup>21</sup> D. R. Crash, *The Techno Revolution*.

creatività. L'articolo fornisce uno sguardo contemporaneo sull'intersezione tra hacking e tecnologia, inquadrandolo nel contesto di una "Rivoluzione Tecnologica". Questo termine cattura il paesaggio in evoluzione della tecnologia e chiarisce come l'hacking sia posizionato all'interno di questo ambiente dinamico. Traendo ispirazione dall'ethos dell'hacking, il Dottor Crash rappresenta l'hacking come un'impresa creativa. Questa interpretazione si allinea con l'idea che l'hacking non è solo un atto distruttivo, ma anche una forma di esplorazione, innovazione e sforzo per trovare soluzioni ingegnose ai problemi tecnologici.

L'opera di Pekka Himanen, *The Hacker Ethic: A Radical Approach to the Philosophy of Business*<sup>22</sup>, rappresenta un contributo significativo all'analisi dei principi etici che definiscono la cultura hacker. L'approccio di Himanen è di natura filosofica, con l'obiettivo di chiarire i principi fondamentali che guidano gli hacker nel loro modo di lavorare, nella creatività e nell'impatto sociale.

Il libro pone particolare enfasi su tre principi chiave che caratterizzano l'etica hacker: creatività, passione e impatto sociale. Questi valori sfidano le norme convenzionali e le concezioni tradizionali dell'etica del lavoro e degli affari. Per gli hacker, la creatività non è semplicemente una competenza, ma un modo di affrontare problemi e sfide. Nel libro, l'hacking è descritto come un approccio innovativo che implica il pensare fuori dagli schemi, trovare soluzioni non convenzionali e spingersi oltre i limiti attraverso il pensiero creativo.

L'etica hacker, secondo Himanen, è alimentata dalla passione. Gli hacker sono mossi da un profondo entusiasmo per il loro lavoro, e questa passione rappresenta la forza trainante del loro impegno. Questo contrasta con le tradizionali etiche del lavoro, che spesso

<sup>22</sup> P. Himanen, *The Hacker Ethic: A Radical Approach to the Philosophy of Business*.

enfaticamente la routine e la conformità. Infine, Himanen sostiene che gli hacker siano intrinsecamente attenti all'impatto del loro lavoro sulla società. Questo implica un senso di responsabilità e una consapevolezza di come le innovazioni tecnologiche possano modellare il mondo. L'etica hacker, quindi, non si limita a obiettivi personali, ma considera anche le implicazioni sociali delle attività di hacking.

Il libro fornisce un'analisi approfondita di questi principi ed esplora le loro implicazioni nel mondo degli affari e nella società. Sfida le concezioni preesistenti sull'etica del lavoro e propone una rivalutazione radicale del modo in cui individui e organizzazioni possono affrontare la creatività, il lavoro guidato dalla passione e le loro responsabilità sociali più ampie. In sintesi, *The Hacker Ethic* è un'esplorazione filosofica che mette in luce i principi fondamentali che guidano gli hacker e il loro potenziale impatto trasformativo nel mondo del business e nella società.

*Hacker Manifesto*<sup>23</sup>, scritto da The Mentor (Loyd Blankenship) nel 1986, è un documento fondamentale che incarna l'essenza dell'etica hacker di quell'epoca. Questo manifesto offre preziose intuizioni sulle motivazioni, frustrazioni e valori che animavano gli hacker del tempo. Il manifesto rappresenta una sorta di istantanea dello spirito e dell'etica della sottocultura hacker. Esso articola la mentalità e i principi che spingevano gli hacker nella loro ricerca di conoscenza ed esplorazione nel mondo dei sistemi informatici. Uno dei passaggi più celebri del manifesto ("noi esploriamo... e voi ci chiamate criminali"), mette in evidenza la percezione, da parte degli hacker, di essere ingiustamente etichettati dalla società. Questo passaggio sottolinea la tensione tra il desiderio di conoscenza degli hacker e le norme sociali che spesso vedono le loro attività come illecite. Pur esprimendo

<sup>23</sup> The Mentor (Loyd Blankenship), *The Hacker Manifesto*.

un certo livello di sfida e ribellione, il testo suggerisce anche una sorta di richiesta di comprensione e di riconoscimento delle reali motivazioni degli hacker. Scritto da un membro attivo della comunità hacker, il manifesto offre una testimonianza diretta e autentica della prospettiva hacker, permettendo ai lettori di comprendere dall'interno il loro pensiero e le loro esperienze. Il documento sfida le norme e le aspettative sociali nei confronti della tecnologia e dell'hacking. Mette in discussione l'ordine costituito, suggerendo che gli hacker siano guidati da valori e obiettivi diversi rispetto a quelli compresi dalla società mainstream. Inoltre, evidenzia la diversità di motivazioni all'interno della comunità hacker, riconoscendo che non si tratta di un gruppo monolitico, ma di individui con ragioni differenti per intraprendere attività di hacking. Un tema centrale nel manifesto è il conflitto percepito tra hacker e norme convenzionali. Il testo sottolinea il divario tra la ricerca di conoscenza degli hacker e la percezione esterna, che spesso li identifica erroneamente come criminali.

I recenti conflitti cyber dimostrano come essi possano affiancarsi alla guerra convenzionale, con conseguenze reali e tangibili. Poiché la tecnologia è sempre più integrata nella vita quotidiana, l'impatto dei cyber conflitti si estende oltre il dominio virtuale. Comprendere e mitigare gli effetti dei cyber conflitti è cruciale per la stabilità globale. In risposta all'evoluzione delle minacce informatiche, è emerso l'hacking etico<sup>24</sup> (o *penetration testing*) come meccanismo di difesa proattivo. Come detto, gli hacker etici conducono attività di hacking autorizzate e responsabili per individuare vulnerabilità nei sistemi e nelle reti. Il loro obiettivo è quello di rafforzare la sicurezza informatica, contribuendo alla resilienza delle infrastrutture digitali.

<sup>24</sup> M. Bishop, *Ethical Hacking*.

### *Conclusioni*

L'esplorazione dell'interazione complessa tra guerra cibernetica, hacking ed etica hacker rivela un paesaggio dinamico plasmato da tensioni geopolitiche, progressi tecnologici e considerazioni etiche. Analizzando gli aspetti principali di questo dominio dalle molteplici sfaccettature, emergono diversi spunti chiave. La nostra analisi sottolinea l'evoluzione della guerra cibernetica come un'arena fondamentale per i conflitti moderni. L'integrazione fluida dell'impegno digitale con la guerra tradizionale richiede una comprensione sfumata delle operazioni cibernetiche sponsorizzate dagli Stati, che comprendono spionaggio, disturbo e campagne di influenza. L'assenza di quadri giuridici ben definiti e universalmente accettati nello spazio cibernetico presenta sfide significative. La complessità nell'attribuire gli attacchi cibernetici e nel definire norme per un comportamento responsabile degli Stati sottolinea la necessità di una collaborazione internazionale per orientarsi in questo terreno complesso.

Una prospettiva storica sul panorama della guerra cibernetica russo-ucraina rivela decenni di operazioni cibernetiche intrecciate con eventi geopolitici. Incidenti come gli attacchi alle infrastrutture critiche e le interferenze nelle elezioni evidenziano la natura multidimensionale dei conflitti cibernetici in questa regione. In relazione a ciò, abbiamo discusso come organizzazioni come il Cyber Peace Institute svolgano un ruolo fondamentale nel documentare e analizzare i conflitti cibernetici. La loro enfasi su un comportamento responsabile degli stati nello spazio cibernetico contribuisce ad aumentare la consapevolezza e la trasparenza, promuovendo una comprensione globale delle minacce cibernetiche e del loro impatto sociale.

La nostra esplorazione abbraccia uno spettro ampio di attività di hacking, dallo spionaggio cibernetico sponsorizzato dallo Stato

all'hacktivismo guidato da obiettivi ideologici. L'inclusione di attori statali e non statali introduce un elemento dinamico con motivazioni e metodi distinti. In questo scenario, l'etica hacker, fondata sulla ricerca della conoscenza, della creatività e della decentralizzazione, forma le basi di una subcultura con legami storici con la curiosità accademica. L'emergere dell'hacking etico, o hacking white-hat, rappresenta un impegno verso un uso responsabile e legale delle competenze di cybersicurezza. Esaminare opere come il Manifesto dell'Hacker fornisce intuizioni storiche sulle motivazioni e le frustrazioni degli hacker, sfidando le norme e le percezioni sociali. I dilemmi etici sorgono dall'equilibrio delicato tra la ricerca della conoscenza, l'adesione all'etica hacker e le potenziali ramificazioni sociali delle azioni cibernetiche.

In conclusione, il nostro viaggio attraverso i domini della guerra cibernetica, dell'hacking e dell'etica hacker illumina sfide complesse e considerazioni etiche. Navigare in questo paesaggio richiede un'integrazione armoniosa di innovazione tecnologica, realtà geopolitiche e un impegno fermo verso una condotta responsabile ed etica nel dominio digitale. Man mano che tracciamo la nostra rotta verso il futuro, lo sviluppo di quadri giuridici chiari e linee guida etiche sarà imperativo per garantire un futuro digitale sicuro ed equo.

## Bibliografia

- N. Anderson - C. Kolas - G. Orosz - G. Pék, *An analysis of cyber threats in the ukrainian conflict*, in «IEEE Security & Privacy», 16, 6, 2018, pp. 64-71.
- M. Bishop, *Ethical Hacking*, in «ACM SIGCSE Bulletin», 33, 1, 2001, pp. 145-152.
- Dr. Crash, *The Techno Revolution*, in «Phrack», 1, 6, 1986, <<http://phrack.org/issues/6/3.html>>.
- S. Goel - B. Nussbaum, *Attribution across cyber attack types: Network intrusions and information operations*, in «IEEE. Open Journal of the Com-

- munications Society», 2, 2021, pp. 1082-1093, <<https://doi.org/10.1109/OJCOMS.2021.3074591>>.
- A. Hilton - J. Hirschmann - C. Deccio, *Beware of ips in sheep's clothing: Measurement and disclosure of ip spoofing vulnerabilities*, in «IEEE/ACM Transactions on Networking», 30, 4, 2022, pp. 1659-1673, <<https://doi.org/10.1109/TNET.2022.3149011>>.
- P. Himanen, *The Hacker Ethic: A Radical Approach to the Philosophy of Business*, New York, Random House, 2001.
- International Cyber Norms: Legal, Policy & Industry Perspectives*, edited by A.-M. Osula and H. Rõigas, Berlin, Springer, 2021.
- F. Kopik - T. Pahi, *Under false flag: using technical artifacts for cyber attack attribution*, in «Cybersecurity», 3, 8, 2020. <<https://doi.org/10.1186/s42400-020-00048-4>>.
- M. Lee, *Attribution*, in *Cyber Threat Intelligence*, Hoboken, Wiley, 2023, pp. 155-174, <<https://doi.org/10.1002/9781119861775.ch6>>.
- S. Levy, *Hackers: Heroes of the Computer Revolution*, Sebastopol, O'Reilly Media, 2010 (1<sup>st</sup> edition 1984).
- M. C. Libicki, *Cyberwar: The next threat to national security and what to do about it*, in «Naval War College Review», 65, 1, 2012, pp. 53-78.
- A. Mishra - Y. I. Alzoubi - M. J. Anwar - A. Q. Gill, *Attributes impacting cyber-security policy development: An evidence from seven nations*, in «Computers Security», 120, 2022, <<https://doi.org/10.1016/j.cose.2022.102820>>.
- R. Nguyen, *Navigating "jus ad bellum" in the age of cyber warfare*, in «California Law Review», 101, 4, 2013, pp. 1079-1129.
- C. Rohith - R. S. Batth, *Cyber Warfare: Nations Cyber Conflicts, Cyber Cold War Between Nations and its Repercussion*, in *2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*, 2019, pp. 640-645, <<https://doi.org/10.1109/ICCIKE47802.2019.9004236>>.
- M. N. Schmitt, *Tallinn manual on the international law applicable to cyber warfare*, Cambridge, Cambridge University Press, 2017.
- B. Smith - C. A. Browne, *Tools and Weapons: The Promise and the Peril of Our Digital Future*, New York, Penguin Press, 2019.
- The Geneva Conventions*, International Committee of the Red Cross, Geneva, 1949.
- The Mentor (Loyd Blankenship), *The Hacker Manifesto*, in «Phrack», 1, 7, 1986, <<http://phrack.org/issues/7/3.html>>.
- United Nations, *Charter of the United Nations*, 1945.
- United Nations, *Group of governmental experts on developments in the field of information and telecommunications in the context of international security*,

*Guerra cibernetica e frontiere etiche*

2015, <<https://documents-dds-ny.un.org/doc/UNDOC/GEN/N15/228/35/PDF/N1522835.pdf>>.

M. C. Waxman, *Self-defensive force against cyber attacks: Legal, strategic and political dimensions*, in «Yale Journal of International Law», 38, 2, 2013, pp. 1-61.

Simon Pietro Romano  
Università degli Studi di Napoli Federico II  
spromano@unina.it  
<https://orcid.org/0000-0002-5876-0382>



FRANCESCO ROMEO

*Intelligenza Artificiale, povertà dell'etica,  
inutilità del giurista*

*Abstract:* Information technology and artificial intelligence (AI) produce culture in a way that is, sometimes, independent of human intervention or its possibility of control. The artificial origin of new cultural products raises problems and questions, provokes conflicts and rejections. In the field of ethics, a cultural gap has grown that is currently being filled by studies financed by private actors, the so-called Ethics Washing, with serious consequences on the reliability of the results. These lead to the current discredit for ethics studies in general, cd. Ethics Bashing, and the denial of the possibility or usefulness of AI ethics. The reaction of scholars generally misunderstands how the AI systems work. The conflict, at least in the legal field, is in part useful in highlighting some misconceptions about law and can help to reframe important institutes of legal theory such as, in particular, the theory of judicial interpretation and application, but also the theory of norm. It is now clear that, on the one hand, machine learning is able to offer new tools for the resolution of social conflicts and, on the other hand, it serves as a spotlight on the misrepresentations of the law enforcement process that doctrine has developed over time. The role of biases in reaching the judicial decision is examined. It is argued that what can reasonably already be required from an AI system is to assist the jurist in the interpretation-application of the law, subject to the control of the reasonableness of the results. It is further argued that in the more recent past, AI systems have begun to insert themselves into the means of production and control of legal rules and their application, in an increasingly ubiquitous manner that is difficult to eliminate or modify. Again, a merely procedural control is not sufficient and perhaps not even possible, however, more in-depth legal reflection is needed than what jurists have been able to offer so far. Finally, it is argued that a correct simulation of the judge's or interpreter's reasoning in an AI system can help to scientifically understand legal reasoning as such and to devise new tools for resolving disputes, or to predict the outcome of disputes, without, however, representing a

simulation of the reasoning and decision itself. Never before it has been possible to test the accuracy or adequacy of many theories about law, and already we can see that many of them are unable to keep up with innovation.

*Keywords:* Ethics Washing, Ethics Bashing, Artificial Intelligence and Cultural Production, Artificial Intelligence and Judgment, Artificial Intelligence and Law Enforcement, Artificial Intelligence and Law Interpretation.

### *Introduzione*

Qualsiasi asserto si intenda avanzare, in tema di Intelligenza Artificiale (IA), esso non può non accettare la premessa che l'informatica e la IA stanno profondamente cambiando la nostra quotidianità. Per quanto le si possa avversare, ne siamo oramai bisognosi e dipendenti. Una innumerevole quantità di prodotti culturali a queste dedicati o da queste immessi nel flusso culturale ci avvolge ormai quotidianamente, senza essere, dalla maggior parte delle persone, compresi nel loro funzionamento o significato. Anche in ambito giuridico, i prodotti culturali non mancano, la normativa in merito è divenuta oramai sovrabbondante. Persino per l'attento e prudente giurista, tradizionalmente diffidente all'innovazione, l'uso di questi sistemi artificiali (SA) è divenuto irrinunciabile: si pensi, ad esempio, alle indagini penali, o anche alla redazione di atti giuridici, o alla consultazione di repertori e massimari. Ma il punto nodale, quello che individua una differenza che cambia totalmente le regole del gioco fin qui valide nella socialità umana, è che questi sistemi producono cultura in modo, alle volte, indipendente dall'intervento umano o dalla sua possibilità di controllo. Una volta era solo l'essere umano a farlo, nessun sostituto o succedaneo dei prodotti culturali umani poteva essere individuato o anche solo pensato credibilmente: oggi i

prodotti culturali dei sistemi artificiali lo sono, almeno di fatto. In linea teorica si possono tenere diverse posizioni, riconoscere o meno creatività e originalità a questi prodotti: si può strenuamente negare doti di intelligenza o di pensiero in senso stretto ai SA, sono tutte questioni di contorno, direi bizantinismi, rispetto al fatto: i prodotti sono lì, da noi utilizzati quotidianamente, i SA producono cultura. È inevitabile che questo incontro di culture diverse, quanto all'origine, ponga problemi e domande, susciti conflitti e respingimenti.

### 1. *Povertà dell'etica*

Da ultimo, il conflitto è sorto nella discussione filosofica, o anche parafilosofica, sui temi dell'etica. Si noti il ritardo: i principali problemi che sarebbero sorti, in questo campo, si erano delineati da più di due decenni, ma la ricerca li ha del tutto disattesi, ritenuti poco credibili o verosimili o irrealizzabili *tout-court*<sup>1</sup>. Non si tratta di una manchevolezza di poco conto. La ricerca applicata, nelle nostre economie, non è patrimonio esclusivo delle università e degli istituti di ricerca pubblici con finanziamenti altrettanto pubblici. Altri soggetti, obbedienti alla logica del profitto ne sono protagonisti e questi non hanno mancato l'appuntamento, sicché, del ritardo etico-poli-

<sup>1</sup> Rilevavo già nel 2002 la lacuna della ricerca nel libro F. Romeo, *Il diritto artificiale*, pp. 155 ss. Su alcuni aspetti vedi già V. Frosini, *L'uomo artificiale*, l'orizzonte del cielo degli albori era già popolato delle nubi che oggi ci minacciano. La ricerca etica rispose all'epoca, appunto, con bizantinismi, tra l'altro sempre gli stessi di oggi come, ad esempio «questi sistemi non potranno mai dire qualcosa che non sia stato programmato», oppure «non riusciranno mai a riprodurre l'attività intellettuale umana», oppure ancora «sono solo macchine», e via dicendo in un florilegio di banalità.

tico-culturale attuale, eticisti e filosofi del diritto debbono rimproverare loro stessi, in primo luogo. Gli altri hanno svolto il loro ruolo in modo impeccabilmente efficiente. Attualmente gli scritti di etica sui SA sono abbondanti, le spiegazioni sul loro funzionamento sono ancora in numero maggiore, ma non v'è concordanza di vedute. Resta un *mood* antagonista, nei confronti di questi sistemi, che forse offusca in parte la realtà, ma, soprattutto, la pericolosità della svolta socioculturale che ci troviamo a percorrere.

