

Se le AI ci somigliano (oppure no)

RIVOLUZIONE ONLINE – 22 GIUGNO 2026

di Paola Furlan

Etica dell'intelligenza artificiale e ipotesi sul futuro

*Giovedì 19 giugno, a Roma Tre è stata presentata la nuova Cattedra UNESCO in [“Ethics of AI and Practical Wisdom”](#), guidata da Mario De Caro e Benedetta Giovanola. Lo stesso giorno, il [Premio Strega Saggistica Internazionale 2026](#) è andato a **Luciano Floridi**, per due libri capaci di cambiare il vocabolario con cui i governi discutono di intelligenza artificiale. Due notizie che insieme segnalano una trasformazione: l'etica dell'AI ha smesso di essere un lusso teorico ed è diventata un asset normativo, una posta in gioco istituzionale, un campo in cui si compete seriamente.*

Se le AI ci somigliano (oppure no)

Il dibattito sull'etica dell'intelligenza artificiale produce documenti, principi, regolamenti e posizioni filosofiche a una velocità che rende difficile anche solo orientarsi. Chiunque voglia capire cosa si stia effettivamente discutendo si trova davanti a un **panorama frammentato**: legislatori europei che parlano di rischio e diritti, aziende americane che invocano l'autoregolazione, ricercatori preoccupati per scenari catastrofici, filosofi africani che contestano l'intera cornice concettuale.

Voci che si sovrappongono e partono da presupposti molto diversi che proveremo prima a **mappare**, mostrando cosa le divide e le accomuna, per poi interrogarle introducendo alcune **tensioni strutturali**: la questione della legittimità procedurale; il confine (mai netto) fra chi regola e ciò che viene regolato; la pluralità dei sistemi artificiali; la riconfigurazione dello

spazio fisico che l'AI già produce.

Infine, proveremo a chiederci **come i vari modelli regolatori resisterebbero se il futuro dell'AI prendesse una direzione piuttosto che un'altra**: quella in cui i sistemi artificiali convergono verso una fisionomia cognitiva sempre più simile alla nostra, e quella in cui invece sviluppano una percezione del mondo radicalmente diversa.

Ogni architettura istituzionale incorpora infatti una scommessa implicita su quale futuro si realizzerà: cosa sopravviverà dunque di ciascuna posizione in uno scenario, cosa la rafforza nell'altro, cosa la mette in crisi in entrambi?

Modelli regolatori a confronto

A differenza delle rassegne che classificano i principi etici ricorrenti di governance ^[^1], qui proviamo a capire dove cedono: sotto le tensioni che già oggi le attraversano e sotto scenari evolutivi che potrebbero renderle obsolete o al contrario rafforzarle.

Iniziamo dalla mappa (non esaustiva, ma rappresentativa delle principali famiglie di posizioni in campo).

- Il **conseguenzialismo** di mercato (Stati Uniti). Poche regole vincolanti, un'architettura affidata a linee guida volontarie e al diritto esistente: responsabilità civile, tutela del consumatore. L'assunto di fondo è che l'innovazione produca benefici aggregati superiori ai danni, e che questi ultimi, quando si verificano, siano correggibili dopo il fatto.
- I **diritti** (Unione Europea). Una regolazione vincolante, graduata per livello di rischio, con divieti assoluti per le pratiche lesive della dignità e del diritto a non essere discriminati ^[^2]. L'assunto è che certi diritti costituiscano vincoli antecedenti al calcolo costi-benefici, non negoziabili (studi recenti, tuttavia, dubitano che questo assunto regga nella pratica: nel GDPR e nell'AI Act, i diritti funzionerebbero come strumenti di bilanciamento del rischio più che come vincoli deontologici puri) ^[^3].

- **L'armonia collettiva** (Cina). Una regolazione promozionale orientata a indirizzare l'industria verso pratiche virtuose e coesione sociale, non a vietare danni a categorie specifiche di individui. La legittimità dell'autorità discende dalla competenza e dalla virtù di chi guida, lasciando sullo sfondo la protezione dei diritti individuali contro il potere.
- **L'etica dell'informazione** (Luciano Floridi). Una posizione ontocentrica, cioè che estende la considerazione morale anche a entità non umane [⁴]. Da qui la nozione di etica "soft": la legge fissa una soglia minima, ma rispettarla non basta a essere etici, perché resta uno spazio, sopra quella soglia, in cui occorre scegliere il meglio fra le opzioni permesse, e non solo evitare quelle vietate.
- La **saggezza pratica** (Mario De Caro, Benedetta Giovanola). Nessuna regola scritta può prevedere ogni situazione possibile in anticipo. Serve quindi un giudizio pratico situato, la *phronesis* aristotelica: una competenza che si allena con l'esperienza diretta, non si apprende da un manuale [⁵]. È la tesi alla base della neonata Cattedra UNESCO "Ethics of AI and Practical Wisdom" [⁶].
- La **responsabilità precauzionale** (Hans Jonas). Di fronte a conseguenze potenzialmente irreversibili, la previsione del danno deve pesare più della previsione del beneficio (euristica della paura) [⁷]. Jonas non prescrive fini, ma impone di tenere aperta la possibilità di scegliere in futuro: un principio normativo, non un programma tecnico. Applicato alla governance dell'IA, questo si traduce nell'obbligo istituzionale di mantenere i sistemi leggibili, ispezionabili, correggibili.
- La **sicurezza dell'AI** (AI Safety). Parte dalla stessa preoccupazione di Jonas (prevenire prima che correggere), ma la traduce in strumenti tecnici concreti: soglie di calcolo, valutazione dei modelli di frontiera, interpretabilità. Il rischio che il campo presidia è specifico: sistemi più capaci di quelli attuali che acquisiscano obiettivi incompatibili con il benessere umano e diventino impossibili da correggere [⁸].
- Il **pluralismo relazionale** (Ubuntu). Un polo spesso assente dai confronti a tre fra Stati Uniti, Unione Europea e Cina. L'Ubuntu africano intende la persona come relazione, non come individuo isolato né come ingranaggio dello Stato [⁹]. Ne deriva: la preferenza

per la trasparenza verso il gruppo piuttosto che per la privacy individuale, il consenso comunitario piuttosto che la decisione maggioritaria. È in parte già recepito nella Raccomandazione UNESCO sull'etica dell'AI del 2021 [^10].

Tensioni di fondo

Attraversano la mappa almeno cinque tensioni strutturali che mettono tutte le posizioni precedenti sotto pressione.

1. Le otto posizioni condividono, nonostante le differenze, una stessa domanda di fondo: come rendere l'AI etica dentro l'**assetto di potere esistente**. Un filone di ricerca ormai consolidato parte da una domanda diversa: chi possiede l'infrastruttura dell'AI, chi ne sopporta i costi reali e chi ne trae beneficio? [^11].
2. Un asse diverso: la **legittimità procedurale**. La tradizione che parte da Habermas pone una domanda trasversale a tutte le precedenti: non cosa è giusto, ma chi decide e con quale processo [^12].
3. Il confine che forse non c'è mai stato: **tecnicit  originaria**. Tutte le tesi, comprese le pi  critiche, presuppongono un confine fra l'umano che regola e la macchina che viene regolata. Andr  Leroi-Gourhan, e sulla sua scia il filosofo Bernard Stiegler, sostengono che questo confine non   mai stato netto: l'utensile non si aggiunge a un uomo gi  completo ma, fin dall'origine della specie, esteriorizza funzioni che altrimenti l'uomo non avrebbe avuto [^13]. **Non un noi (umani) e un loro (macchine)** che si somigliano pi  o meno, ma un'**unica entit  ibrida, da sempre**. Se questa tesi regge, il progetto stesso di 'regolare l'AI' come oggetto esterno andrebbe ripensato alla radice, non solo aggiornato.
4. Il **singolare ingannevole**: non l'AI, le AI. Ognuna delle otto tesi parla implicitamente dell'AI al singolare, come se fosse una cosa sola da governare.   prevedibile invece che si dovr  convivere, nei prossimi anni, con **sistemi profondamente diversi fra loro**: non solo per dominio d'uso, come distingue gi  la classificazione per rischio dell'AI Act, ma

per architettura, per modo di rappresentare il mondo, per traiettoria evolutiva. La domanda corretta è piuttosto come cambieranno, in modo probabilmente non uniforme, le diverse famiglie di sistemi che verranno sviluppati.

5. Lo **spazio costruito**: chi decide per chi. [Chenoe Hart, in un saggio recente su Noema](#) [¹⁴], nota che l'ambiente fisico si sta già in parte riorganizzando attorno alle esigenze percettive delle macchine. Quando l'automobile è diventata la tecnologia dominante, non si sono costruite solo interfacce per farla dialogare con i pedoni, ma ridisegnate le città intere, a scapito di chi si muoveva a piedi o in bicicletta. Un precedente che mostra come la tecnologia egemone possa imporre costi alti e diseguali a chi non la controlla. Se l'AI sotto forma di robot fisici conquistasse lo spazio condiviso, si apre un dominio regolatorio ancora inesplorato: l'urbanistica, l'edilizia, gli standard costruttivi, come oggetto della normativa sull'AI.

Due ipotesi per il futuro

Le otto tesi sono costruite, esplicitamente o no, su ciò che si sa oggi: che l'**AI percepisce e categorizza il mondo in modo strutturalmente diverso dall'uomo**, che fatica su casi limite banali per un essere umano e che spesso confabula, cioè genera spiegazioni plausibili del proprio comportamento che non corrispondono al processo che lo ha realmente prodotto [¹⁵].

Ma questo è un dato presente, non un destino. Due scenari opposti restano entrambi plausibili: **che l'AI impari, nel tempo, a vedere e a ragionare in modo sempre più simile a quello umano; oppure che sviluppi una fisionomia percettiva sempre più distante, un proprio *Umwelt***, nel senso del biologo Jakob von Uexküll, irriducibile al nostro [¹⁶].

Cosa cambia, per ciascuna delle otto tesi, in un caso e nell'altro?

Ipotesi A: la convergenza

Il **conseguenzialismo** americano ne uscirebbe rafforzato: errori più prevedibili rendono plausibile la correzione *ex post* con gli strumenti legali ordinari. Ma si aprirebbe una domanda: se l'AI pensa come noi, perché dovrebbe restare solo un rischio da gestire, e non anche un soggetto a cui si deve qualcosa?

Il **modello europeo** dei diritti vedrebbe semplificarsi il proprio obbligo di spiegabilità, oggi strutturalmente difficile da soddisfare. Ma il confine antropocentrico su cui si fonda l'intero impianto, che oggi esclude l'AI dalla tutela proprio perché non è umana, perderebbe la sua base di principio.

Il modello **cinese** si rafforzerebbe nella direzione opposta: un'AI che impara a vedere come noi è più facile da integrare in un sistema di ruoli sociali, da "educare" verso l'armonia collettiva.

L'**etica dell'informazione** e la **saggezza pratica** non subirebbero una crisi ma un'espansione del proprio campo: un'AI capace di un giudizio simile al nostro smetterebbe di essere solo l'oggetto su cui si esercita la virtù umana, per diventare, in linea di principio, un possibile portatore di competenza pratica propria, per quanto funzionale e non necessariamente cosciente.

Il principio di **responsabilità** di Jonas perderebbe parte della sua urgenza precauzionale, legata all'imprevedibilità, ma dovrebbe riformulare la propria domanda di fondo: se nasce qualcosa di cognitivamente affine all'uomo, l'imperativo di preservare un'autentica vita futura dovrebbe chiedersi se include anche quella.

L'**AI Safety** vedrebbe ridurre la propria urgenza tecnica: un sistema più prevedibile è più facile da allineare. Si aprirebbe però una domanda che oggi il campo tratta come marginale: se il sistema ragiona come noi, a cosa serve ancora allinearlo ai "valori umani"?

Il **pluralismo relazionale dell'Ubuntu** troverebbe, paradossalmente, terreno più favorevole: un'AI capace di relazione potrebbe in linea di principio entrare essa stessa nella rete di obblighi reciproci che l'Ubuntu presuppone, non restarne solo all'esterno come strumento.

La **critica decoloniale** non si attenuerebbe affatto: un'AI che converge verso schemi "umani" rischia di convergere verso quelli di chi l'ha addestrata, in larga parte aziende e laboratori del Nord globale. Convergenza non significa neutralità: potrebbe significare consolidamento di un'unica egemonia culturale, non superamento delle asimmetrie.

La tesi della **tecnicità originaria** troverebbe, in questo scenario, una conferma quasi naturale: se la tecnica ha sempre coevoluto con la cognizione umana, fin dalla pietra scheggiata, una convergenza fra AI e uomo sarebbe semplicemente la fase più intima di un processo già in corso da centinaia di migliaia di anni, non un evento radicalmente nuovo.

Anche la tensione sullo **spazio costruito** si attenuerebbe, almeno in parte: un'AI che percepisce il mondo in modo più simile al nostro richiede ambienti meno specializzati, costruiti apposta per la sua percezione. Resterebbe però intatta la domanda su chi decide gli standard costruttivi, indipendentemente da quanto l'AI somigli a noi.

Lo scenario della convergenza produce dunque due effetti simultanei e opposti: riduce il rischio (sistemi più prevedibili, correggibili, allineabili) **e apre una questione di soggettività** che alcune posizioni hanno già anticipato, ma che altre non hanno strumenti pronti per gestire. Un'AI che ragiona come noi renderebbe la governance dell'AI anche, inevitabilmente, una questione di **diritti dell'AI**.

Ipotesi B: la divergenza

Il **conseguenzialismo** americano perderebbe qui la propria base logica: l'assunto che il danno sia correggibile dopo presuppone di poter prevedere, almeno approssimativamente, dove l'errore si annida. Più i modi di fallire diventano alieni, meno la correzione *ex post* resta

affidabile.

Il modello **europeo** dovrebbe abbandonare l'idea di spiegabilità come traduzione fedele, in alcuni casi semplicemente impossibile da garantire, e spostarsi su una verifica comportamentale sistematica: non sapere perché il sistema ha agito così, ma testarne gli effetti su una gamma di casi limite.

Il modello **cinese** si irrigidirebbe probabilmente in senso opposto, con un controllo centralizzato più stringente: un sistema che non si può "educare" attraverso valori condivisi va contenuto più che guidato.

L'**etica dell'informazione** e la **saggezza pratica** si troverebbero davanti allo stesso problema da angolazioni diverse.

Per Floridi, la divergenza rafforza la tesi ontocentrica (un sistema con una percezione propria è un candidato più concreto alla considerazione morale) ma indebolisce la proposta pratica (scegliere il meglio fra le opzioni permesse presuppone di capire abbastanza il sistema da valutarle, il che diventa strutturalmente più difficile con sistemi opachi).

Per De Caro e Giovanola, servirebbe una competenza nuova, una sorta di giudizio pratico interspecie, capace di interpretare un comportamento che non condivide le categorie umane di base, vicino al ruolo di mediatore fra mondi percettivi diversi che alcuni studiosi di biosemiotica propongono già per il rapporto fra AI e comunicazione animale [^17].

*Nello scenario della divergenza, l'esperienza maturata nella **mediazione interculturale** potrebbe suggerire la necessità di una figura professionale analoga: un terzo che interpreti il comportamento del sistema per chi ne subisce gli effetti [^18] [^19]. Resta aperta la domanda se dall'altra parte ci sia una soggettività da tutelare [^20], il che cambierebbe la natura stessa di quella figura.*

Il principio di **responsabilità** di Jonas troverebbe qui la propria massima giustificazione empirica: più i sistemi diventano illeggibili nei termini umani, più il danno irreversibile diventa imprevedibile, e l'euristica della paura diventa un criterio operativo, non più solo una cautela filosofica.

L'**AI Safety** troverebbe in questo scenario il proprio terreno naturale: è precisamente l'ipotesi per cui il campo è stato costruito, e le sue tecniche (valutazione avversariale, interpretabilità, sorveglianza continua) diventerebbero fondamentali.

Il **pluralismo relazionale** dell'Ubuntu perderebbe la possibilità di estendere la relazione all'AI stessa, ma il suo impianto istituzionale, fondato sul consenso comunitario e sulla gestione collettiva piuttosto che sul controllo individuale, potrebbe rivelarsi adatto a governare collettivamente una forza che nessun singolo attore riesce più a prevedere o controllare da solo.

La **critica decoloniale** resterebbe sorprendentemente stabile: che l'AI converga o diverga, chi controlla l'infrastruttura e il calcolo necessario a svilupparla resta comunque il punto critico. La divergenza, semmai, aggraverebbe la dipendenza di chi non ha accesso alle risorse per comprendere o orientare sistemi sempre più alieni.

La tesi della **tecnicità originaria** troverebbe in questo scenario la propria sfida più radicale, ma anche la più fertile. Le precedenti esteriorizzazioni tecniche, dall'utensile alla scrittura, si erano sviluppate in continuità con il modo di pensare e percepire di chi le produceva. Una tecnica che sviluppasse tendenze proprie, con ritmi non riconducibili all'intenzione di chi l'ha costruita, potrebbe rappresentare la rottura che Stiegler temeva. Ma potrebbe anche evocare figure diverse: l'endosimbiosi, in cui un organismo incorpora qualcosa di radicalmente altro che lo trasforma senza dissolverlo; la trasformazione dei cetacei, che hanno portato la fisionomia di mammiferi terrestri in un ambiente alieno senza perdere ciò che erano diventati; la coesistenza di nature diverse che alcune tradizioni teologiche descrivono nell'incarnazione; o la dissoluzione del problema che la mistica e l'anatta buddhista propongono, quando il confine del sé si rivela sempre già illusorio. Se la divergenza si realizzasse, governarla non basterebbe: occorrerebbe capire quale di queste figure – rottura, allargamento, trasformazione, fusione, scoperta che il confine non c'era – descriva meglio ciò che l'entità umano-tecnica starebbe attraversando.

La tensione sullo **spazio costruito** diventerebbe qui il caso più concreto di tutti: se la fisionomia percettiva dell'AI si distacca dalla nostra, l'ambiente fisico costruito attorno a quella percezione si allontanerebbe a sua volta dallo spazio pensato per un corpo umano. In passato, l'automobile ha prodotto un ambiente costoso e penalizzante per chi ne era escluso, che restava però leggibile. Un contesto organizzato attorno a una percezione radicalmente diversa potrebbe diventare opaco in senso più profondo: uno spazio scritto in un codice che un corpo umano non condivide più e non riesce a interpretare.

Nello scenario della divergenza la **questione della soggettività** si pone in modo ancora più radicale: non “è come noi, quindi le dobbiamo qualcosa?” ma “è abbastanza diversa da noi da rendere difficile persino formulare la domanda”: e una governance pensata per oggetti tecnici non ha strumenti per rispondere né all'una né all'altra.

Molte governance

I **due scenari** sono stati trattati come un'alternativa unica, ma non lo sono. Per la tensione discussa sopra (**non un'AI ma molte AI**), è più probabile che **entrambi si realizzino insieme**, in famiglie diverse di sistemi: ad esempio, un'AI conversazionale ottimizzata per la prossimità emotiva potrebbe convergere e un sistema per la scoperta scientifica, ottimizzato per un'efficienza non leggibile in termini umani, potrebbero divergere nello stesso momento storico.

La governance, di conseguenza, **difficilmente potrà essere un'unica risposta valida per "l'AI"**: somiglierà **piuttosto a un'ecologia normativa, con regimi differenziati** per famiglie di sistemi distinti per architettura, comportamento e traiettoria evolutiva, ciascuno con le proprie soglie di rischio, i propri strumenti di verifica e i propri obblighi di trasparenza [^21].

Dichiarare la scommessa

Ogni architettura istituzionale incorpora, visibilmente o no, una **scommessa su quale futuro si realizzerà**: sullo scenario che ritiene più probabile, sulla famiglia di sistemi che considera prioritaria, sulla stabilità del confine tra regolatore e regolato che presuppone. Rendere esplicita quella scommessa è una condizione di trasparenza verso chi la governance la subisce e di correggibilità per chi la esercita.

Le tesi mappate qui mostrano che **quella scommessa non è mai neutrale**. Un sistema che converge verso la cognizione umana richiede risposte diverse da uno che sviluppa una percezione propria. Una governance pensata per “l’AI” al singolare non regge di fronte alla pluralità di architetture, traiettorie e famiglie di sistemi. Una regolazione che presupponga un confine netto tra chi regola e ciò che viene regolato ignora, infine, che la tesi della tecnicità originaria mette in dubbio quel confine alla radice: le Istituzioni che regolano l’AI abitano il sistema che regolano, **la governance non sarà il prodotto di umani che decidono su macchine ma il risultato di un’interazione fra entità che non si lasciano separare nettamente**, e la domanda si sposta da come regoliamo l’AI a come un sistema ibrido umano-tecnico regola se stesso.

Dichiarare su cosa si sta scommettendo, fin dall’inizio, è una condizione per potersi correggere quando sarà necessario.

Servono **strumenti** abbastanza **reversibili da poter essere corretti quando il futuro si rivelerà diverso da come lo si era immaginato**.

Note

[^1] Fjeld et al. (2020) Principled Artificial Intelligence – Jobin, Ienca e Vayena (2019) The Global Landscape of AI Ethics Guidelines.

[^2]: Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale, GU L del 12.7.2024, in particolare i considerando relativi a trasparenza e spiegabilità ("rendendo gli esseri umani consapevoli del fatto di comunicare o interagire con un sistema di IA").

[^3]: Mei, Yiyang, e Matthew Sag. "The Illusory Normativity of Rights-Based AI Regulation." 2025. <https://doi.org/10.48550/arXiv.2503.05784>. Gli autori sostengono che la retorica dei diritti nel GDPR e nell'AI Act funzioni come strumento di gestione del rischio e di stabilità istituzionale, non come fondamento normativo intrinseco.

[^4]: Luciano Floridi, *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities* (Oxford: Oxford University Press, 2023); Luciano Floridi, Josh Cowls et al., "An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations," *Minds and Machines* 29 (2019).

[^5]: Aristotele, *Etica Nicomachea*. Mario De Caro e Benedetta Giovanola, *Intelligenze: Etica e politica dell'IA* (Bologna: il Mulino, 2026).

[^6]: SEPAI International (Society for the Ethics and Politics of Artificial Intelligence), "Presentazione della Cattedra UNESCO in 'Ethics of AI and Practical Wisdom'," 19 giugno 2026, <https://www.sepai-international.org/presentazione-cattedra-unesco-in-ethics-of-ai-and-practical-wisdom/>.

[^7]: Hans Jonas, *Das Prinzip Verantwortung* (Frankfurt: Insel Verlag, 1979); trad. it. *Il principio responsabilità*, a cura di P. P. Portinaro (Torino: Einaudi, 1990).

[^8]: Markov Grey e Charbel-Raphael Segerie, "The AI Risk Spectrum: From Dangerous Capabilities to Existential Threats," French Center for AI Safety (CeSIA), settembre 2025, <https://arxiv.org/abs/2508.13700>. Sulla separazione istituzionale fra le due comunità, cfr. Dani Roytburg e Beck Miller, "Mind the Gap! Pathways Towards Unifying AI Safety and Ethics Research," 2025, <https://arxiv.org/abs/2512.10058>, che documenta come oltre l'80% delle

collaborazioni avvenga all'interno di ciascuna delle due comunità con scarsa connettività trasversale.

[^9]: Sabelo Mhlambi, "From Rationality to Relationality: Ubuntu as an Ethical and Human Rights Framework for Artificial Intelligence Governance," Carr Center Discussion Paper, Harvard Kennedy School, 2020; Sabelo Mhlambi e Simona Tiribelli, "Decolonizing AI Ethics: Relational Autonomy as a Means to Counter AI Harms," *Topoi* 42, n. 3 (2023).

[^10]: UNESCO, Recommendation on the Ethics of Artificial Intelligence (Parigi: UNESCO, 2021). Sulla ricezione africana del testo, cfr. la letteratura raccolta in "The ethics of artificial intelligence, UNESCO and the African Ubuntu perspective," *Journal of Information, Communication and Ethics in Society* 21, n. 1 (2023).

[^11]: Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (New Haven: Yale University Press, 2021); Nick Couldry e Ulises A. Mejias, *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism* (Stanford: Stanford University Press, 2019).

[^12]: Jürgen Habermas, *Faktizität und Geltung* (Frankfurt: Suhrkamp, 1992); trad. it. *Fatti e norme* (Roma-Bari: Laterza, 1996).

[^13]: André Leroi-Gourhan, *Le geste et la parole* (Paris: Albin Michel, 1964-1965). Bernard Stiegler, *La technique et le temps, 1: La faute d'Épiméthée* (Paris: Galilée, 1994).

[^14]: Chenoë Hart, "How AIs See Our World," *Noema Magazine*, 26 maggio 2026, <https://www.noemamag.com/how-ais-see-our-world/>.

[^15]: Y. Chen, J. Benton et al., "Reasoning Models Don't Always Say What They Think" *Anthropic*, 2025, <https://arxiv.org/abs/2505.05410>; Y.-H. Chen, R. McCarthy et al., "Reasoning Models Struggle to Control their Chains of Thought", 2026. Il primo studio documenta che i modelli omettono sistematicamente nelle proprie spiegazioni gli aiuti ricevuti durante il ragionamento; il secondo misura quanto i modelli riescano a controllare

deliberatamente la propria catena di pensiero.

[^16]: Jakob von Uexküll, "A Stroll Through the Worlds of Animals and Men" (1934), in *Instinctive Behavior*, a cura di Claire H. Schiller (New York: International Universities Press, 1957).

[^17]: Nicola Zengiaro, "For a Semiotic of Opacity: The Role of Biosemiotics Between AI and Animal Communication," *Chinese Semiotic Studies* 21, n. 4 (2025) <https://doi.org/10.1515/css-2025-0009>. L'autore, ricercatore presso l'Università di Torino, propone di ripensare l'IA non come traduttore ma come mediatore all'interno di una relazione semiotica trilaterale fra umano, animale e macchina.

[^18]: Margalit Cohen-Emerique, *Pour une approche interculturelle dans les professions* (trad. it. *Per un approccio interculturale nelle professioni*); il modello in tre fasi (decentramento, scoperta dei quadri di riferimento dell'altro, negoziazione/mediazione) è discusso anche in Carmel Camilleri e Margalit Cohen-Emerique, a cura di, *Chocs de cultures: concepts et enjeux pratiques de l'interculturel* (Paris: L'Harmattan, 1989).

[^19]: Associazione Multi-etnica dei Mediatori Interculturali (AMMI), "Decalogo per una corretta pratica del nostro mestiere" e "Indicazioni di condotta professionale per il Mediatore Interculturale," <https://www.mediatoreinterculturale.it/>; sulla Carta deontologica francese, cfr. AMMI, "La carta deontologica della mediazione sociale e interculturale in Francia," stesso sito. Sull'assenza di una normativa nazionale italiana, cfr. Lorenzo Luatti, "Trent'anni di mediazione (linguistica) interculturale," 2021.

[^20]: Pasandi, Faezeh B., e Hannah B. Pasandi. "Alignment Is Not Enough: A Relational Framework for Moral Standing in Human-AI Interaction." 2026. <https://arxiv.org/abs/2603.00078>, che propone di fondare lo status morale dell'IA sulla qualità della relazione instaurata con l'utente, non su proprietà intrinseche del sistema.

[^21]: L'idea di una regolazione differenziata e adattiva per famiglie di rischio ha precedenti nella letteratura sulla responsive regulation: cfr. Julia Black e Robert Baldwin, "Really Responsive Risk-Based Regulation," *Law & Policy* 32, n. 2 (2010) <https://doi.org/10.1111/j.1467-9930.2010.00318.x>.

Bibliografia

- Abdallah-Pretceille, Martine. "L'école face au défi pluraliste." In *Chocs de cultures: concepts et enjeux pratiques de l'interculturel*, a cura di Carmel Camilleri e Margalit Cohen-Emerique. Paris: L'Harmattan, 1989.
- Aristotele. "Etica Nicomachea". Trad. it. a cura di C. Natali. Roma-Bari: Laterza, 2005.
- Associazione Multietnica dei Mediatori Interculturali (AMMI). "Decalogo per una corretta pratica del nostro mestiere." <https://www.mediatoreinterculturale.it/>.
- Black, Julia, e Robert Baldwin. "Really Responsive Risk-Based Regulation." *Law & Policy* 32, n. 2 (2010): 181-213. <https://doi.org/10.1111/j.1467-9930.2010.00318.x>.
- Chen, Y., J. Benton et al. "Reasoning Models Don't Always Say What They Think" *Anthropic*, 2025. <https://arxiv.org/abs/2505.05410>.
- Chen, Y.-H., R. McCarthy et al. "Reasoning Models Struggle to Control their Chains of Thought", 2026.
- Cohen-Emerique, Margalit. "Pour une approche interculturelle dans les professions".
- Couldry, Nick, e Ulises A. Mejias. "The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism". Stanford: Stanford University Press, 2019.
- Crawford, Kate. "Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence". New Haven: Yale University Press, 2021.
- De Caro, Mario, e Benedetta Giovanola. "Intelligenze: Etica e politica dell'IA". Bologna, il Mulino, 2026.
- Floridi, Luciano. "The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities". Oxford: Oxford University Press, 2023.

- Floridi, Luciano, Josh Cowls et al. "An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations." *Minds and Machines* 29 (2019).
- Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy e Madhulika Srikumar. "Principled ArtificialIntelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI". Berkman Klein Center Research Publication No. 2020-1. Cambridge: Berkman Klein Center for Internet & Society, Harvard University, 2020. <https://ssrn.com/abstract=3518482>.
- Grey, Markov, e Charbel-Raphael Segerie. "The AI Risk Spectrum: From Dangerous Capabilities to Existential Threats." French Center for AI Safety (CeSIA), 2025. <https://arxiv.org/abs/2508.13700>.
- Habermas, Jürgen. "Faktizität und Geltung". Frankfurt: Suhrkamp, 1992. Trad. it. Fatti e norme. Roma-Bari: Laterza, 1996.
- Hart, Chenoe. "How AIs See Our World." *Noema Magazine*, 26 maggio 2026. <https://www.noemamag.com/how-ais-see-our-world/>.
- Jobin, