Annali di studi religiosi
ISBN 979-12-218-2156-7
DOI 10.53136/979122182156729
pp. 521-538 (ottobre 2025)

DOBBIAMO *CREDERE* NELL'INTELLIGENZA ARTIFICIALE? UN DIALOGO PER COMPRENDERE IL PRESENTE CON MICHELA MILANO E PAOLO TRAVERSO

Paolo Costa, Eugenia Lancellotta e Boris Rähme

Abstract: This is the transcript of a round table discussion on the topic of trustworthy Artificial Intelligence (AI) organized and conducted by three FBK–ISR researchers involved in the project "Resilient Beliefs: Religion and Beyond." The experts consulted are Michela Milano and Paolo Traverso, two leading scholars in the field of AI. Starting from the notion of resilient belief, topics such as the difference between human and artificial intelligence, the danger of replacing people with intelligent machines even in jobs traditionally considered creative, the geopolitical dimension of the AI Revolution, the legal regulation of AI hazards and the goal of consistent human oversight, how to create an epistemic and social environment conducive to the human–centric development of new digital technologies, are discussed.

Questa è la trascrizione di una tavola rotonda sul tema della *trustworthy AI* organizzata e condotta da tre ricercatori di FBK–ISR impegnati nel progetto "Resilient Beliefs: Religion and Beyond". Gli esperti consultati sono Michela Milano e Paolo Traverso, due studiosi di frontiera nel campo dell'Intelligenza Artificiale. Proprio a partire dalla nozione di convinzione resiliente vengono discussi temi quali: la differenza tra intelligenza umana e artificiale; il pericolo di sostituzione delle persone con macchine intelligenti anche in lavori tradizionalmente considerati creativi; la dimensione geopolitica della Rivoluzione dell'Intelligenza Artificiale; la regolamentazione giuridica dei rischi dell'IA e l'obiettivo di un sistematico *human oversight*; la creazione di un ambiente epistemico e sociale favorevole a uno sviluppo umanocentrico delle nuove tecnologie digitali.

Keywords: Artificial Intelligence, Resilient Beliefs, Geopolitical challenges, Human-centeredness, EU AI Act

Parole chiave: Intelligenza Artificiale, Convinzioni resilienti, Sfide geopolitiche, Centralità dell'umano, Regolamento UE sull'IA

> Che cosa ci preoccupa davvero nell'Intelligenza Artificiale? Qual è esattamente il problema?
> Nello Cristianini

1. La cornice teorica

Prima di dare il via al dibattito con Michela Milano e Paolo Traverso, vale la pena dire qualcosa sulla cornice dell'evento[1]. Il dialogo con due esperti di una delle trasformazioni tecnologiche che più scuotono l'opinione pubblica oggi è anche l'occasione per onorare la conclusione di un progetto di ricerca biennale intitolato "Resilient Beliefs: Religion and Beyond", finanziato dall'Euregio Science Fund e condotto in collaborazione con l'Università di Innsbruck e lo Studio Teologico Accademico di Bressanone.

Che cosa sono i "resilient beliefs"? I *resilient beliefs* sono le convinzioni tenaci che ciascuno di noi ha e a sostegno delle quali non servono particolari evidenze (pensiamo, per esempio, alla credenza che non sia possibile respirare sott'acqua) o che resistono alle evidenze contrarie (pensiamo, ad esempio, alla fiducia nell'esistenza del libero arbitrio). Siccome le credenze religiose sono un caso paradigmatico di convinzioni resilienti, nel disegnare il progetto abbiamo pensato che potesse essere interessante partire dal modo in cui funzionano queste credenze per investigare altri tipi di convinzioni resilienti, dalla fede nella democrazia alle teorie del complotto, dalla fiducia cieca nella crescita economica alle manie di persecuzione. Uno degli obiettivi della ricerca era proprio l'identificazione di criteri per distinguere tipologie buone e cattive di resilienza delle opinioni.

Più lo utilizzavamo e più ci è parso che il concetto di *resilient belief* avesse la capacità di illuminare una serie di fenomeni interessanti anche dal punto di vista, diciamo così, "pedagogico". Per questo nella primavera del 2024 abbiamo organizzato un breve ciclo di incontri con due classi IV dell'Istituto Artigianelli di Trento. Lavorando fianco a fianco

(1) Pubblichiamo qui la trascrizione, rivista dai partecipanti, della tavola rotonda tenutasi a Trento, nell'Aula grande di FBK, il 21 gennaio 2025 e intitolata "Dobbiamo *credere* nell'Intelligenza Artificiale". La ricerca che sta alla base di questo lavoro è stata condotta all'interno dell'Interregional Project Network IPN 175 "Resilient Beliefs: Religion and Beyond" (2022-2024), finanziato dall'Euregio Science Fund.

con gli studenti abbiamo provato a far emergere le loro convinzioni resilienti riguardo alla trasformazione tecnologica con cui tutti dobbiamo fare i conti oggi: la rivoluzione digitale e, soprattutto, i progressi impetuosi nel campo dell'Intelligenza Artificiale. L'obiettivo, in breve, era aiutare gli studenti a mettere a fuoco la nozione di credenza resiliente; sforzarsi di portare a galla casi esemplari di convinzioni forti; spingerli a confrontarle con quelle dei propri compagni; provare a capire su quali tipo di evidenze poggiassero; ecc.

Dal lavoro svolto in classe sono emersi numerosi spunti che abbiamo deciso di usare come filo rosso per conversare con due esperti del campo, le cui convinzioni resilienti circa i recenti sviluppi della *Computer science* sono della tipologia migliore, quella, cioè, basata su informazioni affidabili e buone ragioni. *Michela Milano* è professoressa ordinaria nel Dipartimento di Informatica, Scienza e Ingegneria dell'Università di Bologna e Direttrice del Centre for Digital Society di FBK. La sua attività di ricerca riguarda l'Intelligenza Artificiale con un focus particolare sui sistemi di supporto alle decisioni. In questo settore ha raggiunto visibilità internazionale collaborando con diversi gruppi di ricerca, dentro e fuori l'Università. È stata vicepresidente dell'European Association of Artificial Intelligence. *Paolo Traverso* è stato Direttore del Centro di Ricerca Tecnologie dell'Informazione e Comunicazione di FBK dal 2007 al 2020. Ha diretto diversi progetti di Intelligenza Artificiale per l'Industria Digitale, per la Salute e il Benessere e per la Pubblica Amministrazione. Ha pubblicato oltre cento articoli scientifici — su riviste internazionali e atti di convegni — ed è anche co–autore di tre libri di testo sulla Pianificazione Automatica e l'Intelligenza Artificiale.

2. Convinzioni resilienti I: la questione dell'intelligenza

Partiamo dal tema dei temi quando si discute di macchine intelligenti. Per cominciare, con gli studenti ci siamo soffermati sull'aura che circonda il concetto di "intelligenza". Molti di loro erano effettivamente inclini a collegare l'intelligenza a qualche forma di creatività. Da questo punto di vista, essere intelligenti significa essere capaci di rispondere in

maniera non banale a una situazione problematica, escogitare soluzioni, per l'appunto, ingegnose a problemi complessi. Una macchina che sa giocare bene a scacchi, sa tradurre velocemente o addirittura scrivere un testo, comporre una canzone o dipingere, suscita meraviglia perché riesce a fare cose che ci aspetteremmo soltanto da una persona.

Molte applicazioni dell'AI sono lì a dimostrarci che l'intelligenza umana può essere simulata da macchine in grado di supervisionare e combinare autonomamente moltissime informazioni in pochissimo tempo per risolvere in maniera efficiente problemi specifici. Ma, hanno osservato molti studenti, si tratta pur sempre di un'intelligenza senza cognizione, cioè senza discernimento, e senza sentimenti, in cui, per esempio, non sembra avere rilevanza la distinzione tra bene e male o quella tra reale e non reale – distinzioni che sembrano invece fondamentali per noi umani. La dimensione più tipica dell'intelligenza umana si direbbe essere, dunque, fuori dalla portata delle macchine. Che tipo di intelligenza è un'intelligenza senza interiorità? È una nuova forma d'intelligenza? Un'intelligenza aliena?

Paolo Traverso: Comincio col dire che sono d'accordo con gli studenti. L'intelligenza artificiale è un'intelligenza senza discernimento e senza emozioni. Certo, abbiamo macchine capaci di fare "sentiment analysis" e stabilire se un testo esprime, per dire, paura o rabbia. Non possiamo però sostenere che le macchine, per quanto intelligenti siano, capiscano che cosa si provi ad arrabbiarsi perché qualcosa che credevamo reale non esiste o indignarsi perché qualcosa a cui attribuiamo un valore speciale viene maltrattato. D'altronde, si tratta di una forma d'intelligenza i cui risultati strabilianti sono stati ottenuti in maniera molto diversa da come avviene nel caso dell'intelligenza umana. Al famoso ChatGPT sono stati dati in pasto miliardi di testi "intelligenti" dai quali, poi, grazie alla teoria della probabilità, è possibile ricavare risposte intelligenti alle nostre domande e curiosità. Il meccanismo, in fondo, è quello di un "riempibuco" statistico.

Possiamo definire tutto ciò "creatività"? Sì, certo, queste macchine possono produrre testi, poesie, canzoni, immagini che prima non esistevano. Un giorno faranno anche film. Ci siamo già molto vicini con Sora,

l'applicazione di OpenAI. Lo stupore è quindi giustificato. È un'intelligenza che emerge ed emerge in maniere diverse, anche perché i metodi di addestramento sono diversi. Ma, come non si stanca mai di ricordarci Nello Cristianini (2025), l'intelligenza prende sempre forme diverse e ha persino poco senso provare a immaginare la natura di un'intelligenza aliena. Sarà sicuramente un'intelligenza totalmente diversa dalla nostra, anche se noi fatichiamo a immaginarcela così *diversa* dalla nostra. È chiaramente un difetto strutturale della nostra immaginazione.

Anche se sono in molti a farlo, non ha perciò senso chiedere scusa a ChatGPT o insultarla. Alle nostre scuse o insulti può infatti reagire, ma non può certo dargli l'importanza che gli diamo noi. E non mi sembra che abbia nemmeno molto senso chiedersi se dovremo prima o poi dare la cittadinanza a un robot. Alla fine, dietro ChatGPT ci sono solo bit, anche se la combinazione di questi bit, grazie all'enorme potere computazionale delle nuove macchine, può produrre effetti stupefacenti. Ma "stupefacente" è poi davvero sinonimo di "creativo"? Sappiamo tutti come scrive le poesie ChatGPT: i programmatori hanno inserito un parametro per cui, anziché cercare sempre la parola più probabile, periodicamente ne viene scelta una a caso. È creatività poetica questa? Non lo so. So però che ci sono persone che usano le poesie scritte con l'IA per dare forma a emozioni ed esprimere sentimenti. C'è un difetto di autenticità in questi casi? A me il sospetto rimane, ma vale comunque la pena di continuare a interrogarci.

Personalmente, comunque, a me preme soprattutto che continuiamo a riflettere sulla possibilità di fare cose buone e utili grazie a queste nuove tecnologie. Restiamo nell'ambito della scuola. Ovvio, nel caso di uno studente che si serve di ChatGPT solo per fare i compiti senza imparare nulla da quello che fa ha poco senso parlare di intelligenza aumentata. Il discorso è diverso per un ragazzo o una ragazza che scelgano di usare questo nuovo strumento allo scopo di accrescere le proprie possibilità di apprendimento. Se ci pensate, quanto sto dicendo riguardo a ChatGPT vale anche per la calcolatrice: se mi sostituisce e basta, perdo competenze, se mi affianca, me le allarga. E quando dico "affiancare" mi riferisco al fatto che chi la usa non deve mai rinunciare alla propria capacità critica e quindi, in caso di bisogno, alla possibilità di discernere un mal funzionamento dello strumento.

Discernimento, appunto. Per quanto riguarda, poi, la distinzione tra il bene e il male, già adesso si può fare in modo che la macchina la incorpori nel proprio funzionamento, ad esempio inserendo dei paletti, i cosiddetti *guardrails*, che escludono dalle risposte corrette quelle che per noi superano i confini della decenza o del decoro morale perché razziste, offensive, violente. Ma sono curioso di sentire che cos'ha da dire Michela in proposito.

Michela Milano: Penso anch'io che le due forme d'intelligenza, naturale e artificiale, siano molto diverse. A dire il vero, fin dall'inizio, e per inizio intendo il 1956, cioè l'anno in cui è stato effettivamente lanciato in una conferenza a Dartmouth il progetto di creare macchine intelligenti, l'obiettivo non era quello di *sostituire*, ma di *simulare* l'intelligenza umana. Poco importa, insomma, cosa ci sta dietro. L'importante è che all'esterno — ai "morsetti" come ci piace dire — il risultato sia simile.

Pensiamo ai sistemi di pianificazione. Pianificare, cioè disporre secondo una sequenza logica un certo numero di azioni, è una cosa che anche i bambini molto piccoli sanno fare benissimo. Per una macchina, tuttavia, pianificare è un compito molto difficile. Per raggiungere lo scopo deve operare in modo sintattico, cioè lineare, mettendo insieme dei blocchetti. Il risultato, alla fine, è lo stesso, ma i modi di arrivarci sono totalmente diversi. Ragionando pragmaticamente, non c'è motivo di scandalizzarsi. Ci limitiamo a prenderne atto.

Lo stesso atteggiamento laico lo adotterei anche rispetto alla questione dei sentimenti. Do per scontato che le macchine non provino sentimenti. Ma questo è necessariamente un difetto? Un'intelligenza senza sentimento potrebbe esserci utile in qualche caso. Pensiamo ai sistemi di IA che stiamo progettando per supportare un processo decisionale, ad esempio per stabilire un ordine di precedenza nelle liste d'attesa per i trapianti d'organo. Non c'è dubbio che il processo decisionale umano è molto più complicato dei processi automatizzati. Ma "complicato" non vuol necessariamente dire più efficiente o giusto. In qualche caso un processo decisionale più asettico, più oggettivo può essere esattamente ciò di cui abbiamo bisogno o ciò a cui aspiriamo. Il che non esclude che, quando serve, non si possano poi inserire nei modelli dei

vincoli etici per evitare che l'asetticità delle decisioni produca delle mostruosità morali (che, per altro, non è che manchino nelle normali decisioni degli esseri umani in carne e ossa…).

Vi racconto un aneddoto per darvi un'idea di quanto le emozioni umane interferiscano con il nostro rapporto con le macchine. A Bologna abbiamo un piccolo robot, il NAO. È un cosino di dimensioni ridotte, poco più di cinquanta centimetri. Lo portiamo in giro per aiutarci nell'opera di divulgazione. Le reazioni che suscita NAO nelle persone sono rivelatrici. Per i bambini è ovviamente un giocattolo. Per i maschi, in genere, è una diavoleria che li appassiona, come fosse una moto di grande cilindrata. Nelle donne spesso suscita l'istinto materno, perché ha due occhioni che ricordano quelli di un bambino. Se tutto ciò succede a questo livello basilare, figuriamoci se possiamo sperare di separare l'intelligenza e i sentimenti nel nostro rapporto con le macchine…

Paolo Traverso: D'altra parte, se posso aggiungere una postilla a quanto ha appena detto Michela, la relazione tra le persone e le cose è sempre carica di sentimenti, proiezioni, strani giochi della mente. Ne abbiamo ragionato proprio in questa aula qualche settimana fa. Le cose non sono mai solo "cose" per noi umani. L'importante è non farsi sopraffare da questo gioco di specchi e non perdere di vista la distinzione tra ciò che è tipicamente nostro e quello che stiamo progettando e costruendo proprio per affiancare, supportare e, se possibile, migliorare le nostre doti e qualità, come pure per compensare i nostri difetti.

3. Convinzioni resilienti II: la competizione con l'intelligenza umana

Discutendo con gli studenti sono emerse spesso — forse addirittura più spesso di quanto ci aspettassimo — inquietudini legate alla diffusione delle nuove tecnologie. Sono convinzioni robuste che riguardano, anzitutto, il loro futuro lavorativo. È viva nei giovani con cui abbiamo dialogato la percezione del rischio di una progressiva sostituzione del lavoro umano. Anche del lavoro creativo. Un esempio ricorrente che ci veniva fatto concerne le mansioni classiche del grafico pubblicitario.

Non di rado, comunque, l'ansia era bilanciata dalla speranza che le nuove tecnologie possano semplificare e quindi migliorare il lavoro, delegando alla macchina le parti più ripetitive e meccaniche o favorendo la gestione simultanea di più processi. Nello scenario più ottimistico, le nuove macchine intelligenti potrebbero addirittura servire da stimolo e spingere le persone a sviluppare nuove competenze, nuovi *skills*.

Un'altra preoccupazione che abbiamo riscontrato ruota attorno al rischio di una crescente dipendenza degli individui dalle macchine, della cessione, cioè, alle macchine di una parte significativa delle capacità umane più preziose (attenzione, pazienza, iniziativa, senso della realtà). Il timore, insomma, è che le macchine finiscano per dettare alle persone i tempi di vita e influire sulla qualità delle loro relazioni. In un gruppo di discussione ha fatto capolino la parola "brainrot", come a dire che i computer hanno in sé qualcosa di vampiresco. In qualche caso, la preoccupazione scaturiva da un'esperienza diretta. Già oggi l'IA è presente nel lavoro scolastico, ma non è sempre vissuta come un sostegno affidabile. Da qui nasce il timore che possa rendere le persone meno intelligenti, impigrendole, rendendole passive, disincentivandole a studiare – timore che si accompagna alla paura per gli effetti che la diffusione dell'IA potrebbe avere sui bambini, sulla loro crescita e capacità di relazionarsi con gli altri.

Michela Milano: Il rischio della sostituzione del lavoro umano è un rischio concreto, non dobbiamo nascondercelo. Se facciamo il confronto con le precedenti rivoluzioni tecnologiche (l'invenzione della macchina a vapore, l'automazione, ecc.), la vera novità è che con l'*AI Revolution* vengono messi a rischio non solo i lavori di fatica, ma anche quelli che una volta chiamavamo i lavori di concetto. Questo è un fatto assodato.

Verso questa rivoluzione si possono poi assumere due atteggiamenti opposti, ambedue legittimi. Il primo è una sacrosanta diffidenza. Uno vuole vederci chiaro e sceglie perciò di muoversi con la massima cautela. Il timore è che possa avvenire una delega completa alla macchina, anche in materia di decisioni che hanno un impatto rilevante sulla vita delle persone. I diffidenti sono spaventati, in particolare, dal rischio di una totale deresponsabilizzazione degli esperti, di coloro che sono

chiamati a dare una direzione alla società. Per fare fronte a questo pericolo, l'UE ha insistito molto sul principio della supervisione umana, dello "human oversight".

Il secondo atteggiamento è più ottimistico. In questo caso, anziché come una rivale, la macchina appare piuttosto come un'alleata. Recentemente sono stata coinvolta in un progetto che mi ha consentito di entrare in contatto con aziende che producono ceramiche. I designer che ho conosciuto erano in genere felici di potersi concentrare solo sulle fasi iniziali del processo creativo, quello più appassionante, allorché si apre effettivamente un nuovo orizzonte. Lo sviluppo dell'intuizione originale, alla fine, è per loro meno interessante, più meccanico, e le macchine possono svolgerlo più velocemente e con precisione assoluta.

La mia impressione è che l'avvento dei *Large Language Models* abbia dischiuso effettivamente nuove possibilità al genere umano. Una di quelle che viene sistematicamente sottovalutata è, secondo me, la possibilità di lavorare un po' meno, tenendoci la parte più creativa del lavoro.

Paolo Traverso: Anch'io vorrei concentrarmi sul lato positivo della faccenda. Ho in mente un esempio che, più che alla sostituzione, ci fa pensare a un allargamento degli orizzonti che non ha precedenti nella storia. Sto pensando al futuro della Sanità pubblica e, vista la crescita esponenziale della popolazione anziana, alla necessità di puntare massicciamente sulla prevenzione delle malattie. Pensate solo a quello che potrebbe fare un sistema d'intelligenza artificiale in grado di analizzare l'immagine della retina di una persona che soffre di diabete e dirci se si tratta di una retina sana o malata. Solo in Italia ci sono più di due milioni e mezzo di persone che soffrono di diabete e corrono il rischio di sviluppare una retinopatia, che è una complicazione classica di questa patologia. Queste persone non sono visitate come si dovrebbe perché non ci sono abbastanza specialisti per effettuare le necessarie visite con regolarità. È un fatto che, senza questo tipo d'innovazione tecnologica, non avremo mai le risorse umane e finanziarie per portare avanti una campagna di prevenzione a tappeto. Solo le nuove macchine intelligenti possono consentirci di tradurre in pratica la massima secondo cui prevenire è meglio che curare, realizzando così il sogno di una sanità veramente pubblica.

Personalmente interpreto in questo modo il principio della supervisione umana sposato dall'*AI Act* europeo. Esseri umani e macchine devono essere alleati, e a questo scopo bisogna fare in modo che i primi affianchino le seconde e le aiutino ad aggiustare sistematicamente i loro risultati. È quello che noi informatici chiamiamo "Human in the loop": un circolo d'intelligenza in cui umano e artificiale sono entrambi indispensabili.

Offrire occhi instancabili ai medici non significa rimpiazzare il loro lavoro, che resta preziosissimo. Ma non fraintendetemi. Con ciò non voglio dire che il pericolo della sostituzione non esista. Molte persone, in fondo, svolgono mansioni "intelligenti", ma non così intelligenti da non poter essere sostituite da macchine intelligenti. Gestire la transizione non sarà facile. Come diceva Michela, bisognerà ragionare su una riorganizzazione complessiva del lavoro, della formazione, e del suo peso nella vita delle persone.

Ritornando ai *resilient beliefs* degli studenti, qualcuno di loro si è aggrappato all'idea che l'IA non possa essere davvero innovativa, che solo l'arte sia veramente creativa. La genialità, però, è patrimonio di pochi. Viene perciò da chiedersi che cosa ne sarà delle forme di creatività più "artigianali", ad esempio il fumetto o la musica pop. Perderanno la loro "unicità", la loro aura, e quindi anche il prestigio e i relativi riconoscimenti economici? Non è che i programmatori, i *computer scientists* finiranno per monopolizzare gran parte della qualità "attiva", non meccanica, del lavoro?

Michela Milano: Personalmente non vedo questo pericolo. Chi escogita gli algoritmi che consentono alle macchine di imparare ha pur sempre bisogno di dati con cui addestrare i computer e questi dati sono il frutto della creatività umana. La fonte primaria dell'intelligenza restiamo noi umani. L'orizzonte non è quindi quello della sostituzione, ma della co-creazione: non *substitution*, ma *co-creation*. Quando parliamo d'intelligenza aumentata pensiamo proprio a questo scenario: mettere a disposizione delle persone più creative strumenti nuovi che potenzino la loro creatività liberando tempo ed energia.

Paolo Traverso: Sì, anch'io concentrerei l'attenzione soprattutto sull'estensione del campo di possibilità. Pensate alla musica. Un nostro

ricercatore, Michele Baldo, ha dato vita a una start-up rivolta proprio ai giovani musicisti. In questo caso la macchina viene in soccorso dei musicisti in erba perché procura loro, a costi infinitamente più bassi, tutti quei supporti tecnici e musicali che una volta erano accessibili solo ai musicisti già affermati. Insomma, dà loro una base su cui esercitare al massimo la propria creatività. La macchina in questo caso funge da supporto e da stimolo. Si affianca al musicista e lo sprona a dare il meglio di sé. Se volete, rappresenta anche una sfida, perché toglie di mezzo ogni ostacolo esterno alla creatività e, quindi, ogni scusa.

4. Convinzioni resilienti III: le implicazioni politiche

Ci resta da affrontare una delle questioni più spinose sul tavolo oggi: quella del rapporto tra Intelligenza Artificiale e politica. Anche gli studenti, con la loro tipica sensibilità per le questioni etiche, hanno l'impressione che oggi siamo chiamati a misurarci con gigantesche questioni geopolitiche. Nello specifico, molti di loro temono che lo sviluppo dell'IA finisca per essere guidato solo dalla ricerca del profitto o che venga usata come strumento di potere, come arma contro specifici gruppi di persone, o per scopi sbagliati (ad esempio per controllare la gente, violandone sistematicamente la sfera della privacy). Sono poi consapevoli del rischio che una simulazione sempre più accurata dell'intelligenza umana possa distorcere la realtà creando confusione nelle persone. Secondo alcuni la capacità di simulazione delle macchine intelligenti andrebbe contrastata con piccole o grandi precauzioni, ad esempio evitando che i robot assumano sembianze umane e facciano così leva sui meccanismi dell'empatia. Il principale pericolo, insomma, è che la nuova tecnologia sfugga al controllo dei suoi creatori, producendo un classico effetto Frankenstein. In generale, molti studenti oscillano tra il timore che l'accelerazione di questa nuova rivoluzione tecnologica aumenti il rischio di catastrofi inedite per il genere umano e il sospetto che, alla fine, anche l'esito dell'*AI Revolution*, come qualsiasi altra rivoluzione tecnologica avvenuta nel corso della storia umana, dipenderà dall'uso che ne faranno le persone che hanno in mano le leve del potere.

Un modo per riassumere la questione è constatare che ci troviamo di fronte a tre grandi alternative: (a) la via del *Big State* (Cina); (b) quella della *Big democracy* (UE); quella del *Big business* (USA). Sappiamo che l'Europa ha sposato la prospettiva di una via democratica, partecipativa, umanocentrica all'IA, promuovendo un enorme sforzo legislativo a più livelli, ispirato all'ideale dell'*accountability* democratica. Sicuramente è uno sforzo di cui andare fieri. Ma qual è il vostro parere sugli effetti reali dell'*AI Act* europeo sulla ricerca? Abbiamo compiuto qualche passo significativo nella direzione di una *trustworthy AI*? Se veramente ci preme avere un'Intelligenza Artificiale centrata sull'umano, non dovremmo cercare anzitutto di impedire la concentrazione del potere politico, economico, mediatico nelle mani di pochi tecnomiliardari?

Michela Milano: Partiamo dall'*AI Act*. Non c'è dubbio che abbiamo assistito a un grande sforzo della Commissione europea. È partito un po' tardi, ma lo sforzo è stato ammirevole. Ci sono lacune, ma è un buon punto di partenza. L'idea generale è che l'IA vada regolata non solo rispetto agli usi intenzionalmente malevoli, ma anche tenendo presente le conseguenze indesiderate che derivano dai pregiudizi diffusi attualmente nella società. I tentativi di correzione non sono facili, ma non per colpa delle macchine. Il problema principale è come siamo fatti noi esseri umani. Il vero guaio sono i nostri *bias*, cioè i pregiudizi contro le donne, contro le minoranze, che sono già all'opera nelle nostre società. Non è certo colpa dei computer se i principali CEO sono tutti maschi e bianchi e le poche donne a capo di grandi aziende sono e si presentano come delle specie di *superwomen*. È forse colpa dell'Intelligenza Artificiale se quando si pensa alla segretaria di un'azienda scatta subito l'associazione con una donna e che i sottintesi sessuali non manchino mai? Per tacere del fatto che per noi italiani un addetto alle pulizie è automaticamente una "donna delle pulizie"…

Da questo punto di vista, vogliamo che l'IA *non* ci assomigli. "Human–centered", quindi, non va inteso come "human–like". "Umanocentrico" significa che la persona deve restare al centro dello sviluppo di queste nuove potentissime tecnologie. Non dimentichiamoci che il chatbot Tay di Microsoft, addestrato su milioni di messaggi di Twitter, è stato spento dopo solo sedici ore nel 2016, perché nelle

sue repliche agli utenti del social network aveva immediatamente perso qualsiasi inibizione. Per come è fatto oggi l'ambiente digitale, il contesto non ci aiuta e non alimenta la speranza o la fiducia. Dobbiamo trovare altre vie rispetto alla mera simulazione di ciò che siamo. Almeno per quanto riguarda i numerosi difetti della nostra specie è meglio, a conti fatti, che l'Intelligenza Artificiale non ci assomigli. L'obiettivo che dovremmo porci è piuttosto quello di orientare gli sviluppi dell'AI in modo da correggere o compensare le ingiustizie presenti nella nostra società. È la direzione in cui si sta muovendo, per esempio, Francesca Rossi, ricercatrice di punta dell'IBM, che ha messo in piedi un consorzio per spingere verso un'integrazione dei principi etici nei sistemi decisionali.

Per quanto riguarda invece l'impatto della nuova normativa europea, stiamo facendo partire in FBK un'iniziativa sull'*AI Act* che dovrebbe aiutarci a capire i suoi effetti concreti sulle ricerche di frontiera nel campo dell'Intelligenza Artificiale.

Paolo Traverso: Il problema geopolitico è reale. I pericoli li conosciamo bene e già l'esito sorprendente del referendum sulla Brexit ci aveva messo in guardia rispetto ai rischi di una distorsione del dibattito pubblico attraverso un uso spregiudicato dell'IA nei social network. Servirebbero più controlli, più regolamentazione, ma questa è spesso ostacolata dall'esistenza di un oligopolio informatico. Il problema non è nuovo nella storia, ma è stato sicuramente esacerbato dalle dimensioni attuali di un mercato globalizzato. Il ruolo di capofila dell'Europa in uno scenario così ricco di insidie è innegabile. Se non possiamo essere la potenza egemone in termini economici o militari, cerchiamo almeno di esserlo in termini di civiltà e progresso. Da questo punto di vista, il riferimento alla centralità dell'umano è prezioso e direi persino imprescindibile. Metterei tuttavia in guardia anche dal rischio di un eccesso di regolamentazione. Non credo che la soluzione sia quella di far andare una Ferrari a venticinque chilometri all'ora, come ho sentito dire da qualcuno.

In Europa, tenuto conto delle nostre dimensioni, della nostra storia e della nostra sensibilità ambientale e democratica, credo che ci sia spazio per interventi meno mastodontici e meno dispendiosi, anche da un punto di vista energetico. Sappiamo, infatti, che qualsiasi innovazione

tecnologica della portata della *AI Revolution* non può essere a costo zero. E i cambiamenti sono appena cominciati. Fra un po' i computer non avranno più le tastiere ed entreremo in quello che, per chi ha la mia età, è uno scenario alla "Star Trek": con le nostre macchine intelligenti parleremo, daremo loro ordini, sbarazzandoci di mouse e tastiere. Anche la lotta tra le Big Tech per i nuovi motori di ricerca è appena cominciata. Con l'IA andremo a caccia d'informazioni in maniera molto più diretta. Tutto ciò avrà dei costi ambientali, che al momento sono prevedibili solo in parte.

Michela Milano: D'altronde, nelle nostre vite non c'è solo il business. In Europa lo capiamo forse meglio che altrove perché il welfare state è stato inventato qui. Una delle priorità del futuro, quindi, sarà usare i sistemi di Intelligenza Artificiale per rendere il welfare più efficiente ed equo. Come diceva prima Paolo, c'è un problema di sostenibilità economica in ambiti della vita in cui l'intervento pubblico resterà fondamentale. Pensiamo agli effetti del cambiamento climatico. In Emilia Romagna l'attenzione è ai massimi livelli per via delle alluvioni susseguitesi nel corso degli ultimi anni. Proprio il settore del *disaster recovery*, della sostenibilità intesa in un senso ampio, inclusa la protezione delle nostre infrastrutture informatiche in caso di catastrofi naturali o artificiali, è un ambito di ricerca in cui investimenti orientati non al profitto, ma al bene comune, potrebbero produrre ottimi risultati.

5. Conclusione: le questioni aperte

A conti fatti, per riassumere quanto abbiamo detto finora, la questione centrale sembra essere quella che passa in genere sotto il nome di "Trustworthy AI", un'intelligenza artificiale capace di creare attorno a sé un'atmosfera di fiducia, anziché di sospetto generalizzato. Per usare il nostro lessico è un'IA che si sviluppa in un ambiente sociale favorevole sia da un punto di vista epistemico (le persone capiscono ciò che sta avvenendo e non se ne fanno un'immagine distorta in senso positivo o negativo) sia da un punto di vista pratico (le persone vigilano sugli usi

dell'Intelligenza Artificiale e si impegnano per favorire quelli che contribuiscono al progresso della società).

Ma esistono modi efficaci per costruire un ambiente epistemico non dominato dallo scetticismo, dalla sfiducia, dal sospetto nei confronti di una rivoluzione tecnologica con cui dovremo fare i conti tutti, senza eccezione, nel futuro prossimo? Non possiamo infatti sorvolare sul fatto che, come ci ha appena fatto notare Michela Milano, questa auspicabile cultura della fiducia deve fare i conti non solo con credenze resilienti ragionevoli, ma con la proliferazione di credenze resilienti più inquietanti che pure, come sappiamo, popolano l'ambiente epistemico delle nostre società: pregiudizi, paranoie, fissazioni, fantasie persecutorie, antipatie viscerali. Esistono convinzioni dure a morire con cui dovete fare i conti nella vostra comunità di ricerca?

Michela Milano: I pregiudizi esistono ovunque. Io, però, vorrei richiamare l'attenzione sui nostri pregiudizi nei confronti delle macchine. Gli esseri umani, come sappiamo, sbagliano. E non sempre si tratta di errori innocenti. I giudici che devono scegliere tra severità e clemenza spesso optano per l'una o per l'altra per motivi che hanno più a che fare con l'umore che con la riflessione ponderata. Abbiamo studi empirici che lo dimostrano. Poi, però, ci indigniamo se il software COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) manifesta dei bias razziali nel valutare i rischi di recidiva dei carcerati, senza chiederci se non siano inferiori a quelli dei giudici in carne e ossa…

Qualcosa di analogo capita con le macchine a guida automatizzata. Siamo circondati da persone che provocano incidenti perché si distraggono per colpa del telefonino, della stanchezza, di un colpo di sonno o di comportamenti ancora più irresponsabili, però gli errori delle macchine ci appaiono inaccettabili e imperdonabili. Perché questo doppio standard? Siamo tornati lì dove siamo partiti: nel campo dei sentimenti. Vale sempre la pena chiedersi, perciò, se l'interferenza delle emozioni ci aiuti o no a mettere le cose nella giusta prospettiva.

Paolo Traverso: A mio avviso il concetto di "trustworthiness" è fondamentale: è la chiave di tutto ciò che abbiamo discusso oggi. Come

facciamo ad alimentare la fiducia e ridurre la sfiducia della gente, che magari conosce poco i temi di cui abbiamo parlato finora? Il mio suggerimento, in parole povere, è andare nella direzione di sistemi che magari ci stupiscano un po' meno, ma siano più affidabili. Una *reliable AI*, che abbia un impatto concreto sulla vita delle persone senza indulgere nell'effetto "Star Trek": questo mi sembra un grande tema di ricerca per il futuro prossimo. Io punterei, quindi, su una Intelligenza Artificiale meno sbalorditiva, basata su un affiancamento costante degli umani che trasferisca alle macchine quel senso del limite che spesso manca anche a noi Homo sapiens del XXI secolo. A questo scopo, è essenziale mettere insieme tutte le competenze (tecnico-scientifiche, ma anche umanistiche) per evitare quello che Stephen Hawking, e altri dopo di lui, hanno paventato, forse esagerando un po', cioè la fine della specie umana per opera dell'IA.

Tornando alla questione della prevenzione sanitaria, il primo esempio che mi viene in mente è il lavoro che stiamo facendo con la dottoressa Maria Chiara Malaguti per monitorare i rischi di peggioramento nei malati di Parkinson. Il nostro obiettivo non è fare miracoli, ma offrire ai medici strumenti, magari imperfetti, ma continuamente correggibili. Non sarà la trasformazione globale dell'ambiente epistemico o sociale che avete in mente voi quando ragionate sulle convinzioni resilienti, ma è comunque una strategia dei piccoli passi che potrebbe familiarizzare gradualmente le persone con l'utilità delle nuove tecnologie in ambiti in cui i progressi si possono toccare con mano e non sono eticamente controversi perché servono per alleviare le sofferenze umane.

Il diritto, le leggi, sono ovviamente importanti per rassicurare le persone e cementare la loro fiducia nel fatto che le nuove tecnologie non verranno usate contro di loro. Parlando in generale, i giuristi oggi oscillano tra entusiasmo e cautela. Non hanno problemi a liberalizzare l'uso di tecnologie dell'IA quando i benefici per la salute umana sono evidenti al di là di ogni ragionevole dubbio. Sulla sostituzione del giudice con una macchina prevalgono invece le perplessità. Se è indubbio, infatti, che i giudici sono pieni di pregiudizi e nella storia dell'umanità ne hanno combinate di tutti i colori, tuttavia la spiegabilità delle sentenze resta un requisito fondamentale per arrivare a una giustizia sempre più

"giusta" e a una forma di *accountability*. Un processo decisionale totalmente automatizzato non sembra garantire questo diritto a sapere come si sia giunti al verdetto in questione. E c'è poi la questione dell'applicazione delle norme in un quadro generale così fluido. Da questo punto di vista, nemmeno uno sforzo di regolamentazione ambizioso come l'*AI Act* può rappresentare un traguardo definitivo.

Paolo Traverso: La sfida secondo me sta nel mettere insieme competenze diverse. Da questo punto di vista la collaborazione con i giuristi è fondamentale e anche noi "tecnici" dovremmo imparare a tenere a bada la nostra impazienza e apprezzare i pregi di un sapere più paziente. Forse ci farebbe bene recuperare un po' di logica aristotelica e fare tesoro di ciò che vi è di saggio nel senso comune, nelle intuizioni e nei bisogni delle persone che subiscono le conseguenze più pesanti dei grandi cambiamenti storici.

Ma, tirando le somme, alla fine le macchine possono pensare o no? E che tipo di pensieri sono in grado di produrre?

Michela Milano: Dovendo esprimermi fuori dai denti, direi chiaro e tondo che secondo me le macchine non pensano. Possono scimmiottare un ragionamento, però l'esperienza del pensiero per come la facciamo noi dall'interno, un pensiero che è *tuo* in senso proprio, non appartiene alla macchina. Intelligenti sì, ma non pensanti, è la mia conclusione, stante così le cose.

Paolo Traverso: Le macchine calcolatrici pensano? Turing se lo chiedeva in un saggio che ha fatto epoca: *Computing Machinery and Intelligence* (Turing 2025). E la sua curiosità si spingeva ben al di là del test che ha poi escogitato e che è passato alla storia proprio come test di Turing. Anche lui era consapevole che, al di là degli agenti conversazionali, resta irrisolto il problema dell'*embodiment*, delle implicazioni dell'incarnazione dell'intelligenza in un corpo e delle sue interazioni col mondo. La macchina magari può arrivare a pensare astrattamente, elaborando dei simboli o identificando dei pattern, ma il fatto che "pensi" non è certo la soluzione a tutti i problemi che ruotano attorno

alla realtà della coscienza umana. Anche le innovazioni più spericolate e sbalorditive hanno bisogno di un contesto per dare i frutti che contengono potenzialmente dentro di sé. Oggi siamo tutti impegnati nel tentativo di creare il contesto più ospitale e fiducioso affinché la nuova forma di intelligenza che abbiamo creato ci aiuti a costruire società migliori, più umane. Speriamo che i nostri sforzi vadano a buon fine.

Riferimenti Bibliografici

Cristianini N. (2025) *Sovrumano. Oltre i limiti della nostra intelligenza*, Il Mulino, Bologna.

Turing A. (2025) *Macchine calcolatrici e intelligenza*, a cura di D. Marconi, Einaudi, Torino.