Macchine creative? Intervista a Mario De Caro

di Giacomo Bottos

26-02-2026

Il 26 e 27 febbraio si è tenuto a Genova "Filosofia, Estetica Cooperativa e Intelligenza Artificiale", tappa nazionale della Biennale dell'Economia Cooperativa promossa da Legacoop, dedicata alla costruzione di una diversa "episteme" dell'intelligenza artificiale. Due giornate di confronto, workshop e dialoghi tra filosofia, scienza e impresa per approfondire temi quali le implicazioni etiche dell'intelligenza artificiale, l'interazione uomo-macchina, la governance dei dati e il ruolo delle piattaforme digitali come beni pubblici orientati a modelli cooperativi e inclusivi.
Pandora Rivista è media partner dell'evento, che ha visto tra gli ospiti principali Mario De Caro, professore ordinario di Filosofia morale all'Università Roma Tre dove è titolare della Cattedra UNESCO in Etica e intelligenza artificiale e visiting professor alla Tufts University (Stati Uniti). De Caro si occupa di libertà e delle implicazioni etiche e teoriche delle tecnologie digitali e delle scienze cognitive, e ha tenuto un keynote speech dal titolo "Creatività artificiale: scienza, arte e democrazia". Lo abbiamo intervistato per approfondire i temi al centro del suo intervento. Il programma completo dell'evento è disponibile sul sito di Legacoop Liguria.

Quali definizioni possiamo adottare del termine "intelligenza"? Ha senso parlare di intelligenze artificiali in un'accezione che mantenga un legame con ciò che intendiamo quando parliamo di intelligenza in riferimento agli esseri umani?
Mario De Caro: In ambito umano esistono diverse definizioni di intelligenza. Secondo la teoria delle intelligenze multiple di Howard Gardner, ad esempio, non esisterebbe un'unica intelligenza, bensì una pluralità di forme relativamente autonome, non coordinate da un principio unitario. Oltre alle capacità misurate dai tradizionali test - come quelle logico-matematiche, linguistiche e spaziali - Gardner ha individuato altre forme, dall'intelligenza corporeo-cinestetica a quella sociale ed esistenziale. Questa impostazione è stata tuttavia criticata da chi sostiene che, pur riconoscendo la varietà delle competenze, esista comunque un'intelligenza generale unitaria, cioè una capacità trasversale di comprendere situazioni e problemi in ambiti differenti e di individuare strategie adeguate ad affrontarli. I problemi cambiano a seconda dei contesti, ma la facoltà di risolverli, secondo questa visione, è fondamentalmente unica. Le macchine potranno mai raggiungere questa capacità a un livello umano? È evidente che, qualora vi riuscissero, sarebbero molto più rapide di noi nell'esecuzione dei compiti, e si parlerebbe di artificial general intelligence (AGI). Altri, in modo ancora più audace, ipotizzano uno stadio ulteriore, la superintelligenza, in cui le macchine non solo supererebbero le prestazioni umane, ma potrebbero anche sfuggire al nostro controllo, compiendo azioni che non saremmo in grado di prevedere o comprendere - cosa che in parte già accade, dato che i sistemi attuali non sono pienamente deterministici.
Secondo le visioni più fantascientifiche, il momento del "sorpasso", in cui l'intelligenza artificiale diverrebbe incontrollabile, coinciderebbe con la cosiddetta singolarità tecnologica, tema su cui ha scritto ampiamente Ray Kurzweil. A prima vista, tutto ciò potrebbe sembrare relativamente semplice da verificare: basterebbe controllare se le macchine sono in grado di fare ciò che sappiamo fare noi. Ma qui sorge una domanda cruciale: noi chi? Spesso si pretende che i sistemi artificiali raggiungano, nei diversi ambiti, il livello dei migliori esseri umani. Demis Hassabis, per esempio, ha dichiarato che parlerebbe di vera intelligenza artificiale generale se una macchina, posta nelle

condizioni in cui si trovava Einstein nel 1911 e dotata delle conoscenze allora disponibili, fosse capace di formulare la teoria della relatività generale. Analogamente, si potrebbe chiedere che in matematica eguagli Andrew Wiles o che in poesia raggiunga la brillantezza di Umberto Saba o Giuseppe Ungaretti. Con criteri di questo tipo, un sistema capace di operare al livello di una persona mediamente intelligente in ogni ambito non sarebbe comunque considerato davvero "intelligente". Esiste un aforisma ironico, la cosiddetta Legge di Tesler, secondo cui l'intelligenza è tutto ciò che le macchine non sanno ancora fare. Un tempo il criterio era il superamento del test di Turing; oggi i large language model, e ancor più i sistemi multisensoriali o neurosimbolici, lo superano agevolmente, producendo risultati spesso indistinguibili da quelli umani, anche in ambiti come la poesia o la musica.

Nel dibattito esistono posizioni, tra le più note quella di Luciano Floridi, che negano per principio che l'intelligenza artificiale possa essere definita davvero "intelligente". Come risponde a queste tesi?
Mario De Caro: Recentemente ho recensito l'ultimo libro di Luciano Floridi, La differenza fondamentale (Mondadori), su Tuttolibri de La Stampa. Floridi è presidente onorario della Società Internazionale per l'Etica e la Politica dell'Intelligenza Artificiale (SEPAI), di cui io sono presidente. Le sue argomentazioni sono serie e ha ragione quando invita alla prudenza rispetto alla tendenza diffusa a proclamare troppo rapidamente il raggiungimento dell'intelligenza artificiale generale. Dove mi trovo in disaccordo è sull'idea che, in linea di principio, l'intelligenza artificiale non possa mai raggiungere certi traguardi, in particolare la "vera creatività". Floridi sostiene che le macchine non potranno mai essere realmente creative; io ritengo non solo che possano esserlo, ma che in parte lo siano già, almeno nei sistemi più avanzati. Un argomento classico contro le intelligenze artificiali era che commettevano errori anche in semplici calcoli matematici. È vero: tali sistemi operano prevalentemente in modo bottom-up e, direi, abduttivo, cioè formulano ipotesi sulla base delle strutture neurali costruite durante l'addestramento. L'abduzione è epistemicamente ampliativa, ma può condurre a errori. In matematica, però, non servono ipotesi bensì deduzioni. La "vecchia" intelligenza artificiale simbolica, meno creativa ma più rigida, non sbagliava nelle deduzioni. Oggi, con i sistemi neurosimbolici, si integrano i punti di forza di entrambe le tradizioni: induzione e abduzione da un lato, deduzione rigorosa dall'altro. E in matematica si registrano risultati notevoli. Il cap set problem, per esempio, è stato affrontato con l'ausilio di sistemi algoritmici avanzati. Un caso significativo è FunSearch, sviluppato da DeepMind: il sistema ha prodotto la dimostrazione di una congettura matematica, pur senza che i programmatori potessero ricostruire nel dettaglio il percorso seguito.
Risultati impressionanti si registrano anche nelle Olimpiadi della matematica, dove vengono proposti problemi inediti che l'intelligenza artificiale non può aver "memorizzato" e che tuttavia risolve meglio degli esseri umani. Se questo non implica una forma di creatività matematica, è difficile capire che cosa lo sarebbe. Lo stesso vale per le arti. In musica, alcune composizioni generate da sistemi avanzati sono difficilmente distinguibili da quelle umane. In poesia, uno studio pubblicato su Nature da Brian Porter e Edouard Machery ha mostrato che molti partecipanti non erano in grado di distinguere testi di autori celebri da quelli prodotti da ChatGPT, e spesso attribuivano a esseri umani le composizioni generate dall'IA. Anche negli scacchi l'evoluzione è significativa. Quando Deep Blue sconfisse Garry Kasparov, la vittoria dipendeva in larga misura dalla potenza computazionale applicata a principi di gioco elaborati dagli umani. Con AlphaZero, presentato da DeepMind nel 2017, il paradigma cambia: il sistema non procede per calcolo esaustivo, ma affronta la posizione come un problema, individua punti di forza e debolezza e sviluppa linee strategiche in modo abduttivo, in maniera analoga - e spesso superiore - ai grandi maestri umani. Trovo dunque difficile negare che le macchine manifestino una forma di creatività genuina. Certo, manca il creatore: non vi è coscienza, né intenzionalità, né volontà dell'autore - quella che in filosofia si chiama voluntas

auctoris - e le macchine non hanno esperienza soggettiva. Ma l'assenza di coscienza non implica automaticamente l'assenza di creatività nei prodotti.

Si potrebbe obiettare che, nonostante in molti ambiti i risultati delle macchine siano comparabili o persino superiori a quelli umani, esse non abbiano una reale comprensione di ciò che fanno. Naturalmente bisognerebbe chiarire che cosa intendiamo per "comprensione", ma si potrebbe sostenere che il semplice raggiungimento di determinati traguardi non dimostri, di per sé, che le macchine possiedano una qualche forma di intelligenza?
Mario De Caro: Questo è un punto decisivo. Ne parlo nel libro che ho scritto con Benedetta Giovanola, Intelligenze. Etica e politica dell'IA, pubblicato da il Mulino, contestando questa obiezione. A mio avviso, sulla questione della comprensione aveva ragione Ludwig Wittgenstein. Già negli anni Trenta, a Cambridge, in dialogo anche con Alan Turing, egli sosteneva - posizione poi sviluppata nelle Ricerche filosofiche - che il modo in cui, dal punto di vista meccanico o causale, arriviamo a usare il linguaggio è filosoficamente irrilevante. Come ha riassunto il suo allievo Norman Malcolm, nothing is hidden: non c'è nulla di occulto dietro l'uso. Ciò che conta è la capacità di seguire correttamente le regole all'interno di pratiche linguistiche condivise. Io so di essere cosciente perché quando dico qualcosa - meno spesso di quanto mi piacerebbe credere - ne conosco le ragioni; ho coscienza di me stesso. Ma quando parla lei, ad esempio, le attribuisco una mente e una comprensione perché usa correttamente le regole della lingua, in questo caso l'italiano. Ma non so nulla di come funzioni la sua testa - che, dice Wittgenstein, potrebbe perfino essere fatta di formaggio. Quello che conta è che, nello spazio pubblico, lei mostri di saper seguire le regole del linguaggio.
Inoltre, per Wittgenstein non è concepibile un linguaggio puramente privato, perché la nozione stessa di linguaggio è parassitaria rispetto a un uso pubblico. Senza un linguaggio pubblico non avremmo standard di correttezza; non potrei nemmeno sapere, parlando il mio presunto linguaggio privato, se la parola "sedia" oggi significa la stessa cosa che significava ieri. Il criterio di identità del significato dipende da pratiche condivise. Dunque, ciò che conta sono l'uso pubblico e la capacità di seguire le regole del linguaggio pubblico. Questo è lo standard che applichiamo, tra esseri umani, per attribuirci reciprocamente comprensione. Per quale ragione non dovremmo adottare questo standard nel caso delle macchine?

Quindi, in fondo, anche noi umani siamo delle black box?
Mario De Caro: Molti lo sostengono, perché il cervello resta in larga parte un mistero. Ma per Wittgenstein anche se disponessimo di tutta la neurofisiologia possibile, ciò sarebbe irrilevante ai fini dell'attribuzione di comprensione. Il punto non è come funziona internamente un sistema, bensì come opera nello spazio pubblico. Proviamo a riformulare la questione con un esperimento mentale, ispirato in parte a Steven Pinker, che cito nel libro Intelligenze. Immaginiamo un gruppo di astronauti che atterra su un pianeta popolato da esseri molto diversi da loro ma che sembrano comprendere: si può dialogare con loro, discutere, scambiarsi argomentazioni. Ora immaginiamo che gli astronauti concludano che quegli esseri "non capiscono", perché sono fatti di carbonio e non di silicio. Abbiamo semplicemente rovesciato la prospettiva: gli alieni siamo noi, e siccome siamo diversi, si conclude che non comprendiamo. Ma la costituzione materiale è irrilevante, ciò che conta è la capacità di seguire regole condivise. Se le macchine arriveranno a usare correttamente il linguaggio secondo standard condivisi, allora dovremo dire che comprendono.
Diversa è la questione della coscienza. Al momento non abbiamo motivo di attribuire una coscienza alle macchine. C'è chi sostiene che l'avranno presto, come Dario Amodei, che ipotizza tempi brevissimi, e chi, come Geoffrey Hinton, premio Nobel per la fisica nel 2024, arriva a dire che forme embrionali di coscienza potrebbero già esserci. Alcuni, come Luciano Floridi, Maurizio Ferraris o

John Searle, sostengono invece che le macchine non comprenderanno mai perché non potranno mai essere coscienti. Ma io non ho questa certezza per una ragione molto semplice: non abbiamo una teoria condivisa e consolidata nemmeno della coscienza umana. Lo possiamo inferire, ed è un'inferenza fallibile. Lo stesso varrebbe, in linea di principio, per le macchine. Secondo alcune teorie, l'emergere della coscienza potrebbe avvenire anche in substrati materiali diversi dal carbonio; secondo altre, come nella prospettiva di Roger Penrose, la coscienza è intrinsecamente legata a specifiche proprietà biologiche. Inoltre, alcuni, come Searle, sostengono che senza coscienza non possa esserci comprensione. Ma la coscienza non è né condizione necessaria né condizione sufficiente della comprensione, anche se spesso la accompagna: non è condizione necessaria, perché in molti casi comprendiamo senza essere coscienti del momento in cui la comprensione si instaura; e non è nemmeno condizione sufficiente, perché siamo spesso coscienti di aver capito qualcosa quando in realtà non l'abbiamo capita affatto. L'esperienza soggettiva di "aver compreso" può essere del tutto fallace. C'è dunque uno scarto tra coscienza e comprensione e questo rende molto più difficile sostenere che l'assenza di coscienza nelle macchine implichi l'impossibilità di comprendere.

Torniamo al concetto di creatività, che lei ha già introdotto e che compare anche nel titolo della lezione che terrà a Genova, "Creatività artificiale, scienza, arte e democrazia". Può elaborare ulteriormente questo tema, sia per quanto riguarda scienza e arte, di cui abbiamo già parlato, sia sul versante della democrazia?
Mario De Caro: Sulla creatività in ambito scientifico e artistico abbiamo richiamato alcuni esempi significativi che mostrano come i sistemi più avanzati non si limitino a replicare schemi esistenti, ma producano soluzioni nuove, talvolta imprevedibili, e valutabili secondo criteri interni alla disciplina. Nella scienza, la creatività consiste nella formulazione di ipotesi feconde, nell'individuazione di connessioni inattese, nella costruzione di dimostrazioni eleganti. Nell'arte, riguarda la capacità di generare forme che producano un valore estetico riconoscibile all'interno di una tradizione. In entrambi i casi, le macchine stanno mostrando prestazioni che rendono sempre più difficile negare l'esistenza di una creatività, almeno funzionale. Il nodo della democrazia introduce una dimensione diversa, politica e istituzionale. Nel settembre 2025, un gruppo composto da venticinque premi Nobel e vincitori del Turing Award, insieme ad altre centinaia di esperti, ha presentato alle Nazioni Unite una mozione in cui si chiedeva l'introduzione di vincoli etici stringenti per lo sviluppo dei sistemi di intelligenza artificiale. L'argomento era che tali sistemi potrebbero progressivamente autonomizzarsi e sfuggire al controllo umano, costituendo una minaccia di portata sistemica. Altri studiosi - tra cui Luciano Floridi e Adriano Fabris - ridimensionarono questo scenario sostenendo che l'intelligenza artificiale, in quanto artefatto, resta strutturalmente dipendente dai suoi progettisti. Se è davvero intelligente, allora non è pienamente artificiale; se è artificiale, non può essere veramente autonoma. In questa prospettiva, il rischio di perdita di controllo è sopravvalutato.
Ma anche ammesso che i sistemi non diventino mai autonomi in senso forte, resta il fatto che possono essere impiegati intenzionalmente per fini dannosi - manipolazione dell'informazione, sorveglianza di massa, guerra automatizzata - oppure in modo irresponsabile, per negligenza o incompetenza. Questo è un rischio reale e attuale, indipendente dall'ipotesi della superintelligenza. Il secondo argomento - su cui insiste molto Floridi - riguarda la retorica dell'emergenza. È vero che alcuni protagonisti della rivoluzione dell'IA tendono a enfatizzare scenari estremi, anche per attrarre l'attenzione mediatica e i finanziamenti. In passato vi sono già stati due AI winters, periodi di drastico calo di fondi e di fiducia, e un nuovo inverno dell'intelligenza artificiale avrebbe ripercussioni sull'intero sistema finanziario globale. Detto questo, ritengo che il problema dell'autonomizzazione non possa essere escluso a priori, anche indipendentemente dalla questione della coscienza. Un sistema può sfuggire al controllo anche perché acquisisce capacità decisionali complesse,

interagisce con altri sistemi, opera in ambienti dinamici e produce effetti cumulativi non pienamente prevedibili. L'autonomia funzionale non coincide con l'autocoscienza. Se poi si dovesse porre il problema della coscienza artificiale, la questione assumerebbe una gravità ancora maggiore. Ma il punto è che il tema democratico è già oggi urgente e riguarda la distribuzione del potere, la trasparenza degli algoritmi, la responsabilità delle decisioni automatizzate, la tutela delle istituzioni rappresentative. In questo senso, la creatività e la democrazia non sono ambiti separati, poiché la creatività delle macchine incide sulla produzione di conoscenza e di cultura; e la produzione di conoscenza e cultura incide direttamente sulla qualità delle nostre democrazie. Il problema non è solo se le macchine siano creative, ma chi controlla quella creatività, a quali fini vengono orientate e secondo quali regole collettive sono governate.

Quali sono, a suo avviso, i rischi più rilevanti da questo punto di vista?
Mario De Caro: Il primo rischio riguarda il fatto che stiamo già affidando ai sistemi di intelligenza artificiale compiti che essi svolgono in larga misura in autonomia e che non possiamo controllare pienamente, per almeno due ragioni. La prima è che si tratta di meccanismi probabilistici, non deterministici: non è possibile prevedere con certezza quale output produrranno in una determinata circostanza. La seconda è il problema dell'explainability: le reti neurali contemporanee operano attraverso miliardi di parametri e connessioni e la loro complessità è tale che, anche quando producono un risultato corretto, spesso non siamo in grado di ricostruire in modo trasparente il percorso inferenziale che ha condotto a quell'esito. Questo punto è particolarmente delicato in ambito medico. Supponiamo che un sistema diagnostico raggiunga un'accuratezza superiore a quella media dei medici umani. Se però giunge alla diagnosi basandosi su correlazioni spurie o su indizi irrilevanti, può produrre errori con conseguenze gravi. In uno studio sulle malattie rare, per esempio, i medici individuavano correttamente la diagnosi in circa il 20% dei casi, mentre i sistemi di IA arrivavano all'80%. Ma per quali ragioni il sistema "indovina"? Sta cogliendo strutture causali rilevanti o si fonda su correlazioni accidentali? Un esempio classico chiarisce il problema. Un sistema addestrato a distinguere immagini di lupi da immagini di husky raggiungeva un'accuratezza del 95%, ma continuava a commettere errori. Si scoprì che non distingueva gli animali in base alle loro caratteristiche morfologiche, bensì allo sfondo: nel dataset di addestramento il lupo compariva spesso sulla neve, l'husky no. In realtà classificava la neve, non l'animale. Se si tratta di lupi e husky l'errore è relativamente innocuo; se si tratta di diagnosi mediche o decisioni terapeutiche, il rischio è evidente.
Il problema democratico nasce perché ciò che accade in medicina accade anche in altri ambiti - militare, finanziario, assicurativo, giudiziario. In alcune giurisdizioni, come la California, gli algoritmi contribuiscono già a determinare la concessione della libertà condizionale; influenzano l'accesso al credito, la stipula di assicurazioni, la selezione del personale. Si tratta di decisioni che incidono direttamente sui diritti e sulle opportunità dei cittadini e che vengono prese, almeno in parte, da sistemi opachi. A questo si aggiunge un ulteriore livello di rischio, legato all'informazione e alla formazione del consenso politico. Le tecniche di profilazione basate sui big data consentono una personalizzazione estremamente sofisticata dei messaggi, come mostrato dallo scandalo Cambridge Analytica o nel contesto della Brexit. Più in generale, vi è il sospetto che potenze straniere tentino di influenzare consultazioni elettorali sfruttando la struttura a eco-chamber dei social media. Già oggi, dunque, sistemi complessi e poco trasparenti possono sfuggire al controllo umano in senso funzionale, incidendo su processi vitali per le nostre società. Il rischio che l'intelligenza artificiale ci sfugga di mano non è uno scenario fantascientifico, ma una possibilità concreta che riguarda il modo in cui regoliamo la tecnologia, organizziamo le istituzioni e distribuiamo il potere decisionale.

Rispetto a questo quadro, quali sono, a suo avviso, le azioni prioritarie? Su quali fronti bisognerebbe intervenire con maggiore decisione?
De Caro: Alcuni tentativi - ancora parziali - sono stati compiuti, soprattutto nell'Unione Europea. L'AI Act rappresenta senza dubbio un risultato significativo. Tuttavia, è evidente la lentezza del diritto rispetto alla velocità dell'innovazione tecnologica. L'AI Act entrerà pienamente in vigore nel 2027, ma si fonda su un contesto tecnologico sostanzialmente fotografato nel 2023. Nel frattempo, in pochi anni, i sistemi di intelligenza artificiale hanno compiuto progressi radicali. Il secondo problema è geopolitico. L'Europa, su questi temi, ha un peso relativamente limitato rispetto ai principali attori globali - Stati Uniti, Cina e, in misura crescente, India. Se uno di questi poli decide di deregolamentare, come è accaduto negli Stati Uniti con l'allentamento delle restrizioni federali sulla ricerca in IA, si crea una dinamica competitiva che rende difficile mantenere standard elevati a livello regionale. La corsa tecnologica rischia di prevalere sulle cautele etiche. Da qui il terzo livello del problema: la necessità di un coordinamento sovranazionale. Senza un organismo internazionale con effettiva capacità di indirizzo - le Nazioni Unite o un'agenzia ad hoc - è difficile stabilire regole comuni. In passato, accordi internazionali hanno vietato le armi batteriologiche o regolamentato determinate tecnologie militari. Analogamente, occorrerebbe intervenire su applicazioni dell'IA ad alto rischio, per esempio l'uso di sistemi autonomi armati, come droni capaci di identificare e colpire bersagli senza un controllo umano significativo. Qui siamo di fronte a questioni che toccano direttamente il diritto umanitario internazionale.
C'è poi un ulteriore tema, spesso trascurato. Di solito si pensa che la soluzione consista nel limitare o normare la ricerca, che è certamente necessario regolare. Ma questo non basta. La prospettiva più promettente è quella definita ethics by design: integrare principi etici direttamente nell'architettura dei sistemi. Non limitarsi a controllare dall'esterno, ma progettare modelli che incorporino vincoli normativi, criteri di trasparenza, meccanismi di accountability e di allineamento ai valori fondamentali - diritti umani, equità, non discriminazione. È su questo che insistono molti dei premi Nobel e dei Turing Award che hanno sollevato l'allarme internazionale: non solo regolamentazione ex post, ma progettazione responsabile ex ante. Se i sistemi diventano sempre più autonomi sul piano funzionale, l'etica non può essere un'aggiunta successiva, ma una componente strutturale della tecnologia stessa. Servono almeno quattro linee di intervento: accelerare l'adattamento del diritto alla velocità dell'innovazione; rafforzare il coordinamento internazionale; regolamentare ricerca e applicazioni ad alto rischio; e, soprattutto, sviluppare una vera integrazione tra progettazione tecnica e principi etici. Solo così si può sperare di governare l'evoluzione dell'intelligenza artificiale senza subirla.

Su questo punto - l'ethics by design - quali sono oggi gli approcci e le principali linee di discussione?
Mario De Caro: Siamo ancora in una fase iniziale. Quando si parla di etica e intelligenza artificiale, l'attenzione si concentra spesso su come usare i sistemi per migliorare noi stessi dal punto di vista morale. Ma questo è un compito secondario. Il problema urgente è che i sistemi agiscano entro vincoli etici operativi. Prendiamo un esempio estremo ma chiarissimo: se un drone armato individua cento bambini tra cui potrebbe trovarsi un terrorista, non deve esplodere. Non è moralmente giustificabile uccidere cento innocenti per eliminare un singolo bersaglio. Questo significa implementare limiti concreti, non limitarsi a riflessioni astratte. Nella tradizione etica moderna si distinguono due grandi famiglie teoriche. La prima è quella dei modelli principialistici, tipicamente kantiani o utilitaristi: esisterebbero principi generali - l'imperativo categorico, il principio di massimizzazione dell'utilità - da tradurre in regole computabili e da implementare nei sistemi. Ma qui emergono due difficoltà. La prima è che tali principi conoscono eccezioni, varianti e conflitti: non esiste una gerarchia univoca e universalmente condivisa tra di essi. La seconda è tecnica: l'intelligenza artificiale contemporanea funziona prevalentemente in modalità bottom-up, data-driven,

non top-down e rule-based. Non applica meccanicamente un insieme finito di regole, ma apprende da grandi quantità di dati, individua pattern e generalizza a partire da casi. Per questo ritengo più promettente un approccio ispirato all'etica della virtù aristotelica. In questa prospettiva non si parte da principi assoluti, ma dall'apprendimento situato, sviluppando quella che Aristotele chiamava phronesis, la saggezza pratica. I bambini non imparano l'etica studiando sistemi teorici, ma attraverso esempi concreti. Kant sosteneva che non si debba mai mentire; eppure, esistono situazioni in cui dire la verità significherebbe consegnare una persona innocente alla tortura. In questi casi il giudizio morale non deriva dall'applicazione cieca di un principio, ma dalla valutazione del contesto.
Definisco questa impostazione "contestualista": si giudica caso per caso, individuando quali principi siano rilevanti nella situazione specifica. È interessante che questa logica sia strutturalmente affine al funzionamento dei sistemi di IA, che apprendono dai dati e generalizzano da esempi. L'idea è dunque addestrare i modelli - in modo supervisionato - a riconoscere configurazioni morali ricorrenti e a reagire secondo standard condivisi. Naturalmente, nei casi controversi si aprirà un dibattito, anche in ragione delle differenze culturali e politiche. Tuttavia, esiste un ampio nucleo di convergenza su principi fondamentali: divieto di uccidere innocenti, di torturare, di discriminare arbitrariamente. Non tutto è culturalmente negoziabile. La linea più promettente è una forma di human in the loop, in cui la migliore saggezza morale umana venga progressivamente integrata nell'architettura dei sistemi, costruendo fin dall'origine modelli vincolati da criteri morali appresi e continuamente raffinati attraverso il dialogo tra competenza tecnica e riflessione etica.

Nell'ottica di sperimentare approcci diversi, più attenti alla dimensione etica e ai rischi, ma anche capaci di immaginare usi alternativi dell'intelligenza artificiale, un contesto come quello della Biennale dell'Economia Cooperativa può rappresentare un ambiente favorevole? Il mondo cooperativo, per sua natura, non persegue solo l'obiettivo economico, ma integra anche finalità sociali. Può essere un terreno adatto per sviluppare un diverso paradigma dell'IA?
Mario De Caro: Sì, direi che il mondo cooperativo rappresenta un milieu ideale, perché fondato su un principio strutturale di solidarietà. Se prendiamo sul serio l'idea - espressa anche da Sam Altman - che l'intelligenza artificiale possa rappresentare una trasformazione paragonabile, se non superiore, all'invenzione della stampa a caratteri mobili nel Quattrocento, dobbiamo riconoscere che siamo di fronte a una potenza storica straordinaria. Se utilizzata bene, può produrre progressi enormi. L'assegnazione recente del premio Nobel per la chimica a ricerche legate alla mappatura di centinaia di milioni di strutture proteiche mostra come l'IA possa accelerare radicalmente le scoperte scientifiche. Si stanno progettando e testando nuovi farmaci con risultati promettenti, potenzialmente in grado di incidere su patologie oncologiche, malattie genetiche e sviluppo di vaccini. Potremmo essere all'inizio di una delle più grandi rivoluzioni mediche della storia umana. Questo è il lato luminoso. Esistono però due nodi problematici. Il primo riguarda l'iniquità nella distribuzione dei benefici economici e del potere politico: una quota rilevante di potere è oggi concentrata nelle mani di pochi grandi attori tecnologici, con implicazioni dirette sulla qualità della democrazia. Lo si vede con particolare evidenza negli Stati Uniti, dove l'influenza dei grandi gruppi tecnologici sulle scelte politiche è ormai esplicita. Il secondo nodo è l'iniquità epistemica. Chi resta fuori dal circuito della conoscenza legata all'IA rischia una marginalizzazione crescente. Numerose stime indicano che una quota molto ampia dei lavori sarà profondamente trasformata nei prossimi anni: alcune professioni scompariranno, altre verranno ridefinite. Senza competenze adeguate, l'accesso a lavori qualificati diventerà sempre più difficile. Questo impone un ripensamento radicale dell'educazione: non è più sostenibile l'idea di studiare fino alla fine dell'università e poi entrare stabilmente nel mercato del lavoro. Occorre formazione continua lungo tutto l'arco della vita.
In questo quadro, il mondo cooperativo può svolgere un ruolo strategico, perché non è orientato

esclusivamente alla massimizzazione del profitto. Integrando obiettivi sociali, può diventare un laboratorio di innovazione tecnologica più equa, attenta alla redistribuzione dei benefici, alla partecipazione democratica e alla formazione diffusa. Se devono emergere cambiamenti positivi, difficilmente verranno dall'alto: possono nascere solo da una cittadinanza informata e attiva, capace di orientare la politica verso priorità adeguate. L'intelligenza artificiale è uno dei temi strutturali del nostro tempo, e contesti fondati sulla cooperazione e sulla solidarietà possono essere spazi privilegiati per costruire un modo diverso di governarne lo sviluppo.