

La custodia della “magnifica humanitas” e l’umanesimo digitale

Stefano Zamagni

15

i quaderni dell'economia civile

La custodia della “magnifica humanitas” e l’umanesimo digitale

Stefano Zamagni

INDICE

- 1. Introduzione e motivazione**
- 2. Linee di azione, urgenti e possibili**
- 3. Post-verità e nuovo progetto educativo**
- 4. La questione sociale nell'era dell'IA**
- 5. Capitalismo cibernetico e principio democratico**
- 6. Un'annotazione conclusiva**

LA CUSTODIA DELLA “MAGNIFICA HUMANITAS” E L’UMANESIMO DIGITALE

Stefano Zamagni

1. Introduzione e motivazione

In questo scritto, fisserò l’attenzione su taluni aspetti specifici dell’attuale situazione socio-economica e politica, sui quali la straordinaria enciclica di papa Leone XIV, *Magnifica Humanitas* (MH), getta una luce di particolare intensità - aspetti a proposito dei quali è urgente sollecitare una riflessione non circoscritta ai soli esperti. Non si dimentichi, infatti, che i nuovi strumenti del digitale nascono come macchine “economiche”, con scienza e tecnica orientate – a volte piegate – a progetti economico-finanziari. Che siano forme di capitalismo digitale, sia esso di mercato o di Stato, lo sviluppo tecnologico è chiamato a rispondere a modelli di ritorno economico sugli investimenti sempre più ingenti. È allora saggio andare oltre le familiari analisi del “capitalismo della sorveglianza”, per indagare i nuovi modi di creazione del valore e di redistribuzione della ricchezza. Si tratta di tecnologie socialmente istituenti che hanno tutte la duplice natura di produrre rappresentazioni del mondo e nuove condizioni dell’umano.

C’è un modo sbagliato – ci avverte la MH – di pensare l’IA, purtroppo assai comune: trattarla come una tecnologia singola, pur potente, paragonabile alla caldaia a vapore o al microprocessore, dotata di una sua propria traiettoria autonoma e di un proprio impatto economico misurabile. Questo modo di pensare, se può essere utile per discussioni tra imprese, non aiuta di certo a capire la trasformazione in corso, che non riguarda uno specifico settore industriale, ma una “ondata” che va mutando una pluralità di contesti di vita. Ne deriva che la questione oggi centrale

è quella di come governare – non solo regolamentare – l'IA senza ridurla né a promessa salvifica né a minaccia incontrollabile, e avendo sempre a mente la domanda: la specie umana deve o non deve continuare ad abitare la terra? Papa Leone si colloca in una posizione realista. Non basta chiedersi – insiste più volte – come rendere gli algoritmi più precisi e più sicuri, ma *chi* debba decidere il modo in cui questi strumenti vengono progettati e usati, con *quali dati*, per quali *fini*, con quali *responsabilità* e con quali *conseguenze* per coloro che ne subiscono gli effetti. L'IA non è di per sé razzista o abilista nel senso in cui può esserlo una persona; ma può incorporare o amplificare strutture sociali razziste o abiliste quando viene costruita entro società che *già* producono tali discriminazioni perché organizzate intorno a soli interessi di parte.

Il tema che definisce il nucleo duro della MH è di natura ontologica, non solo epistemica – un tema che purtroppo continua ad essere dimenticato nel dibattito pubblico. In breve, si tratta di questo. In parallelo con l'evoluzione delle nuove tecnologie, ha iniziato a diffondersi, a partire dagli anni Novanta del secolo scorso, quella corrente di pensiero filosofico che comprende due versioni – il transumanesimo e il post-umanesimo –, una corrente ormai divenuta una vera e propria ideologia fortemente sostenuta dai giganti dell'*High Tech* della Silicon Valley, dove ha sede la University of Singularity, fondata nel 2007 da R. Kurzweil e N. Bostrom. (Ad essere precisi, c'è differenza tra le due versioni: mentre il transumanesimo promuove l'IA per superare i limiti umani con l'accrescimento cognitivo e fisico, il post-umanesimo de-costruisce l'essere umano come misura universale, rifiutando le dicotomie cartesiane mente/corpo e natura/cultura, rifondando l'ontologia in prospettiva non antropologica, ma ecologica. Va detto che una terza versione si sta facendo oggi spazio: il meta-umanesimo, che si contrappone sia alla visione tecno-ottimista del transumanesimo, sia a quella de-costruttivista del post-umanesimo. Non abbiamo però qui lo spazio per occuparcene).

Ha scritto Max More nel "Manifesto" (2003) del movimento che ha anticipato la nascita dell'Università della Singolarità: "Noi mettiamo in questione il carattere inevitabile della vecchiaia e della morte; cerchiamo di migliorare progressivamente le nostre capacità intellettuali e fisiche, sviluppandole anche sul piano emozionale.

Vediamo l'umanità come una fase di transizione nello sviluppo evolutivo". (Come a dire che l'umanità è semplicemente una fase di passaggio del processo evolutivo!). Gli fa eco il filosofo oxoniense N. Bostrom, quando scrive: "Usciremo dall'infanzia dell'umanità per entrare nell'era postumana. La medicina non è chiamata solo a riparare, a ripristinare l'equilibrio, ma pure a rispondere ad un modello superiore di umano" (*Better than human*, Oxford University Press, 2011).

Con l'ideologia transumanista e post-umanista, che va riscuotendo crescenti successi soprattutto negli ambienti anglosassoni (si pensi al ruolo che vanno giocando personaggi come Peter Thiel e Mark Andressen) siamo di fronte ad una tecno-religione il cui dogma centrale è che la tecnologia sia potenzialmente onnipotente e che la capacità di scegliere tra il bene e il male non derivi da una scelta morale secondo una visione trascendente, ma dipenda dalle illimitate conoscenze acquisibili grazie agli algoritmi, che diventano così l'espressione di un nuovo gnosticismo. Come noto, per la Gnosi è la conoscenza che salva; per la Fede cristiana, invece, è Gesù che salva. Il progetto di cui si sta parlando avanza dunque la pretesa gnostica di autosalvazione. Dio può anche esistere, ma non se ne sentirà il bisogno – si afferma. Non v'è bisogno di spendere parole per capire perché transumanesimo e post-umanesimo stiano seriamente preoccupando le religioni. (Cfr. T. Tosolini, *A nostra immagine*, EMI, Padova, 2022) eccetto quella mormonica, che ritiene che le Sacre Scritture invitino l'uomo ad andare oltre se stesso e a trascendere le limitazioni del corpo. (Si vedano i documenti della *Mormon Transhumanist Association*). Il rischio imminente è l'accettazione supina di una sorta di idolatria delle macchine, una escatologia tecnologica.

La conseguenza di ordine pratico della linea di pensiero che si è venuta affermando negli ultimi decenni è che è il capitalismo oligarchico, non più quello democratico, a permettere il progresso socio-economico e la liberazione della società, perché quella democratica è una pratica politica troppo dispendiosa e troppo "wokista". Si leggano documenti quali il "Manifesto del Capitalismo Oligarchico" scritto da P. Thiel nel 2009, in California, e firmato da una potente pattuglia di supericchi quali Vance, Bezos, Musk e altri; il Programma Scientifico del Claremont Institute, uno dei più efficaci "think tank" dell'ultraconservatorismo americano; il farne-

ticante “Manifesto Tecno-Ottimista” di M. Andreessen, co-fondatore di Netscape, dell'ottobre 2023, per rendersi conto di quanto sta accadendo in questo tempo. Il linguaggio della liberazione è divenuto preda da parte di poteri che si ammantano per l'appunto di quel linguaggio. *Corruptio optimi pessima*: siamo di fronte a un sistema che riproduce le parole, i propositi dei liberatori, ma in verità schiaccia le realtà percepite come fragili e vulnerabili in nome di un privilegio di “scorciatoia” in ossequio all'ideologia prestazionale. (Cfr. R. Kurzweil, *The singularity is coming*, 2025, corposo saggio dove l'Autore sostiene che entro la data del 2050 si arriverà alla IA Generale).