Nel campo etico, va rilevato che della materia se ne sono appropriati proprio gli attori privati interessati a quello che è stato chiamato *Ethics-Washing*<sup>2</sup>, l'uso dell'etica come belletto o trucco nascondente o giustificante l'azione o l'uso dei SA. È un'attività fiorente in un campo in cui la quantità di denaro circolante è veramente fuori misura<sup>3</sup>; non nascondiamoci che molti filosofi e colleghi si sono concessi alle richieste: improvvisamente l'etica – campo non incolto ma usualmente assai morigerato – ha conosciuto – e conosce – fasti e ricchezza; la discussione scientifica ne è uscita stravolta e banalizzata, ingigantendo l'avversione di alcuni più seri ricercatori per i SA, chiamati a rispondere di colpe che, più appropriatamente, andrebbero attribuite al '*Sapiens* dietro le macchine'. Gli scritti di etica dell'IA, che insieme agli altri scritti etici avrebbero dovuto riempire scaffali di polverose e voluminose opere colte, oggi sono vendutissimi libelli nelle luccicanti vetrine della belletristica. I contenuti si appaiano alla forma. L'acrimonioso spaesamento del filosofo è palese e i SA vengono giudicati sotto uno sguardo e un *bias* peggiorativo che poco aiuta a recuperare la dignità filosofica:

<sup>2</sup> E. Bietti, *From Ethics Washing to Ethics Bashing*, pp. 1 ss. La qualificazione risale a B. Wagner, *Ethics as an Escape from Regulation*, pp. 84 ss.

<sup>3</sup> Vedi in particolare D. Tafani, *L'«etica» come specchio per le allodole*, pp. 7 ss.

Giudizi e decisioni che hanno effetti rilevanti sulle vite di esseri umani sono oggi affidati, in un numero crescente di ambiti, a sistemi di intelligenza artificiale che non funzionano. In settori quali quello giudiziario, dei servizi finanziari, dell'educazione, dei servizi sociali o del reclutamento del personale, l'uso di sistemi di apprendimento automatico (*machine learning*) nei processi di valutazione e decisione ha dato luogo a esiti ingiusti, nocivi e assurdi – come documenta una letteratura ormai sterminata – con conseguenze che si riverberano a lungo, talora per anni, sulle vite delle vittime. Tali malfunzionamenti non sono occasionali e non sono scongiurabili con interventi tecnici: essi rivelano, anzi, il funzionamento ordinario dei sistemi di apprendimento automatico, utilizzati impropriamente per compiti che non è loro possibile svolgere o che sono impossibili *tout court*. Considerato il ruolo cruciale di tali sistemi nel modello di business delle grandi aziende tecnologiche e gli enormi profitti che queste ne ricavano, non sorprende che, paventando un divieto generalizzato, esse mirino a sottrarre tali prodotti all'intervento giuridico: è nata così, con l'obiettivo di rendere plausibile un regime di mera autoregolazione, l'«etica dell'intelligenza artificiale»<sup>4</sup>.

Gli asserti meritano un forse o un approfondimento, perché l'evidenza conduce a tutt'altre conclusioni. L'autrice ammette la sempre maggiore diffusione dei sistemi di intelligenza artificiale, l'aggiunta, quindi «che non funzionano» appare ingiustificata, così come è posta, anzi, sembrerebbe vero il contrario, si diffondono sempre di più perché finalmente funzionano e se si discute di alcuni aspetti ritenuti indesiderabili, in genere lo si fa per trovare soluzioni che non siano quella, oramai generalmente abbandonata, del rifiuto di questi sistemi. Nel confronto tra danni derivanti da errori o malfunzionamenti, da una parte, e vantaggi apportati dall'uso di questi sistemi, dall'altra, nessuno dubita più della scelta: sono insostituibili. L'autrice nota la crescente letteratura sui danni apportati e sugli «esiti ingiusti, nocivi e assurdi», perdipiù in settori delicati quale quello giudiziario,

<sup>4</sup> Ivi, p. 2.

ad esempio. Anche qui, calcando emozionalmente la parola l'autrice esprime la sua opinione sul sospetto che gli attuali studi sull'etica dell'IA siano favoriti dalle imprese che operano nel settore, anzi che esse stesse ne abbiano cagionato o, quantomeno, favorito la nascita: «è nata così, con l'obiettivo di rendere plausibile un regime di mera autoregolazione»<sup>5</sup>.

L'autrice osserva correttamente il distorto uso autoregolativo dell'etica dell'intelligenza artificiale, è evidente, però, il *bias* emozionale che la porta a ribaltare le caratteristiche reali dei SA. Così vengono attribuite esagerate doti negative, oppure trasformate in negative caratteristiche in grazia delle quali i SA subsimbolici (*machine learning e deep learning*) eseguono egregiamente compiti che i sistemi artificiali simbolici, quali i sistemi esperti, non riescono ad affrontare. La caratteristica principale di questi sistemi è l'adattamento tramite un procedimento di minimizzazione dell'errore. Sono sistemi che migliorano sbagliando e questo li rende sospetti, perché suscettibili di provocare danni, «con conseguenze che si riverberano a lungo, talora per anni, sulle vite delle vittime»<sup>6</sup>. Si pensi all'errore di un SA alla guida di un veicolo a guida autonoma, ad esempio. Che l'errore possa accadere è fuor di dubbio, ma forse non accade anche con l'essere umano? Non è la possibilità dell'errore ad essere rilevante, ma la probabilità dello stesso. Il giudizio sulla utilità o correttezza di un sistema neurale artificiale deve valutare la probabilità del ripetersi dell'errore in confronto alla stessa probabilità in un essere umano.

In altri casi il rilievo negativo nei confronti dei SA diviene positivo se è l'essere umano a utilizzare lo stesso tipo di calcolo o ragionamento. I sistemi di *machine e deep learning* sono progettati sul presup-

<sup>5</sup> *Ibid.*

<sup>6</sup> *Ibid.*

posto della somiglianza strutturale con il cervello umano. Riescono ad operare analogie, induzioni e generalizzazioni che sono modelli di ragionamento essenziali nei procedimenti euristici e che l'essere umano usa abbondantemente. Questi tipi di argomenti logico-statistici o probabilistici non permettono una risposta in termini di vero/falso ma in termini solo percentuali o probabilistici di verità. Così è per l'essere umano e così è per questi sistemi: sono gli strumenti matematici usati che non possono offrire certezza logica, ma sono anche insostituibili nei nostri e nei loro ragionamenti. L'analogia, o ragionamento di somiglianza o argomento a simili, ha un ruolo di primaria importanza nel ragionamento giuridico, non può essere eliminata, ma non può essere annoverata tra i ragionamenti logici deduttivi, non offre certezze ma probabilità. Ogni giurista ragiona per analogie e il risultato del suo ragionamento, chiamato oggi interpretazione o applicazione della legge, è, per molti aspetti, solo probabilisticamente prevedibile e non deduttivo. Nel caso di un interprete-uomo si accetta la probabilità o indeterminatezza della risposta, nel caso di un interprete-SA no. Perché? Perché si chiede ai SA quella certezza e ripetibilità che non solo non sono richiesti a un giudice o avvocato, ma non sarebbero neppure accettati? Evidentemente si guarda a questi sistemi non come mera replicazione dell'operare umano, ma come miglioramento o risoluzione di problemi sin qui irrisolti.

La risposta quindi all'osservazione della Tafani è che errare e adattare il proprio comportamento al fine di minimizzare l'errore è un metodo di apprendimento che si è mostrato utile nella storia della cultura umana, ha permesso al *Sapiens* di comportarsi adeguatamente in un determinato ambiente, massimizzando la propria *fitness*. Lo stesso metodo permette ad alcuni tipi di SA di apprendere e automigliorarsi, questi sono da giudicare utili nei casi in cui le loro percentuali di errore siano inferiori a quelle di esseri umani istruiti per quegli stessi compiti.

La risposta permette di proporre fondatamente un tipo di regolazione di questi sistemi. Essa, infatti, non può non tener conto dell'analisi costi/benefici proveniente dalla comparazione delle percentuali di errore. Se un veicolo a guida autonoma erra meno di un guidatore umano, provocando meno incidenti, non dovrebbero esserci dubbi sulla sua introduzione e sulla sostituzione. Anche qui, il dibattito, spesso emozionalmente guidato, perde di vista il vero problema, che giace non nei sistemi ma nella interazione uomo-sistemi artificiali. Fino a questo punto della sua storia evolutiva, il *Sapiens* ha appreso a prevedere il comportamento di altri *Sapiens* in svariati modi. Questa capacità di previsione è fondamentale per la sua sopravvivenza, per la sua socialità e per il suo successo evolutivo come specie. Le possibilità di previsioni corrette si basano anche sulla somiglianza biologica tra i vari *Sapiens*. Su questa possibilità si è evoluta la nostra cultura. Lo stesso però non vale nei riguardi dei SA. Né gli strumenti di calcolo, né le strutture, né i presupposti biologici sono pienamente condivisi con noi: la prevedibilità, per un *Sapiens*, dell'azione di un SA può quindi essere inferiore, rispetto alla prevedibilità dell'azione di un altro *Sapiens*. Questo è senz'altro un problema in un ambiente socio-tecnico, nel quale sistemi cognitivi naturali ed artificiali interagiscono. Questa interazione soggiace a stimoli diversi che agiscono su strutture diverse. Anche un fattore elementare, come la velocità di reazione, può diventare problematico. Questo è un capitolo della cultura umana, e non solo, tutto da scrivere ex novo<sup>7</sup>.

Un ulteriore rilievo, ripetuto da più parti, riguarda la non controllabilità del ragionamento condotto per arrivare alla conclusione. In questa critica rientrano osservazioni quali il rilievo dato a pregiudizi e *bias*, cattive classificazioni, impossibilità di seguire una procedura

<sup>7</sup> G. Contissa, *Automation and liability*, pp. 17-45.

argomentativa tradizionale, intervento di procedure stocastiche nella minimizzazione dell'errore. Il rilievo riguarda i sistemi subsimbolici, non quelli simbolici, per i quali, invece, l'intera procedura argomentativa è conosciuta. La risposta a questo insieme di importanti rilievi viene dalla constatazione che neppure i processi del pensiero umano sono conosciuti. Come un giudice giunga alla conclusione è cosa diversa da come lo stesso giudice argomenta la propria decisione. L'argomentazione è una fase giustificativa che segue la decisione, ciò che, invece, la precede è largamente sconosciuto. Nella decisione di un giudice non sono presenti solo necessità realizzative dettate da norme dell'ordinamento, ma anche regole di senso comune o buon senso, apparati valoriali personali, preferenze, credenze personali e opinioni: tutti concorrono alla decisione. La decisione di un giudice, ma in verità ogni decisione, è un procedimento del pensiero complesso e tuttora risulta inspiegato nella sua pienezza. La vera 'scatola nera' non sono i SA, anche quelli subsimbolici, ma gli esseri umani. Qui gli esempi sono numerosissimi. È il caso delle decisioni cd. discriminatorie, ad esempio. Dipendentemente dagli esempi offerti al SA in fase di apprendimento, il sistema apprenderà a classificare e quindi distinguere e nominare porzioni di realtà. Il ruolo degli esempi di apprendimento è cruciale. Se in essi sono presenti discriminazioni, il SA le apprenderà. Ma non solo, l'*output* non dipende solo dai dati forniti nella fase di apprendimento. C'è un rilievo più generale e molto più oscuro e imprevedibile nei suoi risvolti. Teoricamente non ci sono limiti alle funzioni matematiche e agli algoritmi utilizzabili per l'apprendimento di una rete neurale. Questa loro versatilità le rende molto performanti, molto più di un essere umano. Possono ragionare con strumenti a noi ignoti e giungere quindi a conclusioni alle quali noi potremmo non arrivare mai. Contemporaneamente, però, questi ragionamenti e queste conclusioni si distaccano da una visione umana (o antropocentrica) del mondo. Le risposte dei SA possono diventare a noi incomprensibili,

pur partendo da premesse impostate da noi e pur essendo vantaggiose nei risultati<sup>8</sup>. Questo è un punto critico di grande rilievo, perché qui il rischio risiede nella perdita del controllo sulla formazione della cultura, dovendoci limitare a un controllo dei risultati. D'altronde, richiedere un continuo controllo procedurale sarebbe una ineffettiva perdita di tempo. Laddove pur si riuscisse a realizzarla, l'unico controllo praticabile e di una qualche efficacia sarebbe quello sui risultati, non il calcolo, che resta fuori dalla nostra portata. Un controllo, quindi, di adeguatezza o ragionevolezza, non di verità rispetto all'impianto argomentativo del SA. Occorre anche notare, che è falsa l'asserzione secondo la quale una rete neurale non argomenta la conclusione. L'argomentazione c'è, ma è non traducibile in termini di linguaggio ordinario. La procedura di minimizzazione dell'errore è l'argomentazione, o meglio, l'euristica. Una argomentazione in termini linguistici è però solo ricostruibile *ex post*, sui risultati già calcolati, quindi non una euristica ma una giustificazione. Questo contrasta con il modello di giudice ideale che parte della dottrina ha costruito, ma non contrasta con una prassi assolutamente monocolore: anche nei giudici la giustificazione della soluzione non corrisponde mai al ragionamento in base al quale si è raggiunta la decisione.

## 2. *Inutilità del giurista*

In ambito giuridico è oramai chiaro che il *machine learning* da un lato è in grado di offrire strumenti nuovi per la risoluzione dei conflitti

<sup>8</sup> Di particolare interesse è il libro di G. Longo, *Matematica e senso*, pp. 32 ss. e *passim*, nel quale viene approfondita, da un punto di vista dei linguaggi matematici, la distanza tra la formazione del significato in noi e nei SA.

sociali e dall'altro funge da evidenziatore delle false rappresentazioni del procedimento di applicazione della legge che nel tempo la dottrina ha elaborato. Queste rappresentazioni si tramandano accademicamente, spesso basandosi più su di un principio di autorità che di effettiva corrispondenza a verità. Porto come esempio l'annoso dibattito sui pregiudizi del giudice. Il modello continentale, figlio della rivoluzione borghese, esige un giudice imparziale esecutore della legge, non fonte del diritto ma applicatore delle norme esistenti nell'ordinamento, frutto della volontà popolare rappresentata nel Parlamento, non della sua volontà. Un'applicazione del diritto che non rispetti questo assioma sarebbe contraddittoria e quindi errata. Ma la rappresentazione è sbagliata, si riferisce a un dover essere, non a un essere, è quindi una direttiva posta al giudice per la risoluzione di controversie. Come tale essa soggiace a tutti i limiti usuali di un sistema cognitivo, quello umano in questo caso, che deve trasformare un insieme di parole in significati e questi in decisioni. Giace in questa attività il nocciolo del problema, giacché i significati soggettivi sono difficilmente separabili dagli aspetti emozionali e valoriali del soggetto stesso. Nella contemporaneità il problema dei diversi sistemi valoriali soggettivi è esploso proprio grazie a questi sistemi, laddove la previsione della decisione di un giudice veniva appoggiata sulla profilazione del giudice stesso più che sulla perfetta descrizione del caso<sup>9</sup>. Conoscendo le opinioni politiche e anche le relazioni sociali e affettive del giudice e sapendo in che relazione stanno con i comportamenti da giudicare, alcuni SA riuscivano a effettuare proiezioni affidabili sulla decisione che quel giudice avrebbe preso, pur senza argomentare in punto di diritto. Insomma se conosco il decisore fin negli *interna corporis* saprò ben prevedere che risposte darà. La risposta del legislatore francese ha vietato la pubbli-

<sup>9</sup> V. A. Laptev - D. R. Feyzrakhmanova, *Application of Artificial Intelligence in Justice*.

cazione del nome del giudice in calce alle sentenze, di fatto impedendo la corretta profilazione dello stesso sui suoi precedenti. Insomma, non nego il dogma teorico, elimino i dati che permettono di negarlo: imbarazzante<sup>10</sup>.

Le reti neurali e il *machine learning* hanno reso i pregiudizi computabili. Tramite esse è possibile simulare quella parte del giudizio umano che non si basa su scelte razionali. Sorprendentemente, in questo modo è possibile simulare le decisioni della giurisprudenza, in parte, con risultati soddisfacenti, pur senza poter ricostruire la stessa argomentazione che offre la sentenza di un giudice, comunque con risultati migliori rispetto alla IA classica. Ma, in ambito giuridico, come si giustifica un giudizio soggettivo o un pregiudizio? Che cosa significa dire che un giudizio soggettivo è corretto o è sbagliato, che è vero o è falso?

Un giudizio soggettivo o un pregiudizio non si costruiscono mai come universali ma, al massimo, secondo un criterio di regolarità, di ciò che per lo più accade, dell'opinione comune e questo criterio è verificabile solo a posteriori e può sempre cambiare ed essere sostituito da un altro, così come ogni opinione cerca con il consenso di catturare le altre. Così alle reti e ai loro risultati non è possibile attribuire un valore di verità se non in senso probabilistico, si può parlare di esperimento robusto, di rete robusta, ma non si può verificare secondo valori di verità il risultato.

Nella nostra cultura giuridica è radicata l'idea che il diritto, ma soprattutto la giustizia, ricostruiscono verità, si fatica a cancellare

<sup>10</sup> Articolo 33 Loi n. 2019-222 du 23 mars 2019 de programmation 2018-2022 et de réforme pour la justice. “Les données d'identité des magistrats et des membres du greffe ne peuvent faire l'objet d'une réutilisation ayant pour objet ou pour effet d'évaluer, d'analyser, de comparer ou de prédire leurs pratiques professionnelles réelles ou supposées”. Vedi anche il controllo a priori di costituzionalità del Conseil constitutionnel, Décision n. 2019-778 DC, 21/03/2019.

dalla mente l'idea di un giudice *bouche de la loi*. Il diritto ha bisogno del certo, e questo certo non può non paludarsi dei paramenti della verità. Che dietro ciò si nasconda, almeno in parte, una finzione può solo essere sussurrato: detto a voce alta, si teme, farebbe tremare le fondamenta stesse delle civiltà giuridiche occidentali. Questo è un problema noto a livello teorico ma sempre rimosso, o occultato. Prima però di discutere le reti come utile sostituto del giudice occorrerebbe affrontare e risolvere questo e altri problemi da un punto di vista teorico generale, scientifico e culturale, sia su quel che il diritto è e dà realmente, sia nel suo aspetto normativo.

Per il diritto non c'è dubbio: questi sistemi aprono possibilità fino ad ora impensabili nella soluzione di conflitti, la loro applicazione al settore giuridico e in particolare giudiziario appare promettente ed inevitabile, nonostante le recenti limitazioni regolamentari introdotte in sede europea. Saranno possibili nuovi strumenti e nuove procedure nei quali il ruolo del giurista si limiterà a quello di mediatore, lasciando agli algoritmi la parte ragionativo-risolutiva. Al contrario, il pesante apparato giudiziario statale, formato da leggi, interpreti, applicatori, e poi sentenze, dottrine, argomenti e linguaggi tecnici potrà venire alleggerito, apportando celerità, prevedibilità ed economicità. È mia opinione che la regolamentazione europea in merito all'utilizzazione dei SA nel processo non servirà ad arrestare questo cambiamento, così come la regolamentazione in materia di privacy non è servita a proteggere la nostra riservatezza. In uno Stato liberista i principi che dominano e impostano le scelte politiche sono dettati dal mercato e questo riguarderà anche l'etica, come in effetti già sta succedendo.

Già agli albori dell'informatica apparve indubbio che una "macchina intelligente" dovesse poter anche simulare alcune capacità umane come quella di apprendere, di fare induzioni e generalizzazioni, non solo calcoli o deduzioni.

Nascono così due idee particolari in riguardo ai sistemi intelligenti. La prima è l'idea dei sistemi probabilistici, cioè di programmi che per un problema in cui parte dei dati sono ignoti riescono a trovare soluzioni 'ragionevoli' o accettabili. La seconda è l'idea dei sistemi che apprendono, di programmi che sulla base dei propri errori si automodificano in modo da non ripetere l'errore. Insomma, il contrario di un programma di intelligenza artificiale classica, che elabora dati di un sistema chiuso, di una realtà già classificata e suddivisa in contenitori concettuali determinati e definiti logicamente. Non si occupa di aprire queste classificazioni né di sussumere la realtà sotto queste classificazioni, bensì solo di ragionare in modo valido all'interno di esse. Tutto il sapere è contenuto in esse e il programma non è in grado di trovare qualcosa di nuovo, ma solo, eventualmente, di esplicitare ciò che è già contenuto nel programma e nella conoscenza fornita ma non ancora raggiunto da esplicite inferenze.

L'essere umano però ha bisogno del confronto con la realtà e da essa trae le informazioni per modificare le vecchie classificazioni, per crearne di nuove e distruggerne di vecchie, anche solo per fornire ai programmi di IA le rappresentazioni e classificazioni entro le quali argomentare: le rappresentazioni umane della realtà sono in continuo mutamento e anche più velocemente della realtà stessa. È la simulazione di questa capacità umana su SA che infastidisce o preoccupa i ricercatori di etica e i giuristi, in parte anche a ragione, vedremo quale, ma prima confrontiamoci con la parte del torto.

La domanda che desidero porre è: sono queste capacità simulate nei SA utili per il giurista? Si noti che utili per il giurista non significa direttamente che esse siano anche necessarie al giurista nell'interpretazione o applicazione della legge oppure che esse siano tipiche del ragionamento giuridico. Dan Hunter dieci anni fa metteva in evidenza il non corretto uso delle reti, seguendo ciò che la teoria generale del diritto teorizzava sul diritto stesso. La conclusione era la seguente:

Le reti neurali possono essere adatte per l'ambito giuridico, ma dobbiamo considerare attentamente la teoria sull'intelligenza artificiale e i modelli teorici giuridici prima di sostenerne l'uso. In particolare, è preoccupante che si assista di nuovo alla tacita accettazione del positivismo normativista. Questo sembra ignorare i reali benefici che un paradigma basato sulla statistica come le reti neurali può apportare al nostro studio del diritto<sup>11</sup>.

L'opinione che sostengo è tripartita. La prima parte ritiene che ciò che si possa ragionevolmente già richiedere a un sistema di intelligenza artificiale è di aiutare il giurista nell'interpretazione-applicazione della legge, ma a condizione del controllo della ragionevolezza dei risultati. Questo è un controllo che entra nel merito del rispetto delle garanzie giuridiche per gli imputati, o le parti, e del rispetto e raggiungimento degli standard di efficienza tipici degli ordinamenti giuridici.

La seconda parte della mia opinione sostiene che nel passato più recente i sistemi di intelligenza artificiale abbiano iniziato a inserirsi nei mezzi di produzione e controllo delle regole giuridiche e della loro applicazione, in modo sempre più capillare e difficilmente eliminabile o modificabile. Anche qui, un controllo meramente procedurale non è sufficiente e forse neppure possibile, però è necessaria una riflessione giuridica più approfondita rispetto a quanto sin qui i giuristi sono stati in grado di offrire. Mi riferisco alla relazione tra l'individuo e i dati o le informazioni che lo riguardano. La discussione è sorta all'indomani degli attentati alle Torri gemelle, ma ancora non c'è soluzione all'orizzonte. La normativa europea è tesa a tutelare più il libero mercato che la privacy dei dati, ma che sui dati relativi a individui sia accettabile – ai fini del mantenimento delle regole democratiche – l'applicazione delle leggi di mercato deve

<sup>11</sup> D. Hunter, *Looking for Law in all the Wrong Places*, pp. 62 ss.

essere rimesso sul tavolo della discussione. È oramai improrogabile quello che la dottrina italiana chiamò *Habeas data*<sup>12</sup>.

La terza parte della mia opinione sostiene che una corretta simulazione del ragionamento del giudice o dell'interprete in un sistema di intelligenza artificiale, possa aiutare, come in parte è già avvenuto, a comprendere scientificamente il ragionamento giuridico come tale, o a prevedere l'esito di una controversia, senza peraltro rappresentare una accettabile simulazione del ragionamento e della decisione stesse. Dal punto di vista scientifico sono degli strumenti rivoluzionari e i tanti arroccamenti delle scuole accademiche, ai quali assistiamo oggi, sono dovuti proprio al fatto che mai come adesso è possibile verificare la correttezza o adeguatezza di molte teorie sul diritto, e molte non tengono più il passo.

Tornando dunque al *machine learning*, esso si basa, abbiamo detto, sull'errore, sull'apprendimento e sulla autoregolazione in contesti variabili, ove non tutto è conosciuto o definito, il *machine learning* può fare analogie, indurre e generalizzare. Alla politica del certo e del passo dopo passo le reti oppongono quella del probabile e della singolarità del ragionamento induttivo. Per questo, fin dall'inizio, le reti hanno stupito più che convinto, però i loro risultati non sono ottenibili con i sistemi di intelligenza artificiale classica. L'intelligenza artificiale subsimbolica opera non seguendo una logica aristotelica o simbolica, ma su algoritmi basati su principi di ottimalità, né si può scegliere diversamente, ché equivarrebbe a rinunciare all'intelligenza artificiale subsimbolica e ai suoi prodotti, traguardo che nelle nostre economie liberiste appare assai velleitario. Ora, se gli attuali sistemi giuridici funzionassero in modo soddisfacente (secondo i loro stessi parametri valutativi) ben ci si potrebbe permettere l'agio di non ac-

<sup>12</sup> V. Frosini, *La protezione della riservatezza*, pp. 37 ss.

cedere a queste logiche diverse, ma non è questo il nostro caso. La domanda sul da farsi resta aperta. D'altro lato, le costrizioni derivanti dalle nostre economie liberiste impongono un mercato efficiente e per ottenerlo già si aprono strade alternative e autonome rispetto alle regolamentazioni (lente e inefficienti) statali. I mercati virtuali mondiali sarebbero impensabili senza i sistemi privati di risoluzione delle controversie e gli altri strumenti di *soft law* elaborati dai grandi *stakeholders* dell'economia digitale. Anche senza giungere alle autorità private, la ricerca scientifica ha già predisposto strumenti algoritmici di risoluzione delle controversie autonomamente implementabili da studi legali, notai o mediatori<sup>13</sup> o anche aziende. A mio avviso la strada è già aperta verso una privatizzazione di ampi settori della giustizia, ma una discesa a rotta di collo su di essa sarebbe rovinosa per le democrazie così come lo è stata nella tutela della privacy. Questi strumenti rivedono tutti i presupposti sui quali si fonda il giudicato in uno Stato di diritto. Nel caso degli algoritmi equitativi<sup>14</sup> i valori sui quali decidere la controversia e dividere beni o diritti li pongono i partecipanti alla procedura in modo del tutto soggettivo e anche, al limite, completamente slegato dal valore di mercato dei beni o diritti in controversia. Il giudice, o il mediatore funge da semplice catalizzatore della decisione algoritmica: invita le parti, le porta a formulare le loro scale di valori, prepara l'algoritmo alla decisione. L'algoritmo decide poi, su dati forniti dalle parti e dal giudice-mediatore, ma, possibilmente, anche in assenza di confronto con la normativa statale. Il

<sup>13</sup> F. Romeo, *Algoritmi di giustizia ed equità*, pp. 1-20; Id., *Algorithmic Conflict Resolution*.

<sup>14</sup> Sulla definizione di "algoritmi equitativi" vedi F. Romeo, *Algorithmical Justice*, pp. 9 ss. Il settore di ricerca risale agli scritti di Brams e Taylor: S. J. Brams, *Negotiation Games*; S. J. Brams - A. D. Taylor, *Fair division: from cake-cutting to dispute resolution*; S. J. Brams, *Game Theory and the Humanities. Bridging two worlds*.

percorso della procedura verso la decisione giuridica (perché risolutiva di una controversia) non guarda a norme esterne rispetto al soggetto per qualificare giuridicamente il caso. Lo sguardo giuridificante<sup>15</sup>, i soggetti in controversia lo volgono verso loro stessi, stabilendo il loro ordine di valori per vederlo affermare dall'algoritmo. In tutti i casi nei quali si tratta di diritti disponibili la procedura può sicuramente saltare la normativa statale e ad essa sostituire la procedura algoritmica e gli ordini valoriali individuali al posto di quelli sociali. Non è questa una procedura che si avvale di principi o regole stabilite dalla volontà popolare in leggi dello Stato e può saltare il controllo statale laddove si chiuda nell'accordo delle parti. Infatti questo è il presupposto e la sfida dei sistemi equitativi di *Fair division*: la possibilità di raggiungere algoritmicamente divisioni nelle quali le parti sono soddisfatte della propria parte e non intendono cambiarla con quella di nessun altro partecipante. Le soluzioni sono dette, per questo, *envy free*. È possibile che una parte abbia ottenuto meno di quanto gli sarebbe spettato per legge, secondo i valori di mercato o di legge, ma la soddisfazione è un apprezzamento in gran parte soggettivo, di questo i sistemi statali di risoluzione delle controversie non fanno conto<sup>16</sup>. La giustizia, del resto, si misura sul metro della felicità o della soddisfazione, non su quello della verità.

### 3. *Meraviglioso sollievo*

Propongo un esperimento mentale: si immagini che improvvisamente ci si svegli da un sonno profondo e da un sogno-incubo rap-

<sup>15</sup> Traggio l'immagine da N. Irti, *Diritto senza verità*, pp. 52 ss.

<sup>16</sup> F. Romeo, *Algoritmi di giustizia ed equità*, pp. 16 ss.

presentato dal presente tecnologico. Al risveglio ci accorgiamo che ogni strumento di intelligenza artificiale è scomparso. Nessun meta-verso all'orizzonte, nessuna predizione sulla criminalità, nessun finto chirurgo-meccanico, nessun chatbot, né interrogazioni di motori di ricerca, né multe automatiche. Nulla di tutto ciò, via, che sollievo, meraviglioso! Accompagnato poi, questo meraviglioso sollievo, dalla soddisfazione di sapere le *big tech* distrutte, in spaventoso *default*. Un ritorno ai bei vecchi tempi, *zurück zu Ethik selbst!* Quanto era bello il passato. Certo ma solo quello non dimenticato, e la gioia si tramuterebbe subito in inconsolabile disperazione, perché l'ultimo tentativo di porre rimedio alle carenze della natura umana sarebbe fallito. Il ritorno a ciò che la speculazione etica ci offriva prima della introduzione della IA significherebbe il fallimento definitivo dell'*Homo Sapiens* e della sua cultura, scompariremmo anche noi, come il *Neanderthalensis*. Se anzi guardiamo sotto una lente l'attuale situazione culturale, certo essa è pessima, ma in essa vi troviamo l'eredità di un percorso che si radica molto indietro nella storia dell'umanità e che la speculazione etica non ha saputo evitare, ha, anzi, in parte, contribuito a costruire. L'uomo che agisce in questo mondo è sempre lo stesso ed è lui il problema, non tanto la IA che lo imita, né i SA che lo supportano.

Se la situazione politico-culturale è critica, la causa è da ricercare nelle teorie che questa situazione hanno costruito. Se, ad esempio, il soggetto di diritto è divenuto merce, è molto probabile che le teorie attuali sul mercato, oggi quelle liberiste, abbiano contribuito a questo esito.

Guardando all'evoluzione della cultura umana, ci troviamo oggi con un gigantesco complesso di rappresentazioni della realtà che, però, tutto spiegano tranne colui che ha creato quelle rappresentazioni e per quale motivo lo abbia fatto. Alcuni studi una risposta ben fondata la danno, considerando la cultura un insieme di strumenti utili a massimizzare l'utile ricavato dal *Sapiens* nel suo agire nella realtà. La cultura si è evoluta in modo rapido e sempre più rapido, non altrettanto è pos-

sibile dire per il *Sapiens*, che è, anzi, rimasto molto vicino alle sue predisposizioni genetiche. Il risultato è quindi, oggi, che quelle predisposizioni, autoritarie, gerarchiche e distruttive controllano degli strumenti culturali che possono essere esiziali per l'intera specie. L'etica dell'uomo contemporaneo esprime ancora quelle predisposizioni etiche originarie sulle quali si è costruita la cultura. L'egoismo genetico domina ancora l'azione del *Sapiens* nucleare computerizzato e non si può pensare di controllarlo ancora con gli strumenti del dover essere, etico o giuridico che sia: lo strumento normativo del dover essere è oramai corrosivo, non penetra fino all'effettività. Torno quindi all'etica e domando quali nuovi strumenti culturali siano disponibili per fungere da controllo alla esplicazione economica della predisposizione all'egoismo del *Sapiens*. Ma tutto tace. Senza optare per lo *ethiks bashing* noto che non è realistico, né adeguato alla sua funzione, scaricare sul diritto e su un suo intervento la ricerca e la responsabilità della soluzione<sup>17</sup> nel senso indicato dalla Tafani e da alcuni altri studiosi di etica. Il diritto non è in grado di elaborare scale valoriali che si fondino in altro che non sia convenzionale. In questo sta il suo nichilismo, come un'autorevole dottrina ha da tempo evidenziato<sup>18</sup>. Il diritto si forma su ciò che viene pensato come condiviso o da imporre da una moltitudine di soggetti<sup>19</sup>. Oggi, il *trend* che si è evidenziato nelle economie liberiste va in direzione di una privatizzazione del diritto, che tenga conto degli aspetti

<sup>17</sup> D. Tafani, *L'«etica» come specchietto*, pp. 11 ss.

<sup>18</sup> N. Irti, *Nichilismo giuridico*.

<sup>19</sup> «La norma fondamentale non è una, assoluta e vincolante per tutti, ma molteplice e discordante: messaggio divino o ragione dell'uomo, tradizione storica, autorità e forza dello Stato e via seguitando. Il politeismo è il politeismo delle norme fondamentali, fra le quali siamo chiamati a compiere la nostra scelta. Ciascuno di noi sceglie la propria norma fondamentale, garanzia di validità dei criteri di giudizio, da noi applicati ai fatti storici». N. Irti, *Diritto senza verità*, p. 60.

non razionali trascurati negli ordinamenti giuridici. Nella logica del profitto, quel che è possibile fare lo si farà. Tocca quindi agli eticisti tornare all'individuo e tornare a dare rilievo ai suoi aspetti emozionali ed irrazionali portandoli nella sua socialità senza che deflagrino in modo dirompente e incontrollato, come è già successo nel Novecento, e senza che restino un dover essere bisognoso di ogni supporto e puntellamento. L'etica deve riallacciarsi allo studio evoluzionista dell'uomo, lì solo potrà trovare la congiunzione tra dover essere etico ed essere, avrà una spiegazione e potrà fare la sua parte. A quel punto il diritto potrà intervenire con un volto nuovo e un braccio efficace.

## Bibliografia

- E. Bietti, *From Ethics Washing to Ethics Bashing, A View on Tech Ethics From Within Moral Philosophy*, <<https://ssrn.com/abstract=3914119> ult. Cons 26/06/24>.
- S. J. Brams, *Negotiation Games, Applying Game Theory to Bargaining and Arbitration*, Oxon Routledge, 1990.
- S. J. Brams - A. D. Taylor, *Fair division: from cake-cutting to dispute resolution*, Cambridge (MA), Cambridge University Press, 1996.
- S. J. Brams, *Game Theory and the Humanities. Bridging two worlds*, Cambridge (MA), the MIT Press, 2012.
- G. Contissa, *Automation and liability: an analysis in the context of socio-technical systems*, in «I-Lex. Rivista di scienze giuridiche, scienze cognitive ed intelligenza artificiale», 11, 1, 2017, pp. 17-45. <<http://www.i-lex.it/index.php/volume-11/fascicolo-11-1/10-automation-and-liability> ult.cons>, 27/06/24.
- V. Frosini, *L'uomo artificiale, etica e diritto nell'era planetaria*, Milano, Spirali, 1986.
- V. Frosini, *La protezione della riservatezza nella società informatica*, in N. Matteucci, *Privacy e banche dei dati*, Bologna, il Mulino, 1981.
- D. Hunter, *Looking for law in all the wrong places: legal theory and legal neural networks*, in H. Prakken - A.J. Muntjewerff - A. Soeteman - R. Winkels, *Legal knowledge based systems. JURIX 94. The Relation with Legal Theory*, Lelystad, Koninklijke Vermande, 1994.

- N. Irti, *Diritto senza verità*, Roma-Bari, Laterza, 2011.
- N. Irti, *Nichilismo giuridico*, Roma-Bari, Laterza, 2004.
- V. A. Laptev - D. R. Feyzrakhmanova, *Application of Artificial Intelligence in Justice: Current Trends and Future Prospects*, in *Human-centric Intelligent Systems*, 4, 2, 2024, <<https://link.springer.com/article/10.1007/s44230-024-00074-2> ult. cons. 30/06/2024>.
- G. Longo, *Matematica e senso. Per non divenire macchine*, Milano-Udine, Mimesis, 2021.
- F. Romeo - M. Dall'Aglio - M. Giacalone (a cura di), *Algorithmic Conflict Resolution: Fair and Equitable Algorithms in Private Law*, Torino, Giappichelli, 2019.
- F. Romeo, *Algorithmical Justice, What is Changing in Law*, in S. Martuccelli - F. Romeo - M. Giacalone, *The European Common Ground of Available Rights*, Torino, Editoriale scientifica, 2020.
- F. Romeo, *Il diritto artificiale*, Torino, Giappichelli, 2002.
- F. Romeo, *Algoritmi di giustizia ed equità nel diritto. Quando razionalità ed emotività convergono*, in «I-Lex», 2021, pp. 1-20.
- D. Tafani, *L'«etica» come specchio per le allodole. Sistemi di intelligenza artificiale e violazioni dei diritti*, in Bollettino telematico di filosofia politica 2023, <<https://commentbfp.sp.unipi.it/letica-come-specchietto-per-le-allodole/> , ult. cons. 27/06/28>.
- B. Wagner, *Ethics As An Escape From Regulation. From “Ethics-Washing” To Ethics-Shopping?*, in E. Bayamlioglu - I. Baraliuc - L. Janssens - M. Hildebrandt, *BEING PROFILED: COGITAS ERGO SUM: 10 Years of Profiling the European Citizen*, Amsterdam, Amsterdam University Press, 2018.

Francesco Romeo  
Università degli Studi di Napoli Federico II  
francesco.romeo@unina.it  
<https://orcid.org/0000-0001-7491-0996>

STEVEN UMBRELLO

*L'intersezione tra il realismo critico di Bernard Lonergan, il bene comune e l'intelligenza artificiale nelle pratiche religiose moderne\**

*Abstract:* Artificial intelligence (AI) profoundly influences a range of contemporary social structures, including religious dynamics. Utilizing Bernard Lonergan's critical realism as an analytical framework, this article investigates the intersections between AI and religious traditions in their shared pursuit of the common good. Drawing on Lonergan's principle that human beings construct their understanding through cognitive processes, we examine how AI-mediated realities align with or challenge traditional religious principles. By exploring specific case studies, we highlight AI's role in reshaping religious symbols, rituals, and even in generating new spiritual meanings. Employing Lonergan's insights on the balance between subjectivity and objectivity, this study analyzes AI's potential to create new sacred spaces and to challenge religious orthodoxy. The core of the discussion centers on the negotiation between religious values and technological innovation, assessing how AI can support religious life while preserving its essential core. Ultimately, this article underscores the significance of the common good in the era of AI-driven religious evolution.

*Keywords:* Bernard Lonergan, Critical Realism, Digital Religion, Artificial Intelligence, Religious Experience, Common Good, Technology.

1. *Introduzione*

In un'epoca segnata da rapidi progressi tecnologici, l'intelligenza artificiale (IA) si pone come una delle forze più trasformative, rimo-

\* Questo articolo, con minime variazioni rispetto all'originale, è la traduzione italiana di un contributo in lingua inglese originariamente pubblicato sulla rivista *Religions* con il titolo *The Intersection of Bernard Lonergan's Critical Realism, the Common Good, and Artificial Intelligence in Modern Religious Practices*.

dellando numerosi aspetti della società, comprese le comunità religiose. In tutto il mondo, dal monitoraggio della comunità Uighur di Xinyang<sup>1</sup> alle discussioni etiche intorno al Libro Bianco dell'Unione Europea sull'IA<sup>2</sup>, le implicazioni dell'IA nei contesti religiosi hanno iniziato a emergere come una sfida e un'opportunità<sup>3</sup>. Le comunità religiose non sono semplici destinatari passivi di questa trasformazione digitale. Si confrontano attivamente con i cambiamenti tecnologici, cercando modi per allineare questi sviluppi con le loro credenze e tradizioni fondamentali. Nel considerare l'interfaccia dell'IA con le tradizioni religiose, è istruttivo riflettere sull'evoluzione storica dell'IA, dai suoi concetti fondamentali nel movimento cibernetico degli anni Cinquanta alle sue attuali sofisticate applicazioni. Questa visione retrospettiva riconosce che i principi della teoria dei sistemi, del feedback e del controllo, stabiliti agli albori della cibernetica, hanno profondamente influenzato lo sviluppo delle tecnologie di IA contemporanee<sup>4</sup>. Comprendere questa eredità dell'IA arricchisce la nostra comprensione dei suoi potenziali impatti e delle considerazioni etiche in ambito religioso. L'adozione di strumenti di IA da parte delle istituzioni religiose è spesso un processo attento e ponderato, sostenuto da riflessioni teologiche e discussioni comunitarie<sup>5</sup>.

Ciò è particolarmente importante in quanto le tecnologie in questione possono influenzare in modo significativo le pratiche

<sup>1</sup> J. Wakefield, *AI emotion-detection software tested on Uyghurs*.

<sup>2</sup> European Commission, *White Paper On Artificial Intelligence - A European approach to excellence and trust*.

<sup>3</sup> D. Löffler - J. Hurtienne - I. Nord, *Blessing robot Bless U2: a discursive design study to understand the implications of social robots in religious contexts*.

<sup>4</sup> T. Marlowe - J. R. Laracy, *Philosophy and Cybernetics: Questions and Issues*.

<sup>5</sup> Y. Song, *Religious AI as an Option to the Risks of Superintelligence: A Protestant Theological Perspective*.

di culto, le interpretazioni dottrinali e il più ampio impegno della comunità. Tuttavia, la comprensione di questa interazione tra IA e dinamiche religiose richiede un quadro di riferimento completo. Tale quadro non deve solo tenere conto degli aspetti tecnici e pratici dell'incorporazione dell'IA nelle istituzioni e nelle pratiche religiose, ma anche approfondire la dimensione filosofica ed etica di questa impresa. Man mano che l'intelligenza artificiale viene integrata nelle attività religiose e nei rituali quotidiani sorgono domande sull'autenticità delle esperienze religiose mediate dall'IA, sui rischi potenziali di un'interpretazione errata e sulle implicazioni più ampie di una visione del mondo religiosa potenziata dall'IA. Affrontare questi problemi richiede un approccio interdisciplinare che unisca le intuizioni tecno-logiche con una profonda comprensione teologica, assicurando che le religioni, quando entrano nell'era digitale, lo facciano con cautela e chiarezza. La filosofia di Bernard Lonergan, caratterizzata dall'enfasi sul "realismo critico", offre un approccio tempestivo e perspicace a questo enigma. Il realismo critico, come lo concepisce Lonergan, si colloca a cavallo dell'intricato equilibrio tra soggettività e oggettività, apprezzando i processi soggettivi che sono alla base della nostra comprensione di una realtà oggettiva<sup>6</sup>. All'interno di questo quadro, l'interazione tra IA e religione diventa meno una questione di accettazione o di rifiuto assoluto e più una questione di discernimento dell'integrazione<sup>7</sup>. La notevole enfasi di Lonergan sul «bene comune» fornisce uno sfondo etico e filosofico, fondando i progressi tecnologici in una visione di prosperità umana. Sfruttando le intuizioni di Lonergan, possiamo analizzare sistematicamente le complessità che permeano il domi-

<sup>6</sup> B. J. F. Lonergan, *A Second Collection* (CWL 13).

<sup>7</sup> P. H. Byrne, *The Ethics of Discernment: Lonergan's Foundations for Ethics*.

nio religioso digitale, dando un senso alle sue sfide, ai potenziali benefici e alle implicazioni generali.

Il contributo è strutturato in modo tale da far emergere progressivamente queste complessità. Inizio collocando la filosofia di Lonergan nel contesto dell'IA e delle dinamiche religiose. Le sezioni successive approfondiscono poi i casi di studio che mostrano come le comunità religiose affrontano le innovazioni dell'IA, utilizzando l'enfasi di Lonergan sul bene comune come pietra di paragone. Nell'esaminare l'uso dell'IA in contesti religiosi, esamino anche le sfide poste da questa tecnologia, soprattutto quando sembra divergere dai principi religiosi, ed esploro come i gruppi religiosi possano conciliare queste tensioni. Concludo sintetizzando i risultati, riflettendo sulle implicazioni più ampie del nesso IA-religione e suggerendo percorsi per una co-evoluzione armoniosa nell'era digitale.

## *2. La filosofia di Lonergan e la comprensione mediata della realtà*

Nella matrice in evoluzione dell'Intelligenza Artificiale e della pratica religiosa, uno sguardo superficiale o sommario non sarà sufficiente. L'intricata danza tra tradizioni religiose millenarie e i rapidi sviluppi dell'intelligenza artificiale richiede una lente interpretativa sofisticata. Richiede una solida struttura filosofica, in grado non solo di articolare le sfumature della situazione attuale, ma anche di immaginare traiettorie etiche e consapevoli per il futuro. Nell'ambito di questa pressante esigenza, il lavoro di Bernard Lonergan, in particolare il suo «realismo critico», diventa particolarmente rilevante e applicabile. Il realismo critico di Bernard Lonergan costituisce una parte cruciale del quadro teorico di questo lavoro. Per fornire un contesto, soprattutto ai lettori che non conoscono la sua opera, è indispensabile collocare le sue idee all'interno del più ampio pa-

norama del realismo critico e del suo sviluppo storico. Il realismo critico, come approccio filosofico alla conoscenza, cerca di bilanciare le dimensioni soggettive e oggettive della comprensione, un tema centrale sia per le scienze naturali che per la teologia. Ian Barbour, fisico sperimentale e teologo protestante, insieme a Lonergan, filosofo e teologo cattolico, hanno entrambi sostenuto un'epistemologia «realista critica»<sup>8</sup>.

Questo approccio sfida diverse posizioni filosofiche come il positivismo, che vede le teorie come semplici sintesi di dati osservabili; lo strumentalismo, che considera le teorie come strumenti utili privi di verità; e l'idealismo, che riduce le teorie a costrutti mentali. Il realismo critico sostiene invece che le teorie offrono una conoscenza parziale, rivedibile e referenziale del mondo, spesso espressa al meglio attraverso modelli e metafore. Questa epistemologia riconosce il ruolo creativo della mente umana, pur affermando l'esistenza di modelli oggettivi in natura. La nozione di realismo critico di Barbour afferma che i concetti validi non sono solo utili ma veri, in quanto rappresentano la struttura degli eventi nel mondo. Questa prospettiva contrasta nettamente con l'epistemologia costruttivista, che enfatizza la conoscenza come attivamente costruita dal soggetto conoscente<sup>9</sup>. Il realismo critico di Barbour, quindi, sostiene la natura astratta della fisica teorica e la necessità dell'indagine sperimentale, riconoscendo che le teorie scientifiche, sebbene incomplete, sono tentativi di rappresentare il mondo oggettivo. L'approccio di Lonergan al realismo critico, sviluppato attraverso il suo Metodo Empirico Generalizzato (GEM), si concentra sul processo di conoscenza

<sup>8</sup> J. R. Laracy, *Theology and Science in the Thought of Ian Barbour*.

<sup>9</sup> J. R. Laracy, *Epistemology and Metaphysics in Interdisciplinary Communication: Insights from Ian Barbour and Bernard Lonergan*.

umana che avviene all'interno della coscienza. Il suo metodo divide questo processo in esperienza, comprensione, giudizio e decisione, sostenendo la necessità di una conversione intellettuale per cogliere le operazioni della propria mente. Il realismo critico di Lonergan emerge da un percorso personale di autoappropriazione filosofica, utilizzando esercizi che stimolano le intuizioni della matematica, delle scienze naturali e del senso comune. Barbour e Lonergan condividono l'impegno a comprendere la conoscenza che integra l'esperienza soggettiva con la realtà oggettiva, rendendo le loro intuizioni collettive particolarmente pertinenti per la nostra esplorazione dell'IA nel contesto delle pratiche religiose.

### 2.1. Il realismo critico di Lonergan

Al centro degli sforzi filosofici di Bernard Lonergan si trova la nozione profonda di «realismo critico»<sup>10</sup>. Non si tratta di un concetto che si può afferrare sfiorando casualmente la superficie. Approfondendo la questione, si scopre una comprensione dettagliata del modo in cui gli esseri umani si impegnano e interpretano il mondo che li circonda. Il realismo critico di Lonergan, tuttavia, presenta un quadro più intricato. Esso accentua l'idea che la comprensione umana della realtà non è diretta, ma piuttosto mediata attraverso una serie di processi cognitivi *a priori*. Questi processi non sono arbitrari, ma seguono un modello strutturato e coerente. Un'analisi più attenta della teoria di Lonergan rivela una catena di passaggi cognitivi che l'individuo compie per arrivare alla comprensione. Si comincia con le esperienze - i dati grezzi della coscienza. Queste esperienze danno

<sup>10</sup> T. J. Tekippe, *What is Lonergan Up to in Insight?: A Primer*.

origine a intuizioni, che vengono ulteriormente analizzate, portando a giudizi sulla veridicità della nostra comprensione. Infine, sulla base di questi giudizi, gli individui prendono decisioni che guidano le loro azioni nel mondo. Nella prospettiva di Lonergan, quindi, il viaggio verso la comprensione non è statico o passivo. Si tratta invece di un processo dinamico e in evoluzione, in cui gli esseri umani non sono spettatori ma piuttosto partecipanti attivi, che costruiscono e ricostruiscono costantemente la loro comprensione del mondo sulla base di nuove esperienze e intuizioni.

### 2.1.1. Coscienza intenzionale: Un impegno attivo

Lonergan non si limita a descrivere i processi cognitivi. Aggiunge un ulteriore livello con il concetto di «coscienza intenzionale»<sup>11</sup>. Questo concetto indica la natura proattiva degli esseri umani nella loro ricerca di comprensione. Piuttosto che essere recettori passivi di informazioni, gli esseri umani sono visti come esseri intenzionali, che cercano intenzionalmente la conoscenza e la comprensione. Questo impegno intenzionale con la realtà significa che la nostra comprensione emerge attraverso un vibrante dialogo tra il nostro mondo interno soggettivo e il mondo esterno oggettivo. Il nostro regno interiore, quello delle cognizioni, delle credenze e dei valori, interagisce costantemente con il mondo esterno degli eventi, delle azioni e delle realtà tangibili. È in questa interazione dialogica che emerge la comprensione.

L'enfasi posta da Lonergan su questa natura attiva e intenzionale della coscienza non è solo gergo filosofico. Ha profonde implicazio-

<sup>11</sup> P. Beer, *An Introduction to Bernard Lonergan: Exploring Lonergan's approach to the great philosophical questions*.

ni, soprattutto nell'epoca attuale dominata da progressi tecnologici come l'IA. Se la comprensione è attiva e intenzionale, allora i modi in cui l'IA si interfaccia con la sfera religiosa e filosofica diventano cruciali. Il realismo critico di Lonergan potrebbe quindi offrire un quadro di riferimento per comprendere l'interazione sfumata tra i regni soggettivo e oggettivo della realtà. Sottolinea inoltre la natura attiva e dinamica della comprensione umana, evidenziando che la comprensione non è semplicemente ricevuta, ma è ardentemente costruita. Questa prospettiva è indispensabile per navigare nelle complesse intersezioni tra IA, religione e cognizione umana.

### 2.1.2. Rilevanza nel contesto dell'IA e della religione

Se consideriamo l'ascesa dell'IA, soprattutto nel campo delle pratiche e delle credenze religiose, le implicazioni di questa comprensione mediata della realtà diventano ancora più evidenti. L'IA, come tecnologia, funziona con algoritmi basati sui dati, mirando a obiettivi specifici attraverso previsioni probabilistiche. Piuttosto che formare modelli completi della realtà, l'IA enfatizza la comprensione specifica del compito, che può mostrare fragilità quando si trova di fronte a contesti diversi dall'ambiente di addestramento<sup>12</sup>. Questi sistemi non mirano a incapsulare la totalità della realtà. Si sforzano invece di ottenere accuratezza, precisione e robustezza nell'ambito dei compiti loro assegnati, basando i loro risultati sui dati specifici su cui sono stati addestrati e sugli algoritmi che regolano le loro operazioni. Questa convergenza tra la comprensione di Lonergan della cognizione umana e il funzionamento dei sistemi di IA fornisce una

<sup>12</sup> O. Kudina, *"Alexa, who am I?": Voice Assistants and Hermeneutic Lemniscate as the Technologically Mediated Sense-Making*.

lente appropriata attraverso la quale possiamo indagare l'impatto dell'IA sulle comunità religiose. Ad esempio, quando un sistema di IA offre un'interpretazione religiosa o facilita un'esperienza religiosa, non presenta una verità religiosa «oggettiva». Al contrario, offre una comprensione mediata, influenzata dalla programmazione, dai set di dati e dai pregiudizi intrinseci<sup>13</sup>.

Le comunità religiose, quando si interfacciano con il regno dell'intelligenza artificiale, non si limitano a navigare nelle acque di un apparato tecnologico neutrale. In sostanza, stanno interagendo con sistemi intricati che contengono le loro interpretazioni e rappresentazioni delle verità e delle pratiche religiose. Questi sistemi di IA, spesso modellati da pregiudizi umani, background culturali e influenze socio-politiche, presentano visioni del mondo specifiche, «mediando» efficacemente i contenuti religiosi nei loro modi unici<sup>14</sup>.

Questo fenomeno dei sistemi di intelligenza artificiale che possiedono le proprie comprensioni mediate risuona con il lavoro di Lonergan. Per Lonergan, il viaggio della conoscenza e della comprensione non è un percorso lineare con un punto di arrivo definitivo. Si tratta invece di un percorso espansivo, iterativo e continuo. Ogni fonte di conoscenza, ogni intuizione e ogni esperienza, sia essa una scrittura tradizionale o un moderno algoritmo di intelligenza artificiale, contribuisce a questa comprensione in continua evoluzione.

Facendo un parallelo, proprio come gli individui assorbono, valutano e integrano le intuizioni provenienti da una miriade di fonti nella loro ricerca della verità, le comunità religiose, quando si con-

<sup>13</sup> H. Thinyane - F. Sassetti, *Towards a Human Rights-Based Approach to AI: Case Study of Apprise*.

<sup>14</sup> P. Bory, *Deep new: The shifting narratives of artificial intelligence from Deep Blue to AlphaGo*.

frontano con l'IA, devono confrontarsi e discernere le intuizioni offerte da queste entità tecnologiche. In un certo senso, l'IA diventa un'altra fonte, un'altra «voce», nel vasto coro di influenze che danno forma alla comprensione religiosa in epoca contemporanea. Inoltre, nell'era dell'IA, questa dinamica diventa ancora più intricata. L'intelligenza artificiale, con la sua capacità di elaborazione dei dati e di riconoscimento dei modelli, ha il potenziale per fornire nuove prospettive su dottrine e pratiche religiose antiche<sup>15</sup>. Queste prospettive potrebbero rafforzare le interpretazioni tradizionali, metterle in discussione o addirittura introdurre dimensioni di comprensione completamente nuove. Le implicazioni sono profonde. Le comunità religiose, nelle loro interazioni con l'IA, sono effettivamente impegnate in un dialogo tra la saggezza antica e le moderne intuizioni computazionali. Per i credenti e i leader religiosi, questo rappresenta sia un'opportunità che una sfida. L'opportunità consiste nello sfruttare le capacità dell'IA per approfondire, espandere e arricchire la comprensione religiosa. Dall'altro lato, la sfida emerge nel discernere quali intuizioni mediate dall'IA si allineano con i principi fondamentali delle loro credenze e quali potrebbero allontanarsi o distorcere le loro verità spirituali. Facendo eco ai sentimenti di Lonergan, questo continuo impegno con l'IA in contesti religiosi esemplifica la natura dinamica e senza fine della nostra ricerca di comprensione. Attingiamo continuamente da una serie di fonti diverse, alla ricerca di chiarezza, coerenza e profondità. Nell'era digitale moderna, i sistemi di intelligenza artificiale, con le loro comprensioni mediate, diventano parte integrante di questo arazzo di conoscenze e credenze. Allo

<sup>15</sup> G. Nachshon - J. Zeleznikow - T. Dowdeswell, *From the Tree of Knowledge and the Golem of Prague to Kosher Autonomous Cars: The Ethics of Artificial Intelligence Through Jewish Eyes*.

stesso modo, anche l'enfasi di Lonergan sul bene comune diventa rilevante in questo contesto. Se le comunità religiose devono abbracciare l'IA, devono farlo in modo da allinearsi con la loro ricerca del bene comune. Ciò comporta un discernimento critico, assicurando che gli strumenti e le piattaforme di IA sostengano i valori, l'etica e le convinzioni fondamentali per queste comunità. La nozione di Lonergan di bene comune non riguarda solo il benessere individuale, ma comprende dimensioni sociali, culturali e persino cosmiche<sup>16</sup>. Pertanto, qualsiasi integrazione dell'IA nelle pratiche religiose deve essere valutata alla luce di questo ampio scenario, considerando non solo i benefici immediati ma anche le implicazioni più ampie per la prosperità umana e l'armonia sociale. Sfruttando le intuizioni di Lonergan sul realismo critico e sul bene comune, possiamo navigare nel complesso terreno in cui IA e religione si intersecano. Mentre le comunità religiose si confrontano con le potenzialità e le sfide poste dall'IA, la filosofia di Lonergan offre potenzialmente una solida base, spingendo queste comunità verso un impegno riflessivo, un discernimento critico e la ricerca del bene comune nell'era digitale.

### *3. L'intelligenza artificiale e la reinvenzione delle comunità religiose in una cultura globalizzata*

Nell'era digitale, le pratiche e le espressioni religiose sono sempre più mediate dalla tecnologia. Heidi Campbell<sup>17</sup> approfondisce questa trasformazione, sottolineando come le comunità religiose si adattino e siano plasmate dagli ambienti digitali. Questo gioco tra

<sup>16</sup> B. J. F. Lonergan, *Insight: A Study of Human Understanding*, (CWL 3).

<sup>17</sup> H. Campbell, *When Religion Meets New Media*.

religione e tecnologia offre un contesto pertinente per esplorare il realismo critico di Bernard Lonergan. La filosofia di Lonergan, che enfatizza l'interazione dinamica tra comprensione e realtà, fornisce un quadro di riferimento per comprendere come le credenze e le pratiche religiose siano reinterpretate e rimodellate nel regno digitale. In quest'ottica, l'avvento dell'intelligenza artificiale nei contesti religiosi può essere visto come un'estensione di questa trasformazione digitale, in cui l'intelligenza artificiale diventa uno strumento per l'espressione religiosa e la costruzione di comunità, proprio come Campbell descrive il ruolo di Internet nella religione.

### 3.1. I nuovi santuari digitali

Nell'era dell'interazione digitale globale, gli algoritmi guidati dall'intelligenza artificiale, soprattutto sulle piattaforme dei social media, sono diventati strumentali per diffondere gli insegnamenti religiosi a un vasto pubblico globale. Tali algoritmi sono progettati per comprendere le preferenze e i comportamenti degli utenti, consentendo così di curare i contenuti religiosi che sono in linea con i singoli utenti<sup>18</sup>. Questa diffusione mirata abbatte le barriere geografiche, le sfumature culturali e le divisioni politiche, facendo emergere l'universalità degli insegnamenti religiosi<sup>19</sup>. L'avvento dei chatbot, dotati di intricati database teologici, ha rivoluzionato il modo in cui i ricercatori si confrontano con i contenuti religiosi<sup>20</sup>. Queste entità

<sup>18</sup> T. Gillespie, *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*.

<sup>19</sup> J. T. Feezell - J. K. Wagner - M. Conroy, *Exploring the effects of algorithm-driven news sources on political behavior and polarization*.

<sup>20</sup> M. Graves, *ChatGPT's Significance for Theology*.

AI si rivolgono a menti curiose, fornendo risposte immediate, guidando gli utenti attraverso complesse domande teologiche e persino facilitando l'esplorazione più profonda delle complessità della dottrina religiosa. La loro disponibilità 24 ore su 24, 7 giorni su 7, assicura che la guida spirituale sia a portata di clic.

L'apprendimento automatico, un sottoinsieme dell'IA, può anche offrire intuizioni uniche nei voluminosi dati generati da forum, blog e discussioni religiose online. Analizzando le caratteristiche, i sentimenti e le tendenze di questi discorsi, i modelli di apprendimento automatico possono potenzialmente identificare i cambiamenti ideologici emergenti, le aree di contesa o le nuove interpretazioni all'interno delle comunità religiose<sup>21</sup>. Tali intuizioni proattive consentono ai leader religiosi e agli studiosi di affrontare i problemi, di affinare le strategie di sensibilizzazione e persino di promuovere dialoghi che mettano in contatto punti di vista diversi.

Tuttavia, l'influenza dell'IA sulle comunità religiose non si limita al mero scambio di informazioni o all'ampliamento della portata. L'IA ha il potenziale per approfondire i legami spirituali all'interno delle comunità e tra di esse. L'analisi predittiva, informata dai comportamenti e dalle preferenze dell'utente, può creare percorsi spirituali personalizzati. Che si tratti di suggerire letture delle scritture, pratiche di meditazione o eventi comunitari, l'IA garantisce che gli individui ricevano una guida in linea con le loro inclinazioni spirituali e traiettorie di crescita. Inoltre, l'intersezione tra IA e realtà virtuale (VR) ha aperto la strada a esperienze immersive prima ritenute impossibili; immaginate di non essere fisicamente in grado di recarvi in un luogo sacro, eppure, grazie alla VR alimentata dall'IA, si pos-

<sup>21</sup> S. Shen - J. Fan, *Emotion analysis of ideological and political education using a GRU deep neural network*.

sono “visitare” questi siti e viverli in modo multisensoriale<sup>22</sup>. Questi pellegrinaggi tecnologici permettono ai credenti di percorrere corridoi sacri, ascoltare antichi canti e persino partecipare a cerimonie religiose lontane, il tutto comodamente da casa. Queste esperienze, pur essendo tecnologiche, risuonano profondamente con il desiderio umano di connessione, appartenenza e arricchimento spirituale<sup>23</sup>. In sostanza, sebbene l’intelligenza artificiale possa sembrare un alleato improbabile per le tradizioni religiose, il suo profondo impatto è innegabile. Dall’espansione della portata alla promozione di connessioni più profonde, l’intelligenza artificiale sta giocando un ruolo fondamentale nel plasmare il panorama religioso del XXI secolo.

### 3.2. Rimodellare i simboli, i valori e le pratiche religiose

L’evoluzione del panorama religioso nell’era dell’intelligenza artificiale richiede un’indagine filosofica sul modo in cui le sfaccettature religiose fondamentali vengono rimodellate. La filosofia di Lonergan, con la sua enfasi sulla natura dinamica ed evolutiva della comprensione, offre un quadro adeguato per questa esplorazione.

I simboli religiosi, tradizionalmente considerati come costanti nel tempo, si trovano ora nel crogiolo dell’innovazione guidata dall’IA<sup>24</sup>. Grazie alla capacità dell’IA di elaborare grandi quantità di dati, i simboli religiosi vengono analizzati, reinterpretati e persino ricreati con una profondità e un’ampiezza prima inimmaginabili<sup>25</sup>. Consideriamo, ad esempio, l’arte religiosa. Algoritmi avanzati di intelligenza

<sup>22</sup> A.P., *VR pilgrimage to holy sites from Mecca to the Vatican, here’s how*.

<sup>23</sup> R. Min, *The metaverse is opening up the world’s holiest sites to virtual pilgrims*.

<sup>24</sup> C. Disbrey, *George Fox and Some Theories of Innovation in Religion*.

<sup>25</sup> J. F. McGrath, *Jesus according to AI*.

artificiale, dopo aver analizzato secoli di opere d'arte religiosa che abbracciano varie culture, sono ora in grado di generare arte che fonde senza soluzione di continuità motivi provenienti, ad esempio, dalle tradizioni cristiana e buddista<sup>26</sup>. Per i puristi, tali fusioni potrebbero apparire come un'esagerazione digitale. Tuttavia, dal punto di vista lonerganiano, queste nuove forme possono essere viste come estensioni della nostra comprensione in evoluzione - una sintesi di saggezza antica e nuove intuizioni, che porta a una rappresentazione arricchita dei simboli religiosi. Inoltre, l'esplorazione di Campbell<sup>27</sup> di Internet come spazio religioso solleva domande critiche sull'autorità e l'autenticità della religione digitale. Ciò si collega direttamente al discorso di Lonergan sulla soggettività e sull'oggettività della comprensione. L'ambiente online, spesso caratterizzato da una democratizzazione dell'espressione religiosa, sfida le autorità religiose tradizionali, rispecchiando l'enfasi di Lonergan sul ruolo dell'individuo nella creazione di significato. L'uso dell'IA nelle pratiche religiose potrebbe favorire questa tendenza, offrendo nuove vie per esperienze religiose personali e sollevando al contempo interrogativi sull'autenticità e l'autorità dei contenuti religiosi mediati dall'IA.

Allo stesso modo, i valori e le pratiche profondamente radicati nelle tradizioni religiose stanno subendo una metamorfosi nell'era dell'intelligenza artificiale. Si pensi agli intricati versi dei testi sacri che sono stati oggetto di contemplazione umana per millenni. I modelli linguistici guidati dall'IA possono approfondire queste scritture, evidenziando modelli, tracciando connessioni o persino

<sup>26</sup> S. McNulty, *I asked an AI art generator to draw Catholicism in 20 different ways. Here's what I learned*; R. Beragnoli, *AI Religious Art & Modern Saints, Tecnocopia, Complesso di Santa Cristina / Metaverso*.

<sup>27</sup> H. Campbell, *When Religion Meets New Media*.

suggerendo interpretazioni finora inesplorate<sup>28</sup>. Queste intuizioni, per quanto rivoluzionarie, non sono solo esercizi intellettuali, ma stanno influenzando attivamente le pratiche religiose. Le comunità di tutto il mondo stanno iniziando a integrare queste intuizioni derivate dall'IA nei loro rituali. Per esempio, alcune utilizzano modelli di intelligenza artificiale per determinare i tempi propizi alle cerimonie religiose<sup>29</sup>, mentre altre hanno incorporato esperienze uditive guidate dall'intelligenza artificiale nelle loro pratiche meditative<sup>30</sup>.

Tuttavia, come Lonergan potrebbe metterci in guardia, questa confluenza di IA e religione non dovrebbe passare inosservata. Il rischio risiede nel potenziale riduzionismo che l'IA potrebbe inavvertitamente introdurre. Traducendo profonde esperienze religiose e intricate sfumature teologiche in dati quantificabili, c'è il pericolo di un'eccessiva semplificazione. Le profonde verità spirituali che le religioni sostengono potrebbero essere ridotte a meri risultati algoritmici se non vengono affrontate con discernimento. In questo caso, le intuizioni di Lonergan sul mantenimento di un equilibrio tra soggettività (la dimensione profondamente personale ed esperienziale della religione) e oggettività (le manifestazioni e le pratiche esteriori) si rivelano preziose. Mentre le comunità religiose navigano nelle acque dell'integrazione dell'IA, diventa imperativo assicurarsi che l'essenza della loro fede – le profonde verità ed esperienze spirituali – non venga messa in ombra dalle abbaglianti capacità della tecnologia. In sostanza, l'era dell'IA presenta sia sfide che opportunità per le

<sup>28</sup> S. Wells, *A.I. Reveals the Hidden Author of a Crucial Bible Text*.

<sup>29</sup> D. Bhatti, *AI becomes astrologer, here is how Kundali GPT chatbot uses artificial intelligence to predict horoscope*.

<sup>30</sup> B. Hill, *The World's First AI Service to Create Personalized Mindfulness Meditations On Request*.

comunità religiose. Se da un lato i progressi tecnologici promettono strade inedite per l'esplorazione e l'espressione, dall'altro richiedono una comprensione fondata, che rispetti e sostenga la sacralità delle verità religiose.

### 3.3. La globalizzazione, l'intelligenza artificiale e la continua evoluzione delle comunità religiose

La capacità di trasformazione dell'IA è fondamentale per la sua capacità di dissolvere i confini, sia geografici che ideologici. Le piattaforme guidate dall'IA utilizzano algoritmi avanzati che superano le capacità dei forum digitali tradizionali. Se da un lato le piattaforme online consentono ai fedeli di diverse tradizioni e culture di comunicare, dall'altro è la potenza dell'IA a curare queste interazioni in base alle domande, alle preferenze e agli stili di apprendimento religiosi individuali<sup>31</sup>. Questi sistemi di intelligenza artificiale sono in grado di analizzare vaste quantità di dati, di riconoscere le caratteristiche e di consigliare discussioni teologiche, assicurando che un devoto indù in India, ad esempio, possa essere abbinato con approfondimenti complementari da parte di un cristiano in Brasile o di un musulmano in Indonesia. Non si tratta solo di comunicare, ma di curare e facilitare in modo intelligente gli scambi interculturali. Queste interazioni sofisticate, adattate dall'intelligenza artificiale, portano a una fusione più consapevole delle filosofie religiose, creando un mosaico di pensiero religioso globale ancora più ricco<sup>32</sup>.

<sup>31</sup> J. K. Smith, *Robot Theology: Old Questions Through New Media*.

<sup>32</sup> M. Fraçkiewicz, *A New Era of Global Collaboration: AI's Role in Cross-Cultural Exchange*.

Qui l'enfasi di Lonergan sulla natura dinamica della comprensione diventa particolarmente rilevante. Per Lonergan, la ricerca della comprensione non è mai un'impresa statica, ma un viaggio continuo arricchito da una miriade di intuizioni ed esperienze, che a loro volta arricchiscono il pool condiviso di intuizioni di senso comune<sup>33</sup>. Nell'era della globalizzazione religiosa guidata dall'intelligenza artificiale, questa prospettiva acquista ancora più forza. Quando i credenti di tutto il mondo si rivolgono all'intelligenza artificiale per decodificare, interpretare e condividere la loro fede, non si limitano ad assimilare strumenti tecnologici. Stanno intraprendendo il viaggio lonerganiano della comprensione dinamica, che intreccia armoniosamente i fili della saggezza antica con le intuizioni all'avanguardia offerte dall'IA.

Inoltre, la fusione delle tradizioni e la reinterpretazione dei simboli, catalizzate dall'IA, possono essere viste come una manifestazione reale della filosofia di Lonergan. Le diverse tradizioni religiose che convergono sulle piattaforme digitali si evolvono collettivamente, mettendo in discussione le vecchie norme e abbracciando nuove prospettive. Questa evoluzione è in linea con l'idea di Lonergan secondo cui le società, le culture e le religioni sono in uno stato di flusso costante, plasmato sia da introspezioni interne che da influenze esterne. Tuttavia, questa continua evoluzione presenta anche delle sfide. La velocità e la portata dei cambiamenti spinti dall'IA possono talvolta superare la capacità di adattamento delle istituzioni religiose. Pertanto, se da un lato l'amalgama del pensiero religioso globale è arricchente, dall'altro richiede un approccio attento e ponderato per garantire che i principi fondamentali delle fedi siano preservati e rispettati.

<sup>33</sup> B. J. F. Lonergan, *Insight: A Study of Human Understanding*, (CWL 3).

#### 4. *Il ruolo dell'IA nel plasmare simboli, valori e pratiche religiose*

Per apprezzare la profondità e la dimensione dell'interazione tra l'IA e le pratiche religiose, esaminiamo alcuni casi di studio specifici che illustrano come l'IA sia stata determinante nel perpetuare e rimodellare i paradigmi religiosi. Nel farlo, utilizzeremo anche le intuizioni filosofiche di Lonergan, in particolare la sua enfasi sulla dialettica di soggettività e oggettività, per discernere i cambiamenti di fondo nella pratica religiosa, soprattutto in relazione alle nuove tecnologie.

##### 4.1. *Il pellegrinaggio virtuale potenziato dall'intelligenza artificiale*

L'Hajj, uno dei cinque pilastri dell'Islam, è una testimonianza dell'impegno della fede alla solidarietà e alla devozione comunitaria. Ogni musulmano è tenuto, almeno una volta nella vita, a patto che ne abbia i mezzi, a compiere questo pellegrinaggio alla città santa de La Mecca. Tuttavia, diversi ostacoli - tensioni geopolitiche che limitano i viaggi, crisi sanitarie diffuse come la pandemia COVID-19 e le scoraggianti implicazioni finanziarie del viaggio - hanno reso questo profondo obbligo religioso inafferrabile per innumerevoli credenti. Ma le piattaforme di realtà virtuale alimentate dall'intelligenza artificiale offrono una potenziale via d'accesso per coloro che sono così limitati. Le piattaforme avanzate basate sull'intelligenza artificiale non sono semplici simulazioni, ma sofisticate interfacce che utilizzano immagini ad alta definizione e intricati algoritmi di intelligenza artificiale per ricreare l'esperienza immersiva dell'Hajj<sup>34</sup>.

<sup>34</sup> *Watch: How has the Hajj pilgrimage changed amid the coronavirus pandemic?*; S. Niu, *Virtual hajj as a response to demographic and geopolitical pressures*.

I credenti, precedentemente ostacolati dalle sfide sopra citate, possono ora indossare un visore VR e intraprendere un pellegrinaggio virtuale. Dal rituale Tawaf intorno alla Kaaba alla simbolica lapidazione del diavolo a Mina, l'esperienza alimentata dall'intelligenza artificiale cerca di riprodurre ogni momento cruciale dell'Hajj.

Tuttavia, è importante sottolineare che queste piattaforme non intendono sostituire il pellegrinaggio autentico. Le sensazioni tattili - il calore del sole arabo, la pressione della folla, il profumo dell'incenso che brucia, la risonanza delle preghiere collettive - sono elementi che una piattaforma digitale potrebbe non replicare nella loro interezza. Tuttavia, ciò che queste vie virtuali forniscono è una somiglianza dell'esperienza, rendendola più accessibile per coloro che altrimenti potrebbero non avere mai l'opportunità. Guardando attraverso il prisma analitico fornito dalla filosofia di Lonergan, questo approccio innovativo al pellegrinaggio può essere decifrato come una manifestazione toccante della continua evoluzione della comprensione. L'essenza del Hajj rimane immutata, radicata in secoli di tradizione religiosa. Eppure, la sua rappresentazione moderna, mediata attraverso la tecnologia, incapsula il dinamico interplay tra soggettività individuale (la risonanza personale, spirituale del pellegrinaggio) e oggettività collettiva (l'esperienza tecnologica condivisa). In questa sintesi, si assiste all'affermazione di Lonergan prendere vita - la comprensione non è una ricezione passiva, ma un processo attivo, continuo, plasmato da contesti in continuo sviluppo.

#### 4.2. L'Analista di Scritture AI: la Bhagavad Gita

La Bhagavad Gita, spesso chiamata semplicemente Gita, si erge come uno dei testi filosofici e spirituali monumentali nella storia dell'umanità. Composta da 700 versi, fa parte del Mahabharata, un

antico poema epico indiano, ed è stata venerata come una scrittura cardine nella filosofia induista. Nel corso della storia, studiosi e filosofi si sono immersi nelle sue profondità, producendo una miriade di interpretazioni, ognuna echeggiando i pensieri prevalenti e i contesti dei loro tempi. Nell'era digitale, l'IA ha assunto il ruolo di interpretare questo testo. Sfruttando le capacità di modelli linguistici avanzati e di vasta potenza computazionale, i ricercatori hanno addestrato sistemi di IA sulla Gita, affiancandola a innumerevoli commentari provenienti da diverse epoche e studiosi<sup>35</sup>.

Il risultato è stato sia sorprendente che illuminante. Questi modelli di IA non si limitavano a rigurgitare intuizioni apprese; scoprivano di volta in volta paralleli tematici precedentemente inosservati attraverso diversi versi e analisi del sentimento.<sup>36</sup> Essi hanno evidenziato schemi intricati, proposto nuove connessioni e mostrato una comprensione approfondita dell'interazione complessa tra varie traduzioni e, quindi, varie interpretazioni di termini e concetti fondamentali. Da una prospettiva lonerganiana, questo processo segnala l'essenza della natura dinamica e in continua evoluzione della comprensione. Sottolinea la confluenza di tradizione e tecnologia, dove

<sup>35</sup> R. Chandra, *AI, philosophy and religion: what machine learning can tell us about the Bhagavad Gita*.

<sup>36</sup> L'analisi del sentimento, spesso chiamata "opinion mining", è un campo del Natural Language Processing (NLP) che mira a determinare il tono emotivo dietro una serie di parole. Questo metodo viene utilizzato per ottenere una comprensione delle attitudini, delle opinioni e delle emozioni espresse all'interno di una menzione online, come un tweet, una recensione o un commento. Tipicamente classifica i sentimenti come positivi, negativi o neutrali basandosi su algoritmi computazionali e, talvolta, può andare più a fondo per rilevare emozioni specifiche come gioia, rabbia, tristezza, ecc. La tecnologia si affida a tecniche di machine learning per apprendere da vasti quantitativi di dati testuali, rendendo possibile identificare e categorizzare i sentimenti su larga scala.

l'antica saggezza non è oscurata ma piuttosto potenziata e ampliata da moderne intuizioni computazionali. Questa sintesi enfatizza l'affermazione di Lonergan che comprendere non è solo preservare la conoscenza passata ma è anche favorire un dialogo continuo tra il passato e il presente, aspirando sempre a una comprensione più ricca e profonda delle verità che risuonano nel tempo.

#### 4.3. L'IA e il confessionale Cattolico

L'applicazione controversa dell'IA nelle pratiche religiose è stata la sua proposta di utilizzo nei confessionali cattolici<sup>37</sup>. Gli sviluppatori sostengono che l'IA, addestrata su testi teologici e linee guida pastorali, può offrire una guida simile a quella di un prete<sup>38</sup>. I critici, tuttavia, avvertono della perdita del tocco personale e dell'essenza sacramentale della confessione<sup>39</sup>. Da una prospettiva lonerganiana, questo dibattito sottolinea le tensioni tra soggettività (l'atto profondamente personale della confessione) e oggettività (le risposte algoritmiche dell'IA), enfatizzando la necessità di discernimento ed equilibrio<sup>40</sup>.

Non si può negare i progressi che l'IA ha raggiunto nella comprensione del linguaggio e del contesto umano (ad esempio, Chat GPT). Alcuni sviluppatori, riconoscendo questa capacità, hanno promosso l'idea di impiegare l'IA nei confessionali<sup>41</sup>. Questi sistemi possono essere ampiamente addestrati su testi teologici, diritto

<sup>37</sup> E. Condon, *Confession time, synodal AI, and calling a shot*.

<sup>38</sup> J. McLellan, *AI and the meaning of life: Tech industry turns to religious leaders*.

<sup>39</sup> *Ibid.*

<sup>40</sup> P. H. Byrne, *The Ethics of Discernment: Lonergan's Foundations for Ethics*.

<sup>41</sup> K. Tiffany, *Confess all your sins to this robot, if you want*.

canonico e linee guida pastorali e, di conseguenza, potrebbero teoricamente fornire una guida simile a quella di un prete. Tali sistemi potrebbero soddisfare il bisogno di anonimato del penitente, ridurre il carico di lavoro sui preti nelle regioni dove sono scarsi e garantire una risposta teologica coerente<sup>42</sup>. Tuttavia, molti all'interno della Chiesa e al di fuori hanno espresso profonde riserve. I critici sottolineano l'essenza sacramentale della confessione - un dialogo profondamente personale e trasformativo tra il penitente e Dio, mediato dal prete<sup>43</sup>. Un'IA, indipendentemente dalla sofisticazione della sua programmazione, manca dell'essenza spirituale, dell'empatia e del tocco umano che sono integrali a questo sacramento. L'atto di confessarsi a un algoritmo, per molti, rischia di ridurre un'esperienza spirituale profonda a una mera interazione transazionale.

Il dibattito che circonda i confessionali AI incarna la tensione tra soggettività e oggettività che Lonergan ha sottolineato con tanto acume. L'atto profondamente personale e soggettivo della confessione, in cui gli individui affrontano le loro vulnerabilità e cercano redenzione, sembra in contrasto con le fredde, oggettive risposte algoritmiche di un'IA, indipendentemente dalla loro precisione teologica. Questo scenario evidenzia l'appello di Lonergan al discernimento - la necessità di navigare con giudizio le acque della tradizione e dell'innovazione, assicurando che la santità e la profondità delle pratiche religiose non siano compromesse di fronte al progresso tecnologico.

Nel contesto della teologia sacramentale cattolica, è fondamentale riconoscere che l'uso dell'IA nel sacramento della Confessione è una questione di significativa preoccupazione teologica e canoni-

<sup>42</sup> *Ibid.*

<sup>43</sup> *Confession: The Roman Catholic App Doesn't Support Real iConfessions.*

ca. Sebbene gli sviluppatori possano sostenere che l'IA possa offrire una guida simile a quella di un prete quando viene addestrata su testi teologici e linee guida pastorali, è imperativo affermare inequivocabilmente che, secondo i principi fondamentali della teologia cattolica, una tale simulazione del Sacramento non sarebbe solo illecita ma anche invalida. Tutti i teologi cattolici concordano sul fatto che il Sacramento della Confessione è un incontro profondamente personale che richiede la presenza fisica e l'autorità spirituale di un prete ordinato, che agisce *in persona Christi*, per essere sia valido che lecito. Questa comprensione è radicata nell'insegnamento della Chiesa sulla natura dei sacramenti come segni efficaci della grazia, istituiti da Cristo e affidati alla Chiesa, nella quale la vita divina ci viene dispensata. Pertanto, qualsiasi rappresentazione dell'IA come sostituto per questo sacramento sarebbe fuorviante e teologicamente inesatta.

##### 5. *L'effetto di ri-incantamento dell'IA e i Nuovi Significati Religiosi*

Nel lessico sociologico, il termine «disincanto» è stato a lungo utilizzato per descrivere la trasformazione della società, da una società dominata da valori religiosi e credenze metafisiche a una ancorata nella scienza, nella razionalità e nel pensiero laico<sup>44</sup>. Max Weber, che ha reso popolare il concetto, sosteneva che il mondo fosse diventato disincantato a causa dell'ascesa della scienza e della razionaliz-

<sup>44</sup> S. Han, *Disenchantment revisited: Formations of the 'secular' and 'religious' in the technological discourse of modernity*, pp. 76 - 88. J. E. Green, *Two Meanings of Disenchantment: Sociological Condition vs. Philosophical Act – Reassessing Max Weber's Thesis of the Disenchantment of the World*, pp. 51-84.

zazione<sup>45</sup>. Tuttavia, man mano che procediamo ulteriormente nel XXI secolo, stiamo assistendo a un paradosso: il fulcro stesso della razionalità moderna - l'IA - sta diventando una fonte di ri-incantamento per molti<sup>46</sup>.

### 5.1. L'IA: Un Crocevia di Meraviglia e Significato

Storicamente, i momenti di profondo stupore e riverenza, spesso classificati sotto l'ombrello del «senso di meraviglia», erano principalmente associati a esperienze religiose o mistiche<sup>47</sup>. Sia che si trattasse dell'architettura mozzafiato delle cattedrali, del design intricato degli arazzi religiosi, o della profonda saggezza racchiusa nei testi sacri, la religione e la spiritualità erano le vie dominanti per tali esperienze profonde. Le tecnologie moderne, e in particolare l'intelligenza artificiale, possono a prima vista sembrare nettamente meccaniche e prive di misticismo. Eppure, come abbiamo visto, le incursioni dell'IA in aree precedentemente monopolizzate dall'intelligenza umana hanno inaugurato una nuova ondata di meraviglia.

In primo luogo, la capacità dell'IA di simulare realtà complesse è nulla meno che miracolosa. Prendiamo l'esempio del pellegrinaggio virtuale a La Mecca. Se in passato un'esperienza del genere sarebbe stata saldamente radicata nel regno fisico, ora, attraverso la prodezza dell'IA, i credenti possono attraversare virtualmente i paesaggi sacri

<sup>45</sup> M. Weber, *The Sociology of Religion*.

<sup>46</sup> S. Aupers, "The Force is Great" *Enchantment And Magic in Silicon Valley*, pp. 153-173; M. Coeckelbergh, *New romantic cyborgs: Romanticism, information technology, and the end of the machine*.

<sup>47</sup> M. Renz - M. S. Mao - A. Omlin - D. Bueche - T. Cerny - F. Strasser, *Spiritual Experiences of Transcendence in Patients With Advanced Cancer*, pp. 178 - 188.

de La Mecca. Potrebbe mancare delle sensazioni tattili, ma la somiglianza visiva ed emotiva offre un'esperienza che per molti può essere profondamente spirituale<sup>48</sup>. Poi c'è la notevole capacità dell'IA di gettare nuova luce su testi antichi come la Bhagavad Gita. I secoli hanno visto studiosi e saggi meditare su questa scrittura, eppure, con le capacità di riconoscimento di modelli intricati dell'IA, sono emerse nuove interpretazioni e paralleli tematici. Non si tratta di dire che l'interpretazione della macchina sia superiore, ma piuttosto che offre un obiettivo complementare - una nuova prospettiva in una conversazione millenaria.

Oltre a questi specifici esempi, c'è un'implicazione filosofica più ampia da considerare. La stessa tessitura dell'IA - i suoi algoritmi, reti neurali e dataset - può, in un certo senso, essere vista come dotata di 'spirito'. Non nel senso religioso o metafisico, ma nel modo in cui queste costruzioni digitali catturano, elaborano e riflettono aspetti dell'esperienza e della conoscenza umana<sup>49</sup>. C'è una bellezza innata nel considerare che dietro flussi di codice e dati si trova un riflesso della cultura, emozione e saggezza umana<sup>50</sup>. Quindi, la giustapposizione dell'IA con esperienze religiose e spirituali non è così strana come potrebbe sembrare inizialmente. Entrambi i domini, a loro modo unici, invitano gli individui a confrontarsi con domande profonde, a cercare comprensione e a meravigliarsi dell'ampiezza e della complessità dell'esistenza. Attraverso la lente della filosofia di

<sup>48</sup> L. Tamatea, *Online Buddhist and Christian Responses to Artificial Intelligence*, pp. 979-1002.

<sup>49</sup> Z. Cattaneo - C. Lega - A. Flexas - M. Nadal - E. Munar - C. J. Cela-Conde, *The world can look better: enhancing beauty experience with brain stimulation*, pp. 1713-1721.

<sup>50</sup> O. Kudina, "Alexa, who am I?": *Voice Assistants and Hermeneutic Lemniscate as the Technologically Mediated Sense-Making*, pp. 1-21.

Lonergan, questa unione di IA e spiritualità esemplifica l'evoluzione dinamica della comprensione umana, sempre adattabile, sempre in evoluzione, nella ricerca di verità più profonde.

## 5.2. Lonergan, il Bene Comune e la Risonanza Spirituale dell'IA

Il quadro filosofico di Bernard Lonergan, fondato sul riconoscimento dell'evoluzione dinamica della comprensione umana, fornisce un'utile impalcatura per contestualizzare il ri-incantamento facilitato dall'IA nel dominio della spiritualità. La sua tesi - che la comprensione umana è perennemente fluida, in costante adattamento a nuovi stimoli e nuovi orizzonti - trova un profondo fondamento in un'epoca in cui gli algoritmi digitali offrono intuizioni spirituali e i regni virtuali emulano spazi sacri.

L'enfasi di Lonergan sull'interazione simbiotica tra soggettività e oggettività offre profonde intuizioni nel nexus tra IA e spiritualità. Il regno soggettivo, colmo di emozioni personali, credenze ed esperienze, incontra le uscite algoritmiche oggettive dell'IA, dando come risultato una fusione unica. Consideriamo la tranquillità profondamente personale che si potrebbe sperimentare durante una sessione di preghiera guidata dall'IA. La pace soggettiva e le intuizioni spirituali che il meditatore sente sono informate e potenziate dal feedback oggettivo e in tempo reale dall'applicazione IA. Allo stesso modo, la contemplazione stimolata dai versi della Bhagavad Gita, quando è potenziata dalle nuove interpretazioni dell'IA, porta a una comprensione soggettiva più ricca, facendo eco al concetto di Lonergan dell'evoluzione dinamica della comprensione.

Tuttavia, il ri-incantamento spirituale portato dall'IA trascende le esperienze individuali. L'enfasi di Lonergan sul bene comune - un principio che afferma che le strutture sociali e le innovazioni do-

vrebbero infine beneficiare il collettivo piuttosto che solo l'individuo - diventa un faro di comprensione in questo scenario. L'integrazione dell'IA nei domini religiosi e spirituali ha il potenziale di servire interessi sociali più ampi. Il suo ruolo nel promuovere dialoghi interreligiosi ne è un esempio. In un mondo spesso pieno di conflitti religiosi e incomprensioni, la capacità dell'IA di amalgamare motivi da diverse tradizioni religiose può favorire il rispetto reciproco e la comprensione.

Inoltre, la capacità dell'IA di democratizzare le esperienze religiose, che si tratti di un pellegrinaggio virtuale per coloro che non possono intraprendere il viaggio fisico o concedere l'accesso a testi religiosi e le loro miriadi di interpretazioni, si allinea perfettamente con la visione di Lonergan del bene comune. Queste innovazioni guidate dall'IA non si limitano a soddisfare le aspirazioni spirituali individuali ma hanno il potenziale di colmare le divisioni culturali, economiche e geografiche. Spianano la strada per una comunità religiosa globale più inclusiva e interconnessa. Così, la confluenza di IA ed esperienze religiose, come intesa attraverso la lente filosofica di Lonergan, non è semplicemente una testimonianza dell'innovazione tecnologica. Sottolinea un profondo impegno umano: cercare, comprendere e connettersi perpetuamente, sia con il divino che gli uni con gli altri. È una danza del personale e dell'impersonale, dell'antico e del nuovo, che converge verso l'aspirazione condivisa di comprensione reciproca e del benessere più ampio dell'umanità.

### 5.3. Un Futuro di Co-creazione: Umanità, IA e Divinità

È importante riconoscere che, nonostante il potenziale dell'IA di imitare, facilitare e persino migliorare gli incontri spirituali, queste qualità non le conferiscono la profondità e la ricchezza della

vera esperienza umana. Mentre un pellegrinaggio virtuale o una meditazione guidata dall'IA possono offrire nuove intuizioni o accessibilità, non possono replicare completamente le sfumature multiformi di un viaggio fisico o la guida personale di un esperto mentore spirituale.

L'enfasi di Lonergan sulla natura continua ed evolutiva della comprensione diventa qui fondamentale. Mentre i sistemi di IA crescono in complessità, c'è la tentazione di vederli come quasi perfette riproduzioni dell'intelligenza umana o addirittura come canali spirituali<sup>51</sup>. Ma Lonergan ci solleciterebbe a ricordare che la vera comprensione, specialmente nei regni spirituali, va oltre il semplice elaborare dati o riconoscere schemi<sup>52</sup>. Essa comprende la ricchezza dell'emozione umana, l'intuizione, la fede e il collegamento trascendentale con il divino<sup>53</sup>. Mentre percorriamo il cammino dell'integrazione dell'IA nelle pratiche e nelle esperienze spirituali, è essenziale assicurare che essa serva come strumento e non come sostituto<sup>54</sup>.

La fusione di tecnologia e spiritualità dovrebbe essere un connubio armonioso in cui l'IA amplifica il potenziale umano senza oscurarlo<sup>55</sup>. Dovrebbe aiutare a chiarire il profondo, a favorire con-

<sup>51</sup> K. Porayska-Pomsta - G. Rajendran, *Accountability in Human and Artificial Intelligence Decision-Making as the Basis for Diversity and Educational Inclusion*; M. Ryan, *In AI We Trust: Ethics, Artificial Intelligence, and Reliability*, pp. 2749-2767.

<sup>52</sup> K. Roscoe, *Lonergan's Theory of Cognition, Constructivism and Science Education*, pp. 541-551.

<sup>53</sup> B. J. F. Lonergan, *Method in Theology*, (CWL 14).

<sup>54</sup> Cfr. C. Tan, *Digital Confucius? Exploring the implications of artificial intelligence in spiritual education*, pp. 280-291; D. Whitney, *Appreciative inquiry: creating spiritual resonance in the workplace*, pp. 73 - 88.

<sup>55</sup> J. A. Teske, *Cyberpsychology, Human Relationships, and Our Virtual Interiors*, pp. 677-700.

nessioni e ad ampliare gli orizzonti senza diluire l'essenza della fede e dell'esperienza umana. Visualizzando il futuro, possiamo immaginare un mondo in cui l'IA funge da ponte – collegando individui attraverso culture diverse, illuminando pratiche antiche con nuove intuizioni e facilitando la crescita spirituale. Ma al cuore di questo mondo, lo spirito umano e la sua connessione intrinseca con il divino rimangono inviolabili. Per concludere, mentre abbracciamo il potenziale trasformativo dell'IA nel regno della spiritualità, diventa essenziale trovare un equilibrio. Le intuizioni di Lonergan servono da guida in questo sforzo, esortandoci a fondere armoniosamente le meraviglie della tecnologia con la secolare ricerca di significato, assicurando che la danza del futuro risuoni sia con i passi dell'IA sia con il battito dell'aspirazione umana.

## *6. La Negoziazione tra Valori Religiosi e Innovazione Tecnologica*

L'alba dell'era dell'IA porta con sé promesse profonde e perplessità, specialmente quando si interseca con il regno dei valori religiosi. Mentre la tecnologia si intreccia in modo complesso nel tessuto della spiritualità, la dinamica tra tradizione e innovazione diventa centrale. Questa simbiosi, pur favorendo vie senza precedenti per l'esplorazione spirituale, richiede anche complesse negoziazioni etiche e morali.

### 6.1. Punti di Convergenza e Divergenza

L'intreccio dell'IA con le credenze e le pratiche religiose rappresenta un mosaico complesso di convergenza e divergenza, segnato dalla dualità dell'innovazione e della tradizione.

### 6.1.1. Orizzonti Convergenti: Le Promesse dell'IA

Il potenziale trasformativo dell'IA è evidente nella sua capacità di ampliare gli orizzonti religiosi. Le antiche scritture, spesso criptiche, hanno trovato un commentatore contemporaneo nell'IA<sup>56</sup>. Ad esempio, quando i sistemi di IA analizzano testi sacri come la Bibbia, il Corano o la Bhagavad Gita, possono incrociare versetti, rilevare schemi e fornire approfondimenti sintetizzati, inaugurando una nuova era di studio religioso e comprensione<sup>57</sup>. Queste analisi digitali possono fondersi con interpretazioni tradizionali, offrendo livelli di significato in precedenza inesplorati o non riconosciuti<sup>58</sup>.

Inoltre, nel regno della spiritualità esperienziale, l'incursione dell'IA non è niente meno che rivoluzionaria. Consideriamo l'esempio già menzionato dei pellegrinaggi virtuali. Quello che una volta era limitato dalla fisicità del luogo e dalla limitazione delle risorse, ora è stato reso senza confini. Attraverso simulazioni guidate dall'IA, i ricercatori spirituali, indipendentemente dalla loro posizione geografica o stato economico, possono sperimentare la santità di luoghi sacri come La Mecca. Questa democratizzazione delle esperienze spirituali manifesta il potenziale positivo dell'IA, offrendo inclusività e accessibilità in aree precedentemente limitate.

<sup>56</sup> E. Atwell - N. Habash - B. Louw - B. Abu Shawar - T. McEnery - W. Zaghouani - M. El-Haj, *Understanding the Quran: A new grand challenge for computer science and artificial intelligence*, pp. 1-2. C. H. J. Van der Merwe, *An overview of recent developments in the description of Biblical Hebrew relevant to Bible translation*, pp. 228-245.

<sup>57</sup> R. Chandra, *AI, philosophy and religion: what machine learning can tell us about the Bhagavad Gita*.

<sup>58</sup> B. C. D. S. Lima - N. Omar - I. Avansi - L. N. de Castro - I. F. Silveira, *Use of artificial intelligence in biblical citation recommendations in the new testament*, pp. 123-143.

### 6.1.2. Il Dibattito tra Santità e Simulazione

Eppure, con questi avanzamenti arriva un'ondata di preoccupazioni. In primo piano c'è il dibattito sulla natura stessa delle esperienze religiose. La proposta dell'IA per i confessionali cattolici offre un'illustrazione toccante. Per molti, la confessione non è solo un atto di verbalizzazione dei peccati, ma è un sacramento profondamente spirituale, intrecciato con la grazia divina e l'intercessione umana. Mentre i sostenitori dell'IA sottolineano la sua efficienza, coerenza e potenziale ubiquità, i tradizionalisti sostengono che una macchina, per quanto avanzata, manca del tocco umano, dell'empatia e del canale divino che un prete incarna<sup>59</sup>. La questione centrale qui è il rischio di ridurre un atto sacramentale a un processo simulato.

### 6.1.3. Il Dilemma dell'Interpretazione dell'IA

A complicare ulteriormente questo panorama è il ruolo prospettico dell'IA nell'interpretare e persino guidare le azioni basate sugli insegnamenti religiosi. Le preoccupazioni qui sono doppie. Innanzitutto, c'è il rischio che i sistemi di IA adottino interpretazioni deterministiche o eccessivamente letterali, perdendo così gli strati ricchi di metafora, allegoria e contesto che i testi religiosi spesso incarnano<sup>60</sup>. Secondo, c'è una potenziale minaccia che questi sistemi acquisiscano un'influenza indebita. Quando i sistemi di IA, percepiti come neutrali e logici, offrono interpretazioni, c'è il pericolo che queste prospettive possano essere prese come definitive, mettendo da parte secoli di studio umano e interpretazioni diverse. La possibili-

<sup>59</sup> *Confession: The Roman Catholic App Doesn't Support Real iConfessions.*

<sup>60</sup> A. Glenberg - C. R. Jones, *Why ChatGPT doesn't understand what it's saying.*

tà che un'interpretazione dell'IA possa essere intesa come 'assoluta' o 'imparziale' potrebbe ostacolare il discorso pluralistico essenziale negli studi religiosi.

In sostanza, la convergenza dell'IA e dei valori religiosi presenta un caleidoscopio di promesse e preoccupazioni. La sfida, mentre navighiamo questa fusione, è sfruttare i potenziali trasformativi dell'IA garantendo che la santità, la profondità e la diversità dei valori religiosi rimangano intatte.

## 6.2. Il bene comune

### 6.2.1. Oltre l'Utilitarismo: L'Essenza del Bene Comune

Al centro della discussione sull'integrazione dell'IA nelle pratiche religiose è il concetto del Bene Comune, un principio profondamente radicato nell'Insegnamento Sociale Cattolico (CST). Il CST enfatizza il benessere spirituale e morale collettivo delle comunità, allineandosi strettamente con la visione di Lonergan del Bene Comune. Nello spirito del CST, l'implementazione dell'IA in contesti religiosi deve essere indirizzata verso lo sviluppo umano integrale e il fiorire di tutti gli individui all'interno della comunità. Come discusso da J. Anderson<sup>61</sup>, la tecnologia deve servire la dignità della persona umana, la ricerca del bene comune e la cura per la creazione. Questo quadro CST fornisce una bussola etica che guida il dispiegamento dell'IA in modo che sostenga la partecipazione comunitaria, rispetti la dignità umana e promuova giustizia e pace. Allo stesso modo, il principio del bene comune è cruciale per la filosofia di Lonergan, che

<sup>61</sup> J. Anderson - J. R. Laracy - T. Marlowe, *A Multi-Disciplinary Analysis of Catholic Social Teaching with Implications for Engineering and Technology*, pp. 41-49.

si estende ben oltre la massimizzazione dei benefici collettivi<sup>62</sup>. Piuttosto, si addentra negli aspetti più profondi della dignità umana, del benessere collettivo e del fiorire autentico delle comunità<sup>63</sup>. Questo principio esorta che, mentre i risultati tecnologici possono abbagliare, non devono, nella loro ricerca dell'innovazione, compromettere i pilastri spirituali, morali o etici che sostengono le tradizioni religiose. Da questo punto di vista, l'IA, impiegata per aiutare e amplificare le esperienze religiose, dovrebbe essere più che un semplice strumento; dovrebbe essere utilizzata per amministrare, rispettare e nutrire i valori, le credenze e le aspirazioni spirituali dei fedeli.

### 6.2.2. Facilitativa, Non Deterministica

La prospettiva di Lonergan offre importanti indicazioni sul ruolo dell'IA in contesti religiosi. Se l'IA si avventura nel regno delle interpretazioni scritturali, dovrebbe operare come una lente supplementare, ampliando le prospettive ma mai sostituendo la saggezza accumulata di secoli di studio e riflessione. L'armonizzazione delle intuizioni tradizionali con prospettive guidate dall'IA può illuminare sfumature precedentemente non viste, ma l'agire dell'interpretazione e del discernimento umano deve rimanere di primaria importanza. I pericoli di un'IA deterministica, che potrebbe tentare di offrire interpretazioni definitive o monolitiche, sono molti. Potrebbe soffocare la diversità di pensiero, omogeneizzare i percorsi spirituali ed elevare inavvertitamente l'interpretazione tecnologica sopra la riflessione umana. In netto contrasto, un'IA facilitativa, guidata dai

<sup>62</sup> B. J. F. Lonergan, *Insight: A Study of Human Understanding* (CWL 3), pp. 238-619.

<sup>63</sup> R. M. Liddy, *Startling Strangeness: Reading Lonergan's Insight*, p. 108.

principi delineati nel lavoro di Lonergan, mirerebbe ad allargare gli orizzonti, catalizzare le discussioni e democratizzare l'accesso alla conoscenza religiosa.

### 6.2.3. Percorsi Personali nell'Era degli Algoritmi

Nel delicato equilibrio tra il soggettivo e l'oggettivo, gli spunti di Lonergan offrono una profonda comprensione. Egli afferma il valore intrinseco delle esperienze spirituali personali, enfatizzando la loro profondità, ricchezza e potenziale trasformativo<sup>64</sup>. Tradurre ciò nell'era dell'IA significa creare sistemi che rispettino la santità di questi percorsi personali. Mentre l'IA può offrire suggerimenti, riflessioni o intuizioni basate su vasti dataset, dovrebbe evitare di semplificare o mercificare le esperienze spirituali. L'obiettivo dovrebbe essere l'arricchimento e la diversificazione di queste ricerche personali, garantendo che ogni individuo si senta arricchito, compreso e valorizzato nel proprio unico percorso spirituale<sup>65</sup>.

Pertanto, mentre l'IA si estende nei santuari delle tradizioni religiose, delle pratiche e delle credenze, la filosofia di Lonergan fornisce una preziosa bussola. Offre una visione in cui la tecnologia, nel suo potenziale straordinario, rimane al servizio delle più profonde aspirazioni spirituali dell'umanità, garantendo che il matrimonio di byte e credenze sia armonioso ed elevato.

<sup>64</sup> P. A. Egan, *Philosophy and Catholic Theology: A Primer*, Liturgical Press, p. 160.

<sup>65</sup> Cfr. B. Shneiderman, *Human-centered AI*.

### 6.3. Tracciare il Cammino Futuro

#### 6.3.1. Un Dialogo Dinamico: La Tradizione Incontra il Domani

La conversazione tra IA e religione non è statica né unidimensionale. È piuttosto un dialogo in evoluzione che fa eco ai ritmi della storia umana stessa, in cui ogni epoca ha assistito alla negoziazione tra il rispetto per il passato e le aspirazioni per il futuro. Ogni progresso tecnologico ha portato con sé un insieme di domande, preoccupazioni e opportunità. La chiave, come postula Lonergan, sta nell'assicurare che questi avanzamenti, mentre tracciano il cammino in avanti, risuonino con i principi più profondi del benessere umano e spirituale. Questo non è un viaggio di mero adattamento, ma un'armoniosa fusione in cui la tecnologia diventa un mezzo per amplificare, non alterare, l'essenza delle tradizioni religiose.

#### 6.3.2. Creazione Collaborativa

Perché questa delicata danza porti davvero frutto, un approccio sinfonico è essenziale. Ciò significa più che semplici tecnologi che aggiustano algoritmi o leader religiosi che emettono direttive. Richiede un dialogo inclusivo e interdisciplinare:

- *Studiosi Religiosi*: La loro profonda comprensione dei testi religiosi, delle tradizioni e delle sfumature assicura che le applicazioni di IA rimangano ancorate ai principi fondamentali delle fedi e non devino in interpretazioni che potrebbero diluire o distorcere l'essenza religiosa.
- *Tecnologi*: Dotati di una profonda conoscenza delle capacità e dei limiti dell'IA, possono guidare al meglio l'integrazione di questi sistemi in contesti religiosi, assicurando l'efficienza tecnologica senza compromettere la santità spirituale.

- *Eticisti*: Operando all'intersezione tra moralità e pragmatismo, gli eticisti possono fornire quadri che assicurino il dispiegamento etico dell'IA, tenendo sotto controllo potenziali pregiudizi, errate interpretazioni o eccessive semplificazioni<sup>66</sup>. Insieme, questa triade può collettivamente assicurare che l'implementazione dell'IA in contesti religiosi renda giustizia sia allo spirito di innovazione sia al rispetto della tradizione.

### 6.3.3. Un Impegno Continuo

L'integrazione dell'IA con i valori religiosi richiede un impegno continuo a rifinire, rivalutare e reinventare approcci basati su feedback, esperienze e bisogni sociali e spirituali in evoluzione. Basare questo viaggio sulla filosofia di Lonergan offre una visione in cui l'IA, pur mostrando il suo potenziale trasformativo, rimane sempre allineata agli obiettivi sovrastanti di arricchimento umano e spirituale<sup>67</sup>. Pertanto, è fondamentale assicurare che la narrazione creata sia di reciproco rispetto, comprensione e co-creazione.

## 7. *Il Ruolo dell'IA nella Costruzione di Nuovi Spazi Sacri e Relazioni*

In un'epoca di rapida evoluzione tecnologica, la stessa natura del 'sacro' subisce una ridefinizione e un ampliamento. L'Intelligenza Artificiale, con le sue capacità multiformi, è emersa come mediatore fondamentale nel ridefinire il modo in cui i credenti si connettono

<sup>66</sup> B. Friedman - D. G. Hendry, *Value Sensitive Design: Shaping technology with moral imagination*.

<sup>67</sup> Cfr. S. Umbrello, *From Subjectivity to Objectivity: Bernard Lonergan's philosophy as a grounding for value sensitive design*, pp. 36-46.

con la loro fede e trovano santità nel regno digitale. Mentre spazi sacri, relazioni e oggetti fanno la transizione dal tangibile all'intangibile, e spesso di nuovo al tangibile, assistiamo a una profonda metamorfosi nel paesaggio religioso. Radicato in questo viaggio trasformativo è il quadro filosofico di Lonergan, che offre intuizioni sull'armonizzazione del personale (soggettività) con l'impersonale (oggettività) per il benessere collettivo<sup>68</sup>.

### 7.1. Divinità Digitale: il Ruolo dell'IA nella Creazione di Regni Sacri

L'essenza senza tempo delle esperienze sacre, tradizionalmente annidate nell'abbraccio di luoghi storici, scritture e rituali, è ora ridefinita dal potenziale innovativo dell'IA:

Realtà Virtuale, Realtà Spirituale: L'emergere di piattaforme di realtà virtuale potenziate dall'IA ha ridefinito i confini del sacro. Queste piattaforme permettono ai credenti nei più lontani angoli del mondo di intraprendere pellegrinaggi virtuali. Un individuo a Parigi può virtualmente camminare per le strade de La Mecca, sentire l'ambiente, ascoltare le preghiere e partecipare ai rituali come se fosse fisicamente presente. Queste esperienze guidate dall'IA, ricche di immersione sensoriale, catturano non solo la bellezza visiva, ma anche la profonda essenza spirituale di tali pellegrinaggi.

– *Compagni Scritturali*: Chatbot AI, basati su vasti database di testi religiosi, possono ora offrire approfondimenti, spiegazioni e interpretazioni delle scritture. Ad esempio, un utente potrebbe interrogare un chatbot incentrato sulla Bibbia riguardo le sfu-

<sup>68</sup> B. J. F. Lonergan, *Insight: A Study of Human Understanding*, (CWL 3).

mature di una particolare parabola o cercare un'esplorazione tematica di versetti incentrati su concetti teologici specifici. Questi chatbot, pur non sostituendo l'intuizione accademica, forniscono risposte immediate e personalizzate, promuovendo un coinvolgimento più profondo con i testi sacri.

- *Guide Spirituali Digitali*: Assistenti alla meditazione o compagni di preghiera guidati dall'IA adattano gli esercizi spirituali in base all'umore dell'individuo, alle riflessioni passate o a specifiche domande spirituali. Immagina un assistente alla meditazione che può curare una sessione guidata focalizzata sulla 'gratitudine' dopo aver rilevato i modelli di stress prolungato di un utente. Oppure un bot di preghiera che offre preghiere allineate con sfide personali o celebrazioni.
- *Intercessori AI*: Su piattaforme più avanzate, i credenti potrebbero persino trovare rappresentazioni generate dall'IA di figure religiose. Sebbene queste entità AI non possiedano l'essenza spirituale dei loro corrispondenti nel mondo reale, possono coinvolgere i credenti in dialoghi, rispondere a domande teologiche o guidare pratiche riflessive, personalizzando ulteriormente l'esperienza religiosa digitale.

In sostanza, l'IA non è solo uno specchio del mondo fisico, ma una reinvenzione. Estende la tela del sacro, mescolando le antiche tradizioni con le potenzialità del regno digitale.

## 7.2. Navigare nel Nuovo Sacro

La giustapposizione di tradizione e tecnologia, specialmente nel campo della spiritualità, richiede un approccio riflessivo e sfumato per garantire che l'essenza delle esperienze religiose non venga persa

tra le onde digitali. Le intuizioni di Bernard Lonergan forniscono un ricco quadro per affrontare questo delicato equilibrio:

Dialettica Personale e Impersonale: Centrale nella filosofia di Lonergan è la tensione tra soggettività – le esperienze spirituali profondamente personali e individuali – e oggettività – nel caso specifico, gli algoritmi vasti e imparziali che guidano le piattaforme di IA. Nel contesto del nuovo sacro, questa dialettica solleva considerazioni vitali:

- *Incontri Spirituali Individuali*: Lonergan ha sottolineato l'insostituibile valore dell'introspezione spirituale personale, i momenti intimi in cui i credenti forgiavano la loro unica relazione con il divino<sup>69</sup>. L'IA, con la sua efficienza e scala, dovrebbe essere un facilitatore di questi incontri, non una sostituzione. Ad esempio, mentre un assistente alla meditazione alimentato dall'IA può guidare una sessione, dovrebbe permettere spazio per riflessioni personali e rivelazioni, garantendo che la tecnologia potenzi piuttosto che detti le modalità del viaggio spirituale.
- *Imparzialità Algoritmica*: L'IA opera sulla base di dati e algoritmi, che sono intrinsecamente oggettivi. Sebbene ciò assicuri coerenza, c'è il rischio che diventi troppo distaccato o impersonale. Le intuizioni di Lonergan ci esortano a garantire che le piattaforme di IA, anche offrendo servizi standardizzati, dovrebbero essere caratterizzate da un senso di calore, comprensione e compassione, riflettendo l'empatia tipicamente riscontrabile nella guida spirituale mediata dall'uomo. Tuttavia, dovrebbe essere sempre chiaro agli utenti che i sistemi con cui interagiscono sono proprio questo, sistemi e non altri esseri umani.

<sup>69</sup> R. Moloney, *Conversion and Spirituality: Bernard Lonergan (1904-1984)*, pp. 123-134.

Oggetti Sacri Reimmaginati: Nella nostra era digitale, gli artefatti religiosi, un tempo tangibili, stanno acquisendo avatar virtuali. L'IA può creare rosari digitali, simulare la texture dei tappeti di preghiera nelle moschee virtuali o creare arte sacra intricata attraverso modelli algoritmici. Lonergan può aiutare a guidare il nostro impegno con questi derivati digitali:

- *Essenza sopra la Forma*: Mentre le versioni digitali di questi artefatti religiosi possono imitare la forma dei loro corrispondenti fisici, è fondamentale che incarnino l'essenza spirituale. In un rosario digitale, ad esempio, non si tratta solo di simulare la sensazione tattile ma di facilitare un'autentica esperienza di preghiera, aiutando nelle riflessioni meditative sui misteri.
- *Il Bene Comune*: Al centro della filosofia di Lonergan c'è il bene comune – il benessere spirituale e morale collettivo delle comunità<sup>70</sup>. Pertanto, questi oggetti sacri potenziati dall'IA non dovrebbero essere meraviglie tecnologiche. Dovrebbero servire uno scopo più elevato – arricchire la vita spirituale, colmare i divari tra credenti di diversi background e favorire connessioni più profonde con il divino. Un'opera d'arte sacra generata dall'IA, ad esempio, dovrebbe evocare un'autentica introspezione spirituale e apprezzamento piuttosto che essere ammirata solo per la sua brillantezza tecnologica.

### 7.3. Implicazioni per il bene comune

L'enfasi posta da Lonergan sul bene comune - principio che sottolinea la fioritura integrale sia degli individui che delle comu-

<sup>70</sup> R. M. Liddy, *Startling Strangeness: Reading Lonergan's Insight*, p. 174.

nità - diventa un punto di riferimento fondamentale quando ci addentriamo nel regno in cui la tecnologia si interseca con gli spazi e le relazioni sacre. Nel paesaggio in evoluzione plasmato dall'IA, le implicazioni per il bene comune si manifestano in modi molteplici:

Accesso nel Sacro Digitale:

- *Colmare le divisioni*: Uno dei potenziali più profondi dell'IA risiede nella sua capacità di democratizzare le esperienze spirituali. Gli insegnamenti o i rituali sacri non sono più confinati a specifiche aree geografiche o a circoli d'élite. Un individuo in un villaggio remoto, con accesso a una piattaforma alimentata dall'intelligenza artificiale, può ora partecipare a congregazioni spirituali globali, rendendo l'appello universale di molte tradizioni di fede una realtà tangibile.
- *Garantire l'autenticità*: Tuttavia, c'è un'avvertenza. Sebbene l'accessibilità sia lodevole, non dovrebbe andare a scapito dell'autenticità. Quando l'IA crea pellegrinaggi virtuali o ritiri digitali, è fondamentale che queste esperienze non siano solo simulacri, ma risuonino con la genuina essenza spirituale delle tradizioni che rappresentano. L'attenzione di Lonergan per il bene comune ci ricorda che l'arricchimento spirituale dovrebbe essere sia ampio che profondo, raggiungendo molti ma senza mai rimanere soltanto in superficie<sup>71</sup>.

Considerazioni etiche sulla spiritualità guidata dall'IA:

- *Pregiudizi*: come per ogni sistema di intelligenza artificiale, c'è il rischio intrinseco che i pregiudizi si insinuino negli algoritmi.

<sup>71</sup> B. J. F. Lonergan, *A Second Collection*, (CWL 13), p. 108.

mi. Quando questi pregiudizi permeano le piattaforme religiose, possono distorcere gli insegnamenti spirituali o favorire certe interpretazioni rispetto ad altre. Sulla base dell'enfasi posta da Lonergan sul bene comune, i sistemi di IA in contesti religiosi devono essere curati meticolosamente per garantire che rappresentino il diverso arazzo di credenze e interpretazioni all'interno delle tradizioni religiose.

- *Privacy e sacralità*: Le esperienze religiose sono profondamente personali e, per molti, sacre. Quando i credenti condividono le loro riflessioni spirituali o chiedono consigli alle piattaforme guidate dall'intelligenza artificiale, c'è un'esigenza fondamentale di salvaguardare questi dati. Questo non può essere soddisfatto semplicemente attraverso le leggi sulla protezione dei dati, ma richiede di sostenere la sacralità del viaggio spirituale di un credente. Qualsiasi uso improprio o mercificazione di questi dati non è solo una violazione tecnologica, ma una violazione della fiducia spirituale.
- *Allineamento etico*: Infine, le piattaforme di IA che operano in ambito religioso devono essere allineate con i principi etici delle tradizioni religiose che servono. Che si tratti di principi di compassione, giustizia o riverenza, queste piattaforme non devono limitarsi a diffondere insegnamenti spirituali, ma incarnarli nelle loro operazioni e interazioni.

In sostanza, poiché l'IA diventa un mediatore del sacro, la sua progettazione, il suo impiego e la sua governance dovrebbero risuonare con il bene comune: un equilibrio armonioso in cui le meraviglie tecnologiche elevano le esperienze spirituali, assicurando che siano accessibili, genuine ed eticamente fondate.

### *Conclusioni*

Nell'intersezione tra le venerabili tradizioni della spiritualità e il regno avanguardistico dell'intelligenza artificiale, assistiamo a un arazzo di sfide e opportunità. Mentre l'intelligenza artificiale si ritaglia il suo spazio nelle esperienze religiose - dai santuari digitali alle interpretazioni scritturali assistite dall'intelligenza artificiale - i confini di ciò che costituisce il regno del sacro vengono ridefiniti in modo dinamico. La filosofia di Bernard Lonergan, con la sua enfasi sull'interazione tra soggettività e oggettività, fornisce un'ancora per il nostro uso degli strumenti tecnologici nella ricerca della spiritualità e del bene comune. Le sue intuizioni ci ricordano che mentre l'intelligenza artificiale può ampliare la portata e la profondità delle esperienze religiose, la santità, l'autenticità e i principi etici delle tradizioni di fede dovrebbero rimanere inviolati.

Il "re-incanto" portato dall'IA presenta un ricco panorama di meraviglie, come il pellegrinaggio virtuale a La Mecca o l'interpretazione sfumata di testi antichi come la Bhagavad Gita. Tuttavia, accanto a questa meraviglia c'è un invito al discernimento, per far sì che in questo nuovo e coraggioso mondo digitale lo spirito umano non venga messo in ombra, ma venga innalzato e arricchito. Il futuro ci invita a una co-creazione armoniosa, in cui l'uomo, le macchine e il divino collaborano alla stesura di un racconto, un racconto in cui le innovazioni tecnologiche non sono solo prodezze di codici e algoritmi, ma risuonano con le aspirazioni più profonde del cuore umano. Dove l'intelligenza artificiale, invece di soppiantare le esperienze spirituali, le integra e le eleva, favorendo un mondo in cui il silicio e lo spirituale si fondono, ampliando i nostri orizzonti di comprensione, riverenza e benessere comune. In questa confluenza di antico e futuristico, ci viene ricordata l'essenza senza tempo della nostra ricerca di significato, connessione

e trascendenza: un viaggio che, con il giusto intento e la giusta guida, può essere meravigliosamente incrementato dagli strumenti della nostra epoca.

## Bibliografia

- J. Anderson - L. Rainie, *As AI Spreads, Experts Predict the Best and Worst Changes in Digital Life by 2035*, Pew Research Center, June 2023.
- J. Anderson - J. R. Laracy - T. Marlowe, *A Multi-Disciplinary Analysis of Catholic Social Teaching with Implications for Engineering and Technology*, in «Systemics, Cybernetics and Informatics», 18, 2020, pp. 41-49.
- VR pilgrimage to holy sites from Mecca to the Vatican, here's how*, in «Gadgets Now», 5 August 2022.
- E. Atwell - N. Habash - B. Louw - B. Abu Shawar - T. McEnery - W. Zaghouani - M. El-Haj, *Understanding the Quran: A new grand challenge for computer science and artificial intelligence*, in «ACM-BCS Visions of Computer Science», 2010, pp. 1-2, <<https://eprints.lancs.ac.uk/id/eprint/71278>>.
- S. Aupers, *"The Force is Great" Enchantment And Magic in Silicon Valley*, in «Marsaryk University journal of law and technology», 3, 2009, pp. 153-173.
- P. Beer, *An Introduction to Bernard Lonergan: Exploring Lonergan's approach to the great philosophical questions*, Glen Waverley, Sid Harta Publishers, 2020.
- R. Beragnoli, *AI Religious Art & Modern Saints, Tecnocopia, Complesso di Santa Cristina / Metaverso*, Bologna, Adiacenze, 2022.
- D. Bhatti, *AI becomes astrologer, here is how Kundali GPT chatbot uses artificial intelligence to predict horoscope*, in «India Today», 4 August 2023.
- P. Bory, *Deep new: The shifting narratives of artificial intelligence from Deep Blue to AlphaGo*, in «Convergence: The International Journal of Research into New Media Technologies», 25, 2019, pp. 627-642, <<https://doi.org/10.1177/1354856519829679>>.
- P. H. Byrne, *The Ethics of Discernment: Lonergan's Foundations for Ethics*, Toronto, University of Toronto Press, 2016.
- H. Campbell, *When Religion Meets New Media*, London, Routledge, 2010.
- Z. Cattaneo - C. Lega - A. Flexas - M. Nadal - E. Munar - C. J. Cela-Conde, *The world can look better: enhancing beauty experience with brain stimulation*, in «Social Cognitive and Affective Neuroscience», 9, 2014, pp. 1713-1721, <<https://doi.org/10.1093/scan/nst165>>.

- R. Chandra, *AI, philosophy and religion: what machine learning can tell us about the Bhagavad Gita*, in «The Conversation», 12 May 2022.
- D.-Y. Chung, *Human Religious Evolution*, in «Open Journal of Social Sciences», 4, 2016, pp. 75-90, <<https://doi.org/10.4236/JSS.2016.43013>>.
- M. Coeckelbergh, *New romantic cyborgs: Romanticism, information technology, and the end of the machine*, Cambridge, MIT Press, 2017.
- E. Condon, *Confession time, synodal AI, and calling a shot*, in «The Pillar», 3 March 2023.
- B. J. F. Lonergan, *Insight: A Study of Human Understanding*, (CWL 3), a cura di F. E. Crowe and R. M. Doran, Toronto, University of Toronto Press, 2013.
- B. J. F. Lonergan, *A Second Collection*, (CWL 13), a cura di F. E. Crowe and R. M. Doran, Toronto, University of Toronto Press, 2016.
- B. J. F. Lonergan, *Method in Theology*, (CWL 14), a cura di R. M. Doran and J. D. Dadosky, Toronto, University of Toronto Press, 2017.
- C. Disbrey, *George Fox and Some Theories of Innovation in Religion*, in «Religious Studies», 25, 1989, pp. 61-74, <<https://doi.org/10.1017/S0034412500019727>>
- P. A. Egan, *Philosophy and Catholic Theology: A Primer*, Liturgical Press, Collegeville, MN, 2009.
- Z. Epstein - S. Levine - D. G. Rand - I. Rahwan, *Who gets credit for ai-generated art?* in «iScience», 9, 2020, 101515, <<https://doi.org/10.1016/j.isci.2020.101515>>.
- Watch: How has the Hajj pilgrimage changed amid the coronavirus pandemic?* in «Euronews», 31 July 2020.
- European Commission, *White Paper On Artificial Intelligence - A European approach to excellence and trust*, Brussels, 19.2.2020 COM (2020) 65 final, pp. 1-27.
- J. T. Feezell - J. K. Wagner - M. Conroy, *Exploring the effects of algorithm-driven news sources on political behavior and polarization*, in «Computers in Human Behavior», 116, 2021, 106626, <<https://doi.org/10.1016/j.chb.2020.106626>>
- M. Frackiewicz, *A New Era of Global Collaboration: AI's Role in Cross-Cultural Exchange*, in «TS2», 7 September 2023.
- B. Friedman - D. G. Hendry, *Value Sensitive Design: Shaping technology with moral imagination*, Cambridge, MIT Press, 2019.
- T. Gillespie, *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*, Cambridge, Yale University Press, 2018.
- A. Glenberg - C. R. Jones, *Why ChatGPT doesn't understand what it's saying*, in «Fast Company», 04 September 2023.
- M. Graves, *ChatGPT's Significance for Theology*, in «Theology and Science», 21, 2023, pp. 201-204, <<https://doi.org/10.1080/14746700.2023.2188366>>.
- J. E. Green, *Two Meanings of Disenchantment: Sociological Condition vs. Philosophical Act – Reassessing Max Weber's Thesis of the Disenchantment of the*

- World, in «Philosophy and Theology», 17, 2005, pp. 51-84, <<https://doi.org/10.5840/PHILTHEOL2005171/24>>.
- S. Han, *Disenchantment revisited: Formations of the 'secular' and 'religious' in the technological discourse of modernity*, in «Social Compass», 62, 2015, pp. 76-88, <<https://doi.org/10.1177/0037768614560960>>.
- B. Hill, *The World's First AI Service to Create Personalized Mindfulness Meditations On Request*, in «LinkedIn», 30 June 2023.
- O. Kudina, «*Alexa, who am I?*»: Voice Assistants and Hermeneutic Lemniscate as the Technologically Mediated Sense-Making, in «Human Studies», 2021, pp. 1-21, <<https://doi.org/10.1007/s10746-021-09572-9>>.
- J. R. Laracy, *Epistemology and Metaphysics in Interdisciplinary Communication: Insights from Ian Barbour and Bernard Lonergan, SJ*, in «Systemics, Cybernetics and Informatics», 17, 2019, pp. 49-54.
- J. R. Laracy, *Theology and Science in the Thought of Ian Barbour*, Berlin, Peter Lang, 2021.
- R. M. Liddy, *Startling Strangeness: Reading Lonergan's Insight*, University Press of America, Lanham, MD, 2006.
- B. C. D. S. Lima - N. Omar - I. Avansi - L. N. de Castro - I. F. Silveira, *Use of artificial intelligence in biblical citation recommendations in the new testament*, in «Revista Científica Multidisciplinar Núcleo Do Conhecimento», 2023, pp. 123-143, <<https://doi.org/10.32749/nucleodoconhecimento.com.br/computer-engineering/biblical-citation>>.
- D. Löffler - J. Hurtienne - I. Nord, *Blessing robot Bless U2: a discursive design study to understand the implications of social robots in religious contexts*, in «International Journal of Social Robotics», 13, 2021, pp. 569-586, <<https://doi.org/10.1007/s12369-019-00558-3>>.
- T. Marlowe - J. R. Laracy, *Philosophy and Cybernetics: Questions and Issues*, in «Systemics, Cybernetics and Informatics», 19, 2021, pp. 1-23.
- J. F. McGrath, *Jesus according to AI*, in «Patheos», 27 July 2020.
- J. McLellan, *AI and the meaning of life: Tech industry turns to religious leaders*, United States Conference of Catholic Bishops, 23 March 2023.
- S. McNulty, *I asked an AI art generator to draw Catholicism in 20 different ways. Here's what I learned*, in «America Magazine», 13 July 2022.
- R. Min, *The metaverse is opening up the world's holiest sites to virtual pilgrims*, in «Euronews», 11 August 2022.
- R. Moloney, *Conversion and Spirituality: Bernard Lonergan (1904-1984)*, in «The Way», 43, 2004, pp. 123-134.
- G. Nachshon - J. Zeleznikow - T. Dowdeswell, *From the Tree of Knowledge and the Golem of Prague to Kosher Autonomous Cars: The Ethics of Artificial Intelli-*

- gence Through Jewish Eyes*, in «Oxford Journal of Law and Religion», 9, 2020, pp. 132-156, <<https://doi.org/10.1093/ojlr/rwaa015>>.
- S. Niu, *Virtual hajj as a response to demographic and geopolitical pressures*, in «Contemporary Islam», 17, 2023, pp. 95-108, <<https://doi.org/10.1007/s11562-023-00512-1>>.
- K. Porayska-Pomsta - G. Rajendran, *Accountability in Human and Artificial Intelligence Decision-Making as the Basis for Diversity and Educational Inclusion*, in «Artificial Intelligence and Inclusive Education. Perspectives on Rethinking and Reforming Education», a cura di J. Knox - Y. Wang - M. Gallagher, Singapore, Springer, 2019, <[https://doi.org/10.1007/978-981-13-8161-4\\_3](https://doi.org/10.1007/978-981-13-8161-4_3)>.
- M. Renz - M. S. Mao - A. Omlin - D. Bueche - T. Cerny - F. Strasser, *Spiritual Experiences of Transcendence in Patients With Advanced Cancer*, in «American Journal of Hospice and Palliative Medicine», 32, 2017, pp. 178-188, <<https://doi.org/10.1177/1049909113512201>>.
- K. Roscoe, *Lonergan's Theory of Cognition, Constructivism and Science Education*, in «Science & Education», 13, 2004, pp. 541-551, <<https://doi.org/10.1023/B:SCED.0000042853.60774.B2>>.
- M. Ryan, *In AI We Trust: Ethics, Artificial Intelligence, and Reliability*, in «Science and Engineering Ethics», 26, 2020, pp. 2749-2767, <<https://doi.org/10.1007/s11948-020-00228-y>>.
- E. Schwitzgebel, *AI systems must not confuse users about their sentience or moral status*, in «Patterns», 4, 2023, 100818, <<https://doi.org/10.1016/j.patter.2023.100818>>.
- S. Shen - J. Fan, *Emotion analysis of ideological and political education using a GRU deep neural network*, in «Frontiers in Psychology», 13, 2022, 908154, <<https://doi.org/10.3389/fpsyg.2022.908154>>.
- B. Shneiderman, *Human-centered AI*, Oxford, Oxford University Press, 2022.
- B. Singler, *The AI Creation Meme: A Case Study of the New Visibility of Religion in Artificial Intelligence Discourse*, in «Religions», 11, 2020, 253, <<https://doi.org/10.3390/rel11050253>>.
- J. K. Smith, *Robot Theology: Old Questions Through New Media*, Eugene, Wipf and Stock Publishers, 2022.
- Y. Song, *Religious AI as an Option to the Risks of Superintelligence: A Protestant Theological Perspective*, in «Theology and Science», 19, 2020, pp. 65-78, <<https://doi.org/10.1080/14746700.2020.1825196>>.
- L. Tamatea, *Online Buddhist and Christian Responses to Artificial Intelligence*, in «Zygon», 45, 2010, pp. 979-1002, <<https://doi.org/10.1111/J.1467-9744.2010.01145.X>>.

- C. Tan, *Digital Confucius? Exploring the implications of artificial intelligence in spiritual education*, in «Connection Science», 32, 2020, pp. 280-291, <<https://doi.org/10.1080/09540091.2019.1709045>>.
- Confession: The Roman Catholic App Doesn't Support Real iConfessions*, in «Tech Crunch», 9 February 2011.
- T. J. Tekippe, *What is Lonergan Up to in Insight?: A Primer*, Collegeville, The Liturgical Press, 1996.
- J. A. Teske, *Cyberpsychology, Human Relationships, and Our Virtual Interiors*, in «Zygon», 37, 2002, pp. 677-700, <<https://doi.org/10.1111/1467-9744.00445>>.
- H. Thinyane - F. Sasseti, *Towards a Human Rights-Based Approach to AI: Case Study of Apprise*, in «Global Souths. IDIA 2020. Communications in Computer and Information Science», vol 1236, a cura di D. R. Junio - C. Kooopman, Cham, Springer, 2020, <[https://doi.org/10.1007/978-3-030-52014-4\\_3](https://doi.org/10.1007/978-3-030-52014-4_3)>.
- K. Tiffany, *Confess all your sins to this robot, if you want*, in «The Verge», 28 April 2016.
- S. Umbrello, *From Subjectivity to Objectivity: Bernard Lonergan's philosophy as a grounding for value sensitive design*, in «S&F\_scienzaefilosofia.it», 29, 2023, pp. 36-46.
- C. H. J. Van der Merwe, *An overview of recent developments in the description of Biblical Hebrew relevant to Bible translation*, in «Acta Theologica», 22, 2004, pp. 228-245, <<https://doi.org/10.4314/ACTAT.V22I1.5462>>.
- J. Wakefield, *AI emotion-detection software tested on Uyghurs*, in «BBC News», 26 May 2021.
- M. Weber, *The Sociology of Religion*, Boston, Beacon Press, 1993 (First published 1920).
- S. Wells, *A.I. Reveals the Hidden Author of a Crucial Bible Text*, in «Inverse», 21 April 2021.
- D. Whitney, *Appreciative inquiry: creating spiritual resonance in the workplace*, in «Journal of Management, Spirituality & Religion», 7, 2010, pp. 73-88, <<https://doi.org/10.1080/14766080903497656>>.

Steven Umbrello  
Università degli Studi di Torino  
steven.umbrello@unito.it  
<https://orcid.org/0000-0003-2594-6313>



*Life-on-life* inaugura un percorso editoriale concepito per accompagnare, potenziare e integrare le attività della Task Force *Human&Future* dell'Università degli Studi di Napoli Federico II. Questo primo volume si configura come un'opera transdisciplinare dedicata a uno dei temi più attuali del dibattito filosofico e scientifico: l'intelligenza artificiale. Il tema viene esplorato da molteplici prospettive – dalla filosofia al diritto, dalla dimensione storica alla pedagogia, dalla politica all'ingegneria informatica – con l'intento di cogliere la complessità e l'impatto profondo che l'IA esercita sulla nostra società. Per la prima volta nella sua storia, l'essere umano è in grado di progettare macchine dotate di ampia autonomia funzionale, capaci di svolgere compiti che un tempo erano prerogativa esclusiva dell'intelligenza umana. Non siamo di fronte a un'innovazione tecnologica tra le tante: si tratta di una vera e propria svolta antropologica, destinata a trasformare radicalmente l'ecosistema sociale di *Sapiens*, intrecciandosi con le tecnologie digitali che ne costituiscono ormai lo sfondo permanente. La lettura del volume offre dunque uno sguardo poliedrico, ma anche profondamente unitario, sull'IA: una tecnologia che rappresenta la prima concreta operazione di *Sapiens* volta alla creazione di enti capaci di agire, competere e generare significato all'interno degli ecosistemi digitali. Nessun ambito dell'esperienza umana resta immune da questa trasformazione, che porta con sé tanto rischi quanto opportunità – entrambi oggetto di attenta analisi nelle pagine del volume.

Mario Cosenza insegna Filosofia Morale presso l'Università degli Studi di Napoli Federico II, ateneo presso il quale è coordinatore della segreteria scientifico-organizzativa della Task Force *Human&Future*; è autore di *All'ombra dei Lumi. Jacques-André Naigeon philosophe* (2020, a breve tradotto in francese), della traduzione e della curatela de *Il Militare filosofo* (2023), nonché di numerose pubblicazioni su questioni etico-politiche della filosofia moderna e contemporanea.

Luca Lo Sapio insegna Filosofia Morale e Bioetica presso l'Università degli Studi di Torino. È coordinatore scientifico dei Gruppi di ricerca Future Eating e BIOETHIKA-TO. Si occupa di Etica applicata, Bioetica e dell'impatto delle nuove tecnologie sul futuro di *Sapiens*. È autore di diverse monografie tra cui *Carne Coltivata. Etica dell'agricoltura cellulare* (2024) e *Cambia la tua vita o affronta l'estinzione. Introduzione a un'etica per la fine del mondo* (2022) e numerosi articoli su temi di etica delle nuove tecnologie, etica dell'astrobiologia ed etica dell'estinzione.

ISBN 978-88-6887-355-4  
DOI 10.6093/978-88-6887-355-4

ISBN 978-88-6887-355-4



9 788868 873554