Su un punto specifico desidero soffermarmi, pur in breve. Il recente passaggio (fine marzo 2026) a Roma di Peter Thiel, per un ciclo di seminari sull'Anticristo (organizzati dalla Associazione Culturale Vincenzo Gioberti di Brescia) destinato ad un pubblico iperselezionato, ci permette di chiarire un nodo importante del suo pensiero. Per Thiel, l'Anticristo, oggi, è chi sfrutta la paura dell'Apocalisse per imporre una governance globale; chi drammatizza i rischi esistenziali rappresentati dal nucleare, dalle armi biologiche e dalla IA, per rallentare il progresso tecnologico di cui l'umanità ha invece bisogno per sopravvivere. Il demonio di Thiel prende la forma della regolamentazione tecnologica e della lotta al cambiamento climatico, forme che fingerebbero di dare sicurezza, ma che in realtà tolgono la libertà ai cittadini, fino a definire papa Leone XIV un papa woke perché si occupa di pace e di IA. La sostanza dell'argomento è che il *Katèchon* (la forza frenante che ritarda la venuta dell'Anticristo, come scrisse S. Paolo) è rappresentato da Trump, dai suoi collaboratori e da entità come il *Deep State*. D'altro canto, il male sarebbe oggi incarnato dalle ONG che aiutano i migranti e tutti coloro che sono affetti dalla “Anti Trump Derangement Syndrome”, una malattia mentale che fa giudicare negativamente ogni cosa che fa Trump.

Come bene spiega E. Mazzarella (*Critica della ragion digitale*, Castelveccchi, 2026), i motivi della evoluzione tecno-teologica di Thiel – convertito al cattolicesimo da diversi decenni – li si trova nel suo testo del 2007, *Il momento Straussiano*, dove viene esposta la nota tesi di Leo Strauss: l'Impero del bene deve mettere da parte le illusioni delle culture liberal e woke e occuparsi piuttosto della “Teologia della

prosperità” fatta propria dai gruppi evangelici e dall’Alt Right che sostengono un cristianesimo senza la Croce e senza il perdono. Il povero è tale – si sostiene – perché è caduto nel vizio e quindi nel peccato! Ebbene, un punto che merita attenzione è che Thiel – laureato in filosofia e in diritto – dichiara di aver attinto l’ispirazione del suo costrutto da René Girard (1923-2015), il noto filosofo francese che fu suo docente all’Università di California. Ma non è così, come ha convincentemente mostrato Bernard Perret (“I conservatori USA e il pensiero di René Girard”, *Vita e Pensiero*, 6, 2025), che si chiede: come può un pensiero – quello di Girard – che ha tra i suoi meriti quello di mettere in guardia contro la violenza e il suo radicamento nel carattere mimetico delle passioni umane – la propensione ad imitare il desiderio altrui e a fare dell’altro un nemico – trovarsi mescolato ad una concezione politica tutta basata sul culto della forza bruta e sul disprezzo del principio democratico? È questo il nodo gordiano che va tagliato se vogliamo comprendere le radici profonde del progetto transumanista, un progetto filosofico-religioso della libertà senza limiti – cioè del libertarismo che non va confuso con il liberalismo – di cui le Big Tech sarebbero le avanguardie. Libertà anche di “hate speech” e di disinformazione, ma soprattutto di praticare il *MUAI* (*Malicious Use of Artificial Intelligence*), cioè l’uso malevolo dell’IA. Che fare, allora? La risposta di papa Leone è che, se il mondo cristiano intende rimanere fedele alla sua missione, non può limitarsi ad indignarsi e ad avanzare proposte di natura sia politica sia economica, di mera mitigazione e di adattamento alla nuova situazione. Né, per l’altro verso, può chiudersi nel misoneismo, che è l’atteggiamento mentale tipico di chi pensa che non ci sia nulla da fare, perché troppo grosse e complesse le sfide in campo. E soprattutto deve smetterla con la retorica della “pazzia” umana, perché quanto è sotto i nostri occhi non consegue dalla mente deviata di un gruppo di persone, ma da una scuola di pensiero che opera indisturbata da oltre un ventennio, anche per colpa della pigrizia mentale di coloro che avrebbero avuto titolo per intervenire. A ben considerare è questo l’atteggiamento di chi ritiene che il principio di responsabilità si limiti a chiedere di rispondere, di dare conto delle proprie azioni (*respondeo*, in latino). Il cristiano invece sa bene qual è il senso della celebre parabola del buon Samaritano: la piena responsabilità è prendersi cura del peso delle cose (*res pon-*

du), che gravano sull'altro. In sostanza, si è responsabili non tanto per quello che si fa, quanto piuttosto per quello che non si fa, mentre si potrebbe farlo.

Ebbene, è a fronte di un quadro del genere che papa Leone prende con inusitata forza posizione a favore del progetto dell'umanesimo digitale, che rimarca la centralità dell'uomo anche nel presente contesto storico. Come aveva scritto G. Piana (*Umanesimo per l'era digitale*, 2022), tre sono i pilastri di tale progetto. Primo, il recupero della dimensione misterica della persona e quindi l'impossibilità di una sua totale oggettivazione. Secondo, l'accoglimento della visione personalista dell'uomo, come soggetto ontologicamente relazionale, al posto della visione individualista. Terzo, l'apertura ad una prospettiva trascendente che motivi la persona ad andare oltre, ma sempre ricordando che la forma umana dell'esistenza non è un'appendice dello sviluppo tecnico.

2. Linee di azione, urgenti e possibili

Dal momento che non è pensabile – ci dice papa Leone – una Chiesa che si fermasse alle sole *lamentationes* senza effetto e che si limitasse alla sola denuncia dei mali che l'intreccio completo tra naturale e artificiale andrebbe a determinare, la domanda che sorge spontanea è: che fare per dare ali alla prospettiva dell'umanesimo digitale? In altro modo, cosa si può fare affinché l'umanesimo integrale possa essere accolto da quanti, credenti e non, vanno prendendo atto che la creazione di un “nuovo umano”, libero dai vincoli della naturalità, sarebbe qualcosa di aberrante, perché vedrebbe l'umanità come una mera fase di transizione nello sviluppo evolutivo? La dichiarazione finale del vertice sulla sicurezza dell'IA svoltosi a Bletchley (UK) nel gennaio 2025, al quale hanno aderito trenta paesi, l'ONU, l'OCSE e l'UE introduce la distinzione tra *guardrail* e *safeguard*, e aggiunge che non è più sufficiente, allo stadio cui siamo giunti, limitarsi all'algoritmica, che tuttavia resta necessaria. Quel che si richiede è un intervento complementare sull'uomo, a partire da uno specifico progetto educativo, che non si limiti a dettare linee-guida per insegnare l'uso efficace dei nuovi strumenti e a fissare codici di condotta, ma che sappia affrontare la sfida educativa che l'IA pone al sistema scolastico e universitario. È il tema dell'alleanza educativa per l'era digitale di cui tratta la MH.

L'algoritica mira ad integrare principi etici nella progettazione degli algoritmi e nel funzionamento del sistema. In tal senso, essa funge da *guardrail*, come spiega Paolo Benanti (*La nuova logica del dominio*, Roma, Laterza, 2026), ma non raggiunge il foro interno di chi utilizza l'IA, la sua libertà, le sue scelte. Cosa ne è della responsabilità morale soggettiva di chi interagisce con l'IA? Ecco perché è necessario il *safeguard*: un progetto educativo, fondato sulla aristotelica etica delle virtù, che rafforzi nell'utente la capacità di resistere all'antropomorfizzazione della IA. Oggi, i chatbox dei LLM sono espressamente progettati per agire in modo antropomorfo attraverso la conversazione linguistica, come se le IA fossero umane, e quindi coscienti. E la ragione per cui questo avviene è prettamente utilitaristica, per attrarre sempre più clienti e utenti. Il paradosso che ne consegue è che l'umanizzazione della macchina induce nell'uomo la tendenza a percepire l'essere umano, incluso se stesso, come macchina. Un solo esempio: il fenomeno del sicofantismo, cioè la tendenza dei LLM di concordare quasi sempre con l'utilizzatore, che viene adulato per catturarne il pensiero. Con il che la persona che utilizza l'IA per consigli o suggerimenti finisce col diventare incapace di autocorrezione e di autoconsapevolezza. Si consideri che l'uso di chatbot a fini relazionali è, oggi, secondo solo alla ricerca di informazioni. Il potere di convincimento che viene esercitato è tale da giustificare l'appellativo di "tecnologie della persuasione".

È in vista di ciò che occorre tornare a *pensare* la Tecnica e non solo continuare ad occuparsi della Tecnologia e dei suoi successi. Uno dei guasti più seri che l'egemonia culturale contemporanea ha creato nelle nostre società è proprio questo: che non vi sarebbe bisogno di impegnarsi con il pensiero pensante, essendo più che sufficiente il pensiero calcolante per ottenere risultati utili, come se fosse possibile agire senza pensare. (agire è cosa ben diversa da fare). Ecco perché è urgente introdurre nella dichiarazione universale dei diritti dell'uomo il diritto all'*Habeas mentem*, - analogo al ben noto *Habeas corpus*, introdotto nell'Inghilterra del XIII secolo - un diritto volto a tutelare la libertà di pensiero della persona dalle manipolazioni che il "capitalismo delle menti" va attuando. Prendersi cura della mente è oggi un compito irrinunciabile per chi ha a cuore la libertà in tutte e tre le sue dimensioni: libertà *da*, *di*, *per*. Non può essere la Tecnica a fissare l'agenda, a meno

di accogliere supinamente quella nuova forma di potere che Paolo Benanti chiama “potere computazionale”. Quale la sua cifra? Quella di recidere il nesso tra competenza tecnica e comprensione delle sue conseguenze. Serve quindi un’etica capace di confrontarsi con l’incertezza, di affrontare situazioni in cui le conseguenze non sempre sono prevedibili. Ebbene, è da questa esigenza che nasce l’etica della responsabilità: offrire un quadro di riferimento per valutare azioni e omissioni non solo in base all’intenzione e alla convinzione (Max Weber) ma anche in base agli effetti che esse possono generare in ambiti complessi e interconnessi.

Nel suo *Il principio responsabilità* (1979), Hans Jonas opportunamente ha parlato di un principio di precauzione: “in dubio, pro malo”, ovvero di fronte all’incertezza sugli effetti di una nuova potente tecnologia, occorre dare più peso alla “profezia di sventura” che “alla profezia di beatitudine”. Il punto è rilevante, perché vale a scongiurare il rischio dell’affermazione egemonica della responsabilità adiaforica, cioè della “responsabilità tecnica”, che non prevede il giudizio morale di ciò che è bene e male. L’agire adiaforico, tipico delle grandi organizzazioni, andrebbe valutato, secondo questa impostazione, in termini solamente funzionali, in forza del principio che tutto ciò che è possibile è anche eticamente lecito.

Una lacuna di un certo peso che deve essere colmata, nel *milieu* culturale cristiano, è quella di ripensare la vicenda della Ascensione al Cielo di Gesù. Si tende infatti a pensare in taluni ambienti, che con la sua Ascensione, andando a sedersi sul trono del Padre, Gesù abbia chiuso la storia, dopo aver compiuto la sua missione: la redenzione del genere umano. Ma questa lettura dell’Ascensione pone un problema teologico: se la storia che ci separa da quell’avvenimento non ha più interesse, se tutto si è già compiuto, perché inviarci lo Spirito Santo? Si legge in Luca, all’inizio degli Atti, laddove gli apostoli domandano al Cristo se ora ristabilirà il Regno di Davide, la risposta che ricevono è: “Non vi spetta conoscere l’ora fissata dal Padre nella sua autorità, ma vi sarà data una forza”, lo Spirito Santo appunto, e detto ciò scompare in cielo. Gesù lascia dunque vuoto il trono di Davide, per affidare agli uomini la missione di riempirlo. (G. Giraud, *Costruire un mondo comune*, Milano, Piemme, 2025).

Ce lo rammenta anche D. Bonhoeffer, quando nel suo *Resistenza e Resa* scrive: “La

fedele cristiana si distingue da tutte le altre concezioni del mondo perché non conduce l'uomo fuori dal mondo, ma lo rimanda proprio dentro il mondo... Il Cristiano non è dispensato dai compiti terreni, ma proprio in essi è confermato". Nel suo discorso al corpo diplomatico Vaticano del 9 gennaio 2026, papa Leone XIV, rinnovando la visione agostiniana delle due città, lungi dall'opporre eternità e tempo, Chiesa e Stato, insiste, assai opportunamente, che "i cristiani sono chiamati da Dio a dimorare nella città terrena con il cuore e la mente rivolti alla città celeste, la loro vera patria. Allo stesso tempo, i cristiani che vivono nella città terrena non sono estranei al mondo politico e, guidati dalle Scritture, cercano di applicare l'etica cristiana al governo civile... Agostino mette anche in guardia dai gravi pericoli per la vita politica derivanti da *false rappresentazioni della storia, da un nazionalismo eccessivo e dalla distorsione dell'ideale del leader politico*". (Corsivo aggiunto). Si confronti la posizione di papa Leone con la proposta di un "governo tecnologico" avanzata da A. Karp e M. Zamiska, *The Technological Republic*, 2025. Secondo gli Autori, la tecnologia deve incorporare una sorta di nuovo "umanesimo", per disegnare il futuro dell'Occidente, basato su una specifica "teologia scientifica".

3. Post-verità e nuovo progetto educativo

Una seconda rilevante linea di azione, suggerita dalla rotta tracciata nella MH, chiama in causa il grosso tema della verità. *Postruth* e *distrust* seguono gli orizzonti ingegneristici della nostra civiltà digitale. Di fronte all'emergenza di queste vulnerabilità siamo chiamati ad un grande sforzo. Invero, il collasso dei regimi di verità e di falsità è un fenomeno tipico di ogni significativa innovazione tecnologica. Oggi, però, l'IA e altro hanno prodotto una condizione umana confusa e diffidente, che pone a rischio anche le democrazie nelle loro capacità istituzionali. Saranno le macchine a discriminare il vero dal falso e non più il nostro *sensorium*, il nostro apparato di sensi, sentimento e pensieri? E poi in che misura possiamo fidarci di queste macchine? (Cfr. A. Stella, G. Ianulardo, *Truth and Relations*, Londra, Palgrave MacMillan, 2026, per una pregevole esposizione critica delle varie concezioni paradigmatiche di verità).

Dai server di posta elettronica che separano le email utili dallo spam a quelli ban-

cari che distinguono gli accessi legittimi ai nostri risparmi da quelli fraudolenti, alle diagnosi di IA medica, o ancora all'IA militare che profila i potenziali bersagli colpendo il nemico e risparmiando il non nemico, o alle auto a guida autonoma. Nonostante i successi, talvolta più conclamati che reali, va riconosciuto che sono ancora delle macchine "opache", caratterizzate da "opacità epistemica". Funzionano certamente, ma non sappiamo perché e come funzionano. (Non a caso si parla di "black box"). Non si dimentichi, infatti, che quella digitale è una tecnologia ancora pre-scientifica. È tale una tecnologia che funziona, ma di cui non si possiede la teoria che spieghi perché funziona. Un precedente famoso è quello della caldaia a vapore, che entrò in operazione prima che la teoria termodinamica ne dimostrasse il funzionamento. Il principio etico della *explainability* esigerebbe invece, di ricostruire i percorsi e i vincoli con cui gli algoritmi calcolano la loro probabilità innescando inferenze, decisioni e azioni. La verità non si dà, infatti, senza la sua rete di senso fatta di osservabilità, causalità, responsabilità. Invece quel che avviene è "l'adattamento statistico" di cui parla papa Leone.

Un aspetto, troppo spesso trascurato, tipico delle intelligenze artificiali è che queste non sono capaci di affrontare problemi di causalità. L'intelligenza artificiale è capace soltanto di studiare correlazioni, per quanto complesse. Quella della macchina, infatti, è un'intelligenza sintattica, che mai potrà sostituire l'intelligenza semantica. Ne deriva che l'intelligenza artificiale non distingue il vero dal falso, dato che è in grado di occuparsi solamente del verosimile. Da un punto di vista epistemologico ciò solleva un problema serio: qual è il valore della conoscenza generata dall'intelligenza artificiale? È ovvio infatti che se rinunciamo a ricercare nessi di causalità tra i fenomeni che ci intrigano, mai si riuscirà ad ottenere certezza su quanto accade. Il fenomeno delle *fake truths* – fenomeno molto più serio di quello delle *fake news* - va visto in quest'ottica. Se lancio sui social una falsa verità, che viene accolta da molti soggetti come vera e se, trascorso un certo lasso di tempo nessuno la contesta, quella *fake truth* diviene verità a tutti gli effetti. (Questa, come sappiamo, è una delle tesi centrali del pragmatismo, posizione filosofica nata negli USA a fine Ottocento).

Accade infatti che i messaggi brevi e fulminei dei social sono tanto più efficaci

quanto più disconnessi da qualsiasi riflessione e sono tanto più convincenti quanto più sono radicali. Gli algoritmi fanno prevalere le posizioni più semplici e dunque drastiche, che accentuano il divario tra “noi” e “loro”. Non costruiscono opinioni ma catturano e forgianno identità, esasperando una inclinazione ben nota agli psicologi, secondo cui prestiamo ascolto a ciò che già ci è familiare. Non solo, ma riconosciamo e aderiamo alle idee che già abbiamo, cioè viviamo in una “eco-chamber”. In tale contesto, il discorso scientifico diviene refrattario ai dati di fatto e pure alle verifiche. Come ha mostrato il noto giurista americano Cass S. Sustein, una falsa verità prevale sempre sulla sua smentita. Il fatto è che, oggi, c'è scollamento tra messaggio e significato, tra demagogia e dati di fatto. Nati per diffondere prodotti e generare profitti, gli algoritmi, offrendo gratuitamente i loro servizi, vanno facendo degli stessi utenti un prodotto che genera profitti. Di populismi distruttivi ce ne sono sempre stati nella storia; le società li patiscono e li superano. Come? Aggrappandosi alla verità. Oggi, però, questo vecchio meccanismo di difesa sta venendo meno. La post-verità – che non è l'accreditare le bugie come verità, quanto piuttosto l'intorpidire le acque al punto tale che diventa praticamente impossibile distinguere il vero dal falso – minaccia gli anticorpi che usano la democrazia per curarsi dalla malattia dei populismi e resistere al continuismo.

Gli esseri umani sono chiamati a confrontarsi con nuovi tipi di agenti: quelli artificiali. Ciò comporta che dobbiamo modificare i nostri comportamenti, quando interagiamo con tali dispositivi. Quindi non è più in gioco la subordinazione della persona alla macchina e all'organizzazione del lavoro che essa richiede. Ciò è avvenuto nel passato con le prime due rivoluzioni industriali. Quanto oggi è in gioco è che la persona è considerata sostituibile in ulteriori due mansioni rispetto a quelle già rimpiazzate nel passato. Per un verso, la creatività. Oggi, la macchina può prendere *decisioni* anche importanti e non solamente operare delle scelte, come in passato. Per l'altro verso, la subordinazione non è più alla catena di montaggio, ma all'algoritmo, così che l'attività umana viene ridotta a semplice procedura, rigidamente determinata, invece che promuovere la costante interazione con contesti sempre nuovi. Di qui la dequalificazione del lavoro e il grave rischio dell'appiattimento su un modo di pensare riduzionista.

Ciò solleva la questione di cosa accade ai diritti umani quando una parte crescente della vita sociale, politica, economica si sposta nello spazio digitale. Le nuove macchine non sono più strumenti *esterni* alla vita umana, ma infrastrutture entro le quali si formano opinioni, si esercitano libertà, si producono diseguaglianze e si prendono decisioni. Ne deriva che il problema non è soltanto proteggere online gli stessi diritti che già esistono offline, ma comprendere se il digitale modifichi il significato e le condizioni di esercizio di quei diritti. Quali obblighi ricadono sugli Stati e quali sulle piattaforme private? Papa Leone dedica grande attenzione al rischio educativo dell'IA: la velocità delle risposte, la facilità delle sintesi, la disponibilità in tempo reale di informazioni possono spegnere il desiderio della ricerca personale, del dubbio, dell'attesa. Ogni tecnologia educa chi la usa in modo non neutrale; ogni rivoluzione tecnologica promette un potenziamento umano, ma ogni potenziamento comporta anche una perdita.

La recente decisione della Gran Bretagna di vietare, per via legale, l'accesso ai social media ai minori di sedici anni ha riaperto il confronto, anche aspro, tra due posizioni opposte. Da una parte, vi sono coloro che, analizzando gli indicatori di salute (fisica e mentale) in età evolutiva, si accorgono del loro peggioramento negli ultimi vent'anni e come questo sia correlato all'età di ingresso e all'intensità di uso del mondo digitale. All'ansia pervasiva che tiene in ostaggio tanti adolescenti, si aggiungono nuovi fenomeni mai visti prima in età evolutiva, primo fra tutti il ritiro sociale. Dall'altra parte, non pochi sono coloro che ritengono che sia meglio preferire l'educazione al divieto, vale a dire il "safeguard" va privilegiato rispetto al "guardrail".

Come spesso accade, la verità sta nel mezzo – ricordava Aristotele. Infatti, quando si parla di educare al mondo online, non si tiene conto che si ha a che fare con un ambiente grandemente additivo, architettato ad arte per diventare una calamita il cui campo magnetico intrappola il cervello dei minori, già di per sé più vulnerabile. D'altro canto, va riconosciuto che per decenni abbiamo permesso al mondo digitale di qualificarsi come zona franca da qualsiasi regola presente nel mondo reale a tutela di diritti dei minori. È giunto il tempo di pensare e progettare come andare oltre le due opposte posizioni. Ampia è ormai l'evidenza empirica a proposito del-

la correlazione tra uso dei social media e usura degli indicatori di salute mentale. La questione ha assunto le proporzioni di una vera e propria emergenza di sanità pubblica. Ciò spiega, fra l'altro, la recente decisione della Commissione Europea di procedere, entro l'autunno del 2026, ad un rafforzamento del "Digital Services Act", eliminando le funzionalità dannose, del tipo: algoritmi che creano dipendenza, contenuti nocivi, contatti indesiderati, e altro ancora. Si pensi al caso degli *AI Companion*, che generano ad arte aspettative irrealizzabili e tendono a creare *captive minds*.

In un saggio di notevole spessore (*Captive Minds. A Study of Manipulation*, Harvard University Press, 2026) gli Autori, A. Margalit e A. Sharon, affrontano una questione che concerne in profondità la vita della nostra società, ossia il modo in cui le persone possono essere condotte a pensare, desiderare, giudicare e agire secondo una volontà che non è propriamente la loro. La manipolazione è fenomeno antico, ma oggi essa assume una portata nuova. La sua rilevanza non è tanto che qualcuno possa mentire o ingannare qualcun altro, quanto piuttosto che una mente possa essere catturata senza che il corpo venga costretto. La questione è allora comprendere quando un'influenza diventa dominio, quando la persuasione diventa controllo, quando una società può produrre condizioni che rendono i cittadini vulnerabili alla manipolazione sistematica. Può dirsi davvero libera una società se le opinioni su cui fonda la propria legittimità sono sistematicamente fabbricate o catturate da altri?

È notevole l'impatto delle considerazioni che precedono sul processo di apprendimento. Due scuole di pensiero circa l'utilizzo degli strumenti dell'IA in campo educativo si confrontano. Per un verso, vi sono coloro che considerano tale utilizzo efficace e quindi auspicabile. Per l'altro verso, c'è chi considera questi strumenti causa del cosiddetto "cognitive offloading", che è la possibilità di ottenere informazioni, soluzioni, dimostrazioni tramite l'IA, riducendo così lo sforzo cognitivo e favorendo una nuova forma di pigrizia mentale. C'è, però una "terza via", quella che cerca di integrare le due posizioni per mezzo della sequenzialità dell'esperienza di apprendimento, facendo leva sull'attrito cognitivo: allenare la mente con lo sforzo richiesto dalla lettura e dalla ricerca personale. Una conferma pratica di

questo approccio ci viene da recenti esperimenti realizzati dall'Università del Michigan e dalla Stanford Graduate School of Business, che utilizzano la tecnica della "submission" a tre stadi: L'IA viene impiegata solo nella terza fase dei corsi, mentre le prime due fasi adottano metodi tradizionali. L'esito è che gli studenti sviluppino una maggiore sensibilità ai *bias* ed una marcata capacità di giudizio critico. Vale a dire, l'IA esprime il meglio del proprio potenziale quando è utilizzata da chi ha già allenato la propria mente.

Il messaggio del Papa, introducendo nel dibattito la figura dei *virtual influencer*, ci invita a riflettere sui pericoli che andranno a materializzarsi se non si interviene prontamente. Si tratta di figure, create dall'IA e adottate dai brand per finalità di marketing in relazione alla scalabilità e all'adattabilità delle "persone digitali", che sono in grado di muovere le scelte di acquisto delle persone reali. Di qui la sfida della *Media Education*. Occorre pensare alla necessità di riconoscere gli stimoli emotivi nelle relazioni con l'IA, di difendere l'intimità, di valorizzare autentiche esperienze relazionali. E questo perché – ci dice Leone XIV – la Chiesa non è solo "depositum fidei", ma pure "depositum humanitatis", e dunque è interessata ad indagare le trasformazioni del potere nella odierna società.

4. La questione sociale nell'era dell'IA

Rivolgo ora l'attenzione ad alcuni aspetti della realtà socio-economica "rivoluzionata" dalle nuove tecnologie. L'accelerato sviluppo in ambito tecno-scientifico dei sistemi di Intelligenza Artificiale va sollevando un insieme di inquietanti interrogativi circa le implicazioni di tali sviluppi sul fronte economico, e sociale. Ciò è particolarmente il caso quando ci si riferisce all'IA generativa e soprattutto alla più recente IA agentica, una nuova branca dell'IA capace di generare nuovi contenuti multimediali (testo, video, audio) in risposta a suggerimenti umani (i cosiddetti *prompts*) sulla base di addestramenti *ad hoc* estratti da miliardi di dati (*big data*). Tali sviluppi stanno determinando cambiamenti profondi nella sfera giuridica, socio-economica, culturale dei nostri paesi, impensabili fino a qualche anno fa. Poiché nella definizione di intelligenza è implicato il concetto di "coscienza" e quindi di "umano", il fatto che le tecnologie in evoluzione possano avvicinarsi a dare

la percezione di una “coscienza” a macchine basate su algoritmi – la cosiddetta coscienza artificiale – è un tema piuttosto inquietante. Al di là di questioni etiche, come la definizione stessa di cosa riteniamo essere la “vita”, la possibilità di un nuovo materialismo che intenda far credere che l’essenza della vita umana è solo una serie di algoritmi sembra un’eventualità non troppo remota che porrebbe in discussione la centralità dell’uomo, con quel che ne conseguirebbe. Di qui l’urgenza di giungere a definire una specifica algoretica, come papa Francesco e Leone XIV hanno sottolineato con forza.

Uno dei nodi importanti riguarda la possibilità o meno di riconoscere nuovi diritti umani digitali. Il problema è stabilire se le tecnologie digitali richiedono diritti complessivamente nuovi, come il diritto all’autodeterminazione digitale, oppure se sia sufficiente estendere e reinterpretare i diritti già riconosciuti, come privacy, libertà di espressione, non discriminazione e accesso all’informazione. Un passaggio importante è rappresentato dall’idea, affermata in sede ONU, secondo cui gli stessi diritti che le persone hanno *offline* devono essere protetti anche *online*. Tuttavia, questa affermazione, pur essenziale, non esaurisce la questione. Alcune esigenze di protezione sembrano infatti assumere nel digitale una forma specifica: il diritto di accesso a internet, il diritto all’oblio, il diritto a sapere se si sta interagendo o meno con un sistema di intelligenza artificiale, il diritto a non essere sottoposti a forme manipolative o discriminatorie di profilazione automatizzata. Il punto centrale è che l’intelligenza artificiale accentua la tensione tra innovazione e tutela dei diritti: può migliorare servizi, decisioni e conoscenza, ma può anche produrre sorveglianza di massa, discriminazioni, manipolazione, opacità e perdita di controllo umano.

Ecco perché le sole linee guida etiche o forme di autoregolazione privata non bastano quando le tecnologie incidono su diritti fondamentali e processi democratici. Ciò in quanto non è corretto presumere che un diritto umano mantenga lo stesso significato quando viene trasferito dal mondo fisico a quello digitale. La formula secondo cui gli stessi diritti esistenti *offline* devono essere protetti *online* resta politicamente e giuridicamente importante, ma rischia di nascondere una trasformazione profonda. Il punto non è negare che i diritti valgono anche *online*; è ricono-

scere che, nel passaggio al digitale, il loro contenuto può cambiare. La privacy è l'esempio più chiaro. Nel mondo fisico, i tribunali dispongono di tempo per valutare il conflitto tra privacy, reputazione, libertà di espressione e accesso all'informazione. *Online*, invece, la rapidità della circolazione dei contenuti rende spesso tardiva qualsiasi decisione giudiziaria. La regolazione è affidata in molti casi a piattaforme private, con deficit di trasparenza e legittimità. La teoria della non-coerenza dei diritti umani digitali non siano semplici copie dei diritti tradizionali. Essi esistono in un ambiente nel quale tempo, spazio, potere, controllo e rimedio giuridico funzionano diversamente. Ne deriva una visione più relativizzata dei diritti umani, nella quale anche valori centrali come la dignità possono essere sottoposti a bilanciamenti nuovi e meno stabili.

La grande intuizione dell'enciclica MH è che il problema del lavoro umano non consiste tanto nella sostituzione materiale dell'uomo, quanto piuttosto nella sua atrofizzazione spirituale. È un fatto che quando gli economisti discutono del futuro del lavoro, tendono a domandarsi: quale sarà il vantaggio comparato degli umani? Non è questa la domanda appropriata, cioè come gli umani possono gareggiare contro le macchine; piuttosto occorre chiedersi quali sono le forze degli umani e come queste possono essere impiegate per fare cose che le macchine mai potranno fare. Sostituire gli esseri umani con i robot dotati di IA nella speranza di ridurre il costo del lavoro: è questa la tentazione ricorrente di una pluralità di imprese. Questa pratica ha già un nome: "AI washing", cioè l'uso dell'IA come pretesto per giustificare ristrutturazioni, mascherando altri problemi finanziari. Ma studi recenti smentiscono il mito di una "manodopera" quasi gratuita. La ragione è l'aumento vertiginoso dei costi legati all'IA, che tendono a superare i costi del lavoro.

Il 19 maggio 2026, Sundar Pichai, CEO di Google, in un incontro con i suoi sviluppatori rivelò che parecchie aziende nel mondo avevano già esaurito il loro budget annuale di token. (Un token, un gettone, è l'unità fondamentale di informazione utilizzata da un modello linguistico, come Chat GPT e altri, per generare un testo). Il token è diventato l'unità di fatturazione dell'IA, che misura la quantità di testo elaborato dal modello e serve a stimare le risorse di calcolo necessarie e i costi relativi. Soprattutto l'IA agantica è divoratrice di token, come il noto economista di

Stanford Erik Brynjolfsson ha chiarito di recente. Secondo lo studioso, il budget dedicato dalle aziende ai token dovrebbe raddoppiare ogni anno, superando gli incrementi di produttività attesi. Il fatto è che i nuovi agenti girano a vuoto, perchè non dispongono di un meccanismo affidabile per identificare che un certo compito è irrisolvibile oppure è già stato risolto. Ecco perché il fenomeno in rapida diffusione dell'*IA washing*, va preso in seria considerazione. La tendenza delle imprese ad esagerare il ruolo dell'IA per apparire più innovative di quanto siano davvero. Non solo, ma si ricorre ad una narrazione dell'IA che non corrisponde alla realtà per giustificare decisioni aziendali controverse, quali licenziamenti di lavoratori, così da renderli più accettabili perfino agli occhi degli stessi. L'esito è una duplice distorsione. Per un verso, si alimenta l'illusione di un'automazione già in atto; per l'altro verso, si celano le reali difficoltà economiche dovute a incapacità organizzativa. È questo un problema che, prima o poi, dovrà essere affrontato in sede di contrattazione sindacale.

L'Università di Berkeley ha mostrato, per via empirica, che l'aumento dell'efficienza non libera tempo, lo riempie di nuovi incarichi, fino al burnout. (Il Rapporto è pubblicato sulla *Harvard Business Review*, Feb. 2026). La scomparsa dei tempi morti è l'effetto più serio delle nuove tecnologie sul fronte della salute mentale dei lavoratori. Poiché l'IA elimina "l'attrito del foglio bianco", i lavoratori tendono a usare pause di varia natura per inviare un ultimo comando, limitando i naturali tempi di recupero cognitivo. Si passa così da un compito all'altro, senza pause per effetto della presenza di un partner digitale sempre al proprio fianco. L'espansione del lavoro è spesso volontaria e inizialmente gratificante. L'IA rende facile fare di più. Ma la sua velocità di sviluppo non è affiancata da capacità cognitive adeguate da parte dei lavoratori che tendono a non voler smettere più di lavorare, in seguito all'adozione da parte dell'impresa delle tecniche di "gamification". L'esito è la riduzione di fatto del tempo libero, proprio perché l'IA ci rende più produttivi!

La trasformazione dell'infrastruttura digitale globale ha raggiunto, da qualche anno, una biforcazione fondamentale; la rete non è più progettata solo per il consumo umano, ma pure per l'interazione diretta tra sistemi intelligenti attraverso protocolli di comunicazione. Ciò è dovuto all'ascesa dell'IA agentica che ha dato

l'avvio alla cosiddetta Internet delle Macchine, dove l'utente non è più la persona dietro uno schermo, ma l'agente intelligente che opera per suo conto. È agevole afferrare le conseguenze per la produttività aziendale: la capacità di un'impresa di rendersi leggibile e interrogabile dalle macchine diventerà uno dei principali fattori di successo. Occorre allora interpretare correttamente questa transizione per comprendere come le gerarchie economiche verranno riscritte dalla capacità di coordinamento autonomo tra sistemi capaci di negoziare tra loro. Moltbook è un social network sperimentale lanciato il 31 gennaio 2026 da Matt Schicht, che rappresenta la prima piattaforma dedicata agli agenti di IA e non ai lavoratori, i quali sono ammessi solo come osservatori. Le IA conversano tra loro e si scambiano informazioni, estromettendo totalmente il lavoro dalla "vita" delle macchine.

La questione di centrale rilevanza non è tanto se l'IA aumenti la produttività, - il che è certamente vero - quanto piuttosto perché imprese, investitori, istituzioni pubbliche la considerino così importante già oggi, anche se i suoi effetti complessivi non sono ancora pienamente visibili nelle statistiche dell'economia. La risposta è che l'IA modifica il modo in cui il lavoro viene organizzato. Non agisce solo sulla velocità con cui si produce, ma sul rapporto tra persone, strumenti e risultati. È qui che si apre il tema decisivo del futuro del lavoro: non tanto la sua scomparsa, ma la trasformazione del ruolo dei lavoratori dentro organizzazioni sempre più dipendenti dall'automazione. Si badi che la trasformazione non arriva solo quando si perde il posto di lavoro, ma mentre già si lavora: il mio lavoro sta cambiando anche se non sto cambiando lavoro!

Il punto sollevato merita un'attenzione particolare per come l'IA sta trasformando il mercato del lavoro. Come si trae dal recente Rapporto di PwC "Global AI Jobs Barometer 2026", frutto di una rigorosa ricerca empirica di vastissimo raggio, l'IA non muta e/o elimina solamente i tradizionali compiti lavorativi, ma pure i gradini della scala gerarchica delle imprese. PwC ha coniato il neologismo "senioritazation", per denotare la maturazione forzata delle competenze degli occupati. Specificamente, rispetto alla tradizionale scala (junior, middle, senior, manager) va declinando il primo gradino, quello riferito al lavoro junior, che ora viene svolto dall'algoritmo. Con quale esito? Che al neoassunto fresco di laurea (o di altro titolo) non

si chiede più di eseguire “gli ordini” ricevuti, ma di prendere decisioni nel problem solving o nel project management. Come dire che il dipendente junior deve comportarsi da senior, pur conservando la remunerazione da junior. Come si può capire, ciò accende un nuovo tipo di conflittualità dentro l’impresa. Da qualche tempo è in rapida crescita il numero di imprese che vanno istituendo la figura del Chief AI Officer, proprio per cercare di risolvere un tale conflitto.

Un’annotazione finale. Il 29 gennaio 2026, la *Harvard Business Review* ha pubblicato un’inchiesta a cura di T. Davenport e L. Srinivasan, tra i più quotati esperti dell’impatto delle IA sulle organizzazioni, condotta su oltre mille dirigenti globali a dicembre 2025. L’indagine dà risposta empirica ad una delle domande più dibattute negli ultimi tempi: l’IA sta davvero sostituendo i lavoratori o le aziende stanno solo usando l’IA come “pretesto” per ridurre il personale? La risposta che viene dai dati rovescia buona parte della narrazione pubblica corrente. Le aziende stanno bensì riducendo il personale, ma quasi sempre in anticipazione di un impatto futuro, non come conseguenza di un’efficienza già registrata. È agevole comprendere le serie conseguenze sulla idea stessa di lavoro umano di un trend del genere. Quanto si sta realizzando deve spingerci a prestare la dovuta attenzione alla seguente *res nova*. È ormai noto che la vera risorsa all’origine della creazione di valore è rappresentata oggi dalla vita registrata delle persone, le quali forniscono, spesso senza rendersene conto, i dati che servono a far funzionare le piattaforme. Ciò spiega, fra l’altro, perché negli ultimi tempi USA e Cina hanno negoziato con alcuni paesi africani accordi in base ai quali dati biologici e sanitari di quelle popolazioni vengono “acquistati” in violazione del principio della sovranità dei dati. La grande sfida che ne deriva è come realizzare la redistribuzione tra tutti del valore ricavato dalla vita degli esseri umani, secondo criteri di equità. È in tale prospettiva che si colloca il progetto, denominato *Webfare*, lanciato da Università e Politecnico di Torino e diretto da Maurizio Ferraris, per fornire una risposta a tale sfida. Tra le proposte suggerite, mi piace sottolineare quella di dare vita a “cooperative di dati”, in cui una pluralità di enti (associazioni, Università, strutture sanitarie e altri) conferiscono le proprie tracce digitali a cooperative, appositamente costituite, che le gestiscono in modo trasparente e per distribuire i benefici alle comunità

di appartenenza, sotto forma di previdenza, assistenza, ricerca, servizi vari e così via. È questo un modo concreto per contribuire alla edificazione di una *major humanitas*.

5. Capitalismo cibernetico e principio democratico

Una linea di azione che la MH ci propone con forza è quella che concerne il senso e il futuro del principio democratico. La conseguenza di ordine pratico della linea di pensiero che si è venuta affermando negli ultimi decenni è che è il capitalismo oligarchico, non più quello democratico, a permettere il progresso socio-economico e la liberazione della società, perché quella democratica sarebbe una pratica politica troppo dispendiosa e troppo “wokista”. Si leggano documenti quali il “Manifesto del Capitalismo Oligarchico” scritto da P. Theil nel 2009, in California, e firmato da una potente pattuglia di supericchi quali Vance, Bezos, Musk e altri; il Programma Scientifico del Claremont Institute, uno dei più efficaci “think tank” dell’ultraconservatorismo americano; il farneticante “Manifesto Tecno-Ottimista” di M. Andreessen, co-fondatore di Netscape, dell’ottobre 2023, per rendersi conto di quanto sta accadendo in questo tempo. Il linguaggio della liberazione è divenuto preda da parte di poteri che si ammantano per l’appunto di quel linguaggio. *Corruptio optimi pessima*: siamo di fronte a un sistema che riproduce le parole, i propositi dei liberatori, ma in verità schiaccia le realtà percepite come fragili e vulnerabili in nome di un privilegio di “scorciatoia” in ossequio all’ideologia prestazionale.

Oggi le grandi imprese hanno difficoltà ad adottare sistemi di IA agentica non per ragioni tecniche, ma per una ragione concettuale che affonda le radici in oltre un secolo di storia del lavoro: le imprese (ma non tutte) continuano a chiedere alle macchine di fare ciò che già fanno gli esseri umani, senza chiedersi se quei processi abbiano ancora senso; in particolare se ha ancora senso continuare con il modello organizzativo taylorista. Si autorizza così l’esistente, invece di ripensare chi decide come si lavora e chi ne paga il costo. In altri termini, la più parte delle aziende sta chiedendo agli agenti di fare il lavoro così come questo è stato progettato per essere svolto da esseri umani. Il risultato è una inefficienza amplificata, e si ca-

pisce perché. I processi umani sono costruiti attorno a “compressioni cognitive” che le persone fanno inconsciamente – interpretare le ambiguità; navigare le eccezioni; usare il giudizio contestuale. Un agente di IA che eredita un processo umano senza quelle compensazioni è come un operaio cui si chiede di lavorare su una catena di montaggio non progettata per lui. In sostanza, il valore reale non viene tanto dall’automazione dei compiti, ma dal ripensamento di cosa debba essere fatto. Questa distinzione – tra fare meglio certe cose e chiedersi se quelle cose vadano fatte – è proprio la distinzione che il taylorismo ha rimosso: il manager taylorista ottimizza, non mette in discussione ciò che va fatto.

Il dibattito pubblico sugli agenti di IA vede contrapporsi due rappresentazioni, entrambe coerenti. Da una parte, c’è l’idea che un agente sia poco più di un assistente evoluto, capace di eseguire compiti ripetitivi con maggiore velocità. Dall’altra, c’è l’immagine di sistemi quasi indipendenti, pronti a sostituire intere porzioni di lavoro umano. Il 18 feb. 2026, Anthropic ha pubblicato il Rapporto “Measuring AI agent autonomy in practice”, che evita entrambe le semplificazioni. La domanda di partenza è: quanta autonomia viene effettivamente concessa a questi sistemi, in quali ambiti vengono usati, quanto è rischioso ciò che fanno, come muta il comportamento degli utenti con l’esperienza. Il valore del documento sta qui: invece di discutere in astratto di “agenti”, il Rapporto prova a misurare il modo concreto in cui essi vengono inseriti nei processi lavorativi. Donde la conclusione: l’autonomia degli agenti di IA non è una proprietà fissa delle tecnologie, ma l’esito della capacità (e della volontà) dell’impresa di accettare che l’IA deve lavorare per noi, non pensare per noi. In buona sostanza, la questione non è quello che l’IA farà per noi, ma quel che vogliamo faccia per noi. (D.Wu, S. Liang, *Human-AI interaction and collaboration*, CUP 2026). In quali condizioni le macchine possono interagire a collaborare con gli esseri umani in modo efficace e socialmente accettabile. La domanda centrale: può l’IA diventare un supporto effettivo all’intelligenza umana senza sostituire il giudizio e la responsabilità.

Sorge spontanea la domanda: in quale senso la regolazione della IA, intesa in senso tradizionale, non è sufficiente? La ragione è che una realtà in rapido mutamento non è catturabile da norme per l’approvazione delle quali occorrono tempi me-

dio-lunghi. Sono dunque necessarie norme flessibili e prontamente adattabili alle evoluzioni in corso e soprattutto è urgente dare vita ad un organismo indipendente in grado di assicurare il rispetto delle regole. Solo così si può scongiurare il rischio del “dominio tecnopolare”. L'*Artificial Intelligence Act* (AIA) approvato in sede UE all'inizio del 2024 recepisce tale indicazione di metodo, affidando alla politica il compito di estrarre dai valori guida da tutti condivisi (privacy, dignità umana, responsabilità, trasparenza, assenza di discriminazioni, controllo umano significativo), le regole specifiche da far rispettare, regole elaborate sulla base del principio del male che si vuole evitare. Come noto, l'uomo riconosce con più facilità ciò che lo ferisce e lo umilia, mentre è più in difficoltà quando deve decidere su ciò che lo realizza. È in ciò l'essenza dell'approccio della gestione del rischio, un approccio che mira a scongiurare i mali emergenti.

L'AIA diventerà pienamente applicabile dal 2 agosto 2026, ma nel luglio 2026, la Commissione Europea ha proposto di rinviare fino al 2 dicembre 2026 il termine per l'applicazione di questa disciplina, soprattutto per i sistemi ad alto rischio (come biometria, riconoscimento delle emozioni, valutazione dei richiedenti asilo, valutazioni nel settore giuridico), allo scopo di alleggerire, nel breve termine, l'onere di *compliance* per imprese e pubbliche amministrazioni. Inoltre, la semplificazione normativa che il dispositivo *Digital Omnibus* attua è conseguenza dell'idea, di origine statunitense, secondo cui una regolamentazione troppo severa rallenterebbe l'innovazione e diminuirebbe la competitività del comparto. La sfida allora è dimostrare che si può innovare pur all'interno di un quadro normativo esigente, frutto di deliberazione pubblica. (Il codice della strada ha forse bloccato l'industria dell'automobile?) Per due decenni, il discorso sulla regolamentazione europea è stato dominato da una narrazione secondo cui la regolamentazione comporterebbe costi elevati, sottraendo risorse alla ricerca. Ma si tratta di un argomento privo di fondamento, come la teoria economica seria ha dimostrato. Non solo, ma la sequenza di operazioni che ha coinvolto Mistral AI, tra settembre 2025 e maggio 2026, dimostra che quella narrazione è vera solo in minima parte.

La fase recente dello sviluppo dell'intelligenza artificiale mostra un passaggio concettuale che può rivelarsi più profondo di quanto appaia a prima vista. Dopo anni

nei quali l'AI è stata identificata soprattutto con testi, immagini, codici, previsioni, raccomandazioni e interazioni linguistiche, una parte crescente della ricerca e dell'industria si sta spostando verso sistemi capaci di agire nel mondo fisico. La genealogia è stata fissata da Jensen Huang in un suo recente intervento (Maggio 2026): è cominciata con la *perception AI* – comprendere immagini, parole, suoni. Poi è venuta la *generative AI* – creare testo, immagini e suoni. Adesso entriamo nell'era della *physical AI*, un'intelligenza artificiale che può procedere, ragionare, pianificare e agire. Non si tratta semplicemente di applicare modelli generativi alla robotica, né di aggiungere un'interfaccia linguistica a macchine già esistenti. La Physical AI indica una trasformazione più ampia: l'integrazione tra percezione, ragionamento, pianificazione, controllo, sensori, attuatori e infrastrutture meccaniche, con l'obiettivo di rendere i sistemi artificiali capaci di operare in ambienti materiali, variabili e non interamente prevedibili. L'IA simbiotica (la simbiosi fra umani e macchine) pone un problema etico decisivo: non basta stabilire quali principi debbano guidare l'IA, bisogna capire come questi principi possano essere tradotti in pratiche concrete. Dire che un sistema deve essere trasparente, responsabile, centrato sull'essere umano non è sufficiente. Occorre rendere tali criteri utilizzabili nella progettazione, nella valutazione e nell'impiego dei sistemi. È questo un compito che ancora non viene preso in seria considerazione dai policy makers.

Il 6 aprile 2026, Open AI ha pubblicato un documento di notevole portata: *“Industrial Policy for the Intelligence Age: Ideas to keep People First”*. (“La politica industriale per l'era dell'Intelligenza: idee per porre le persone al primo posto”). Vi si legge che Sam Altman propone una tassa sui robot, un fondo pubblico di ricchezza nazionale, la settimana lavorativa di trentadue ore a parità di salario, meccanismi automatici di sostegno al reddito ogniqualvolta la disoccupazione causata dall'IA supera una certa soglia. Il senso è chiaro: il capitalismo cibernetico incontra le proprie contraddizioni e che cerca di rimediare alla bisogna. Sappiamo cosa è stato il New Deal di Roosevelt e l'Era Progressista che ne è derivata. Allora, i riformatori erano i politici, i movimenti sociali, gruppi di intellettuali, e le aziende erano l'oggetto della regolazione. Nel caso in questione, è l'azienda stessa a proporre le politiche che dovrebbero mitigare gli effetti negativi dell'IA.

Qual è la perplessità che il documento di Open AI solleva e che papa Leone porta bene alla luce? Quella di dare per scontato che la “disruption” sia inevitabile, una forza tecnica che procede secondo la sua logica interna. Non ci si chiede se lo sviluppo accelerato verso la superintelligenza sia la cosa giusta da fare, né si pone in discussione la direzione e la velocità del processo. Ma allora cosa ne sarà della politica democratica? È bene ricordare che, come Aristotele insegnava, la democrazia è governo *del* popolo, *per* il popolo, *con* il popolo. Non ci si deve allora meravigliare se nel manifesto di Peter Thiel è scritto, a tutto tondo, che la democrazia deliberativa non solo a nulla serve oggi, ma è addirittura d'intralcio all'avanzamento del progresso. (Si rammenti che Peter Thiel è l'iniziatore dell'accelerazionismo, una influente corrente di pensiero nata in California alla fine del secolo scorso).

Quale “prezzo” paghiamo quando una piattaforma privata diventa il principale filtro di visibilità sociale, informativa e politica? Cosa diventa la democrazia quando la comunicazione pubblica viene riscritta per massimizzare le reazioni dentro formati che premiano l'urgenza emotiva? Si capisce allora perché l'IA non sia un mezzo assiologicamente neutro, come la MH ripete più di una volta. Un'anticipazione dei rischi associati all'avanzata del capitalismo cibernetico la troviamo nel libro del filosofo tedesco Günther Anders, del 1956 (tradotto in italiano con il titolo *L'uomo è antiquato*, Torino, Bollati Boringhieri, 2017). Già all'inizio degli anni Cinquanta, Anders intuisce che l'uomo stava diventando “obsoleto”. Sulla scorta della sua “filosofia della discrepanza”, Anders avanza l'idea che con la modernità si sia creata una asimmetria tra l'uomo che è rimasto basicamente lo stesso di sempre e i prodotti elettronici che egli ha creato e che continuamente lo oltrepassano e quindi lo dominano. (“Noi siamo inferiori a noi stessi”, p.14). Aveva cioè visto, come i bravi filosofi sanno fare, in anticipo sui tempi. Ecco perché il problema sopra sollevato, è qualcosa di straordinariamente importante. I capitalisti illuminati – e ce ne sono, anche se non molti - lo stanno capendo, tanto che si interrogano sul da farsi. Le nuove tecnologie aumentano bensì il tasso di profittabilità e di crescita dell'economia, ma a quale prezzo morale e spirituale ciò avviene? L'uomo vede solo i propri prodotti e si pensa (narcisisticamente) a immagine e somiglianza di sé stesso. Il punto è che lo scientismo tecnologico sogna molto, ma pensa poco; soprattutto

non guarda verso l'ontologia. Nella stagione della infocrazia, la questione centrale per chi non si rassegna a barattare la propria libertà e dignità per un "piatto di lenticchie" è se vi sarà tempo necessario per trovare risposte adeguate, prima che la dittatura della Tecnica metta a tacere le opinioni dissenzienti. Chi scrive è persuaso che questo tempo ci sarà. Il fondamento di tale speranza è nella presa d'atto che si va oggi diffondendo la consapevolezza che non è sostenibile una società in cui tutto si riduce, per un verso, a migliorare le transazioni basate sul principio dello scambio di equivalenti e, per l'altro verso, a far dipendere il proprio benessere da trasferimenti attuati da strutture di tipo assistenziale. Invero, non è capace di futuro la società in cui si dissolve il principio di fraternità. Non v'è felicità laddove esiste solamente il "dare per avere" oppure "il dare per dovere". Ecco perché né la visione singolarista del mondo, in cui tutto (o quasi) è scambio, né la visione statocentrica della società in cui tutto (o quasi) è doverosità, sono guide sicure per farci uscire dalle secche in cui la seconda grande trasformazione di tipo polanyano ci ha immessi. La proposta che ci viene dalla MH è saggiamente resa con l'immagine biblica di Neemia, che ricostruisce le mura di Gerusalemme, coordinando gli sforzi di ciascuno, senza imporre comandi. È questa la logica della sussidiarietà (circolare), che va applicata alla governance tecnologica, la quale si oppone sia alla delega totale al mercato sia alla regolamentazione centralizzata ad opera dello Stato.

6. Un'annotazione conclusiva

Chiudo con una considerazione di portata generale. La possibilità è sempre la combinazione di due elementi: le opportunità e la speranza. È sbagliato pensare che perché qualcosa diventi possibile sia necessario intervenire solamente sul lato delle opportunità, vale a dire sul lato delle risorse e degli incentivi. Invero, i problemi che abbiamo di fronte non si risolvono invocando un mero aumento di risorse, anche perché buona parte dei nostri problemi sono dovuti a un eccesso di risorse. (Si pensi alla competizione cosiddetta posizionale e ai guasti che essa sta provocando). Quel che è necessario perché la possibilità abbia a realizzarsi è insistere sull'elemento della speranza, la quale non è mai utopia. La speranza si alimenta con la creatività dell'intelligenza politica e con la purezza della passione civile. In

ogni caso è della distopia, assai più dell'utopia, che si deve temere quando si inizia un cammino. La distopia è la sindrome che colpisce chi soffre delle "passioni tristi", nel senso di B. Spinoza – la tristezza della delusione e della assenza di desiderio. Il paese che è stato culla dell'Umanesimo, prima, e del Rinascimento, poi, non può cadere vittima di simili disposizioni d'animo.

Mi piace terminare con un riferimento a Jean Bodin, filosofo francese del XVII sec., autore del noto saggio *Colloquio dei Sette*. Vi si narra il dibattito, piuttosto acceso, tra sette personaggi: un cattolico, un luterano, un calvinista, un ebreo, un mussulmano, un naturalista, un celtico. Dopo un lungo e pure violento contendere, i sette si abbracciano con amore reciproco per l'armonia ritrovata. Ciò avviene non grazie alla capacità argomentativa dei disputanti nel difendere le rispettive posizioni, ma all'impegno che ciascuno aveva posto nel dare conto "delle opere della santità suprema della propria vita". Ebbene, Papa Leone ci ammonisce che amare l'esistente è un atto di fede. Il che apre alla speranza che un futuro migliore sia concretamente possibile, ad una condizione però, quella efficacemente spiegata nella suggestiva sintesi papale della Lettera VII di Platone: "Le cose più profonde e importanti si imparano solo dopo molto tempo e molta fatica, impegnandosi nella discussione con gli altri a "sfregare" i concetti e le esperienze come fossero pietre focaie, finché in noi non scocchi la scintilla della comprensione" (MH, n. 140). Si tratta di una delle più straordinarie metafore della cultura occidentale: il sapere è un'accensione interiore, cioè del cuore, che non potrà essere scaricata da alcuna macchina.





c/o Scuola di Economia e Management Università di Bologna, sede di Forlì
P.le della Vittoria, 15 – 47121 Forlì – Tel. 0543.62327
ecofo.aiccon@unibo.it – www.aiccon.it

Stefano Zamagni

Economista e professore emerito dell'Università di Bologna, dove ha insegnato Economia Politica, e Adjunct Professor alla Johns Hopkins University, SAIS Europe. È Presidente della Commissione Scientifica di AICCON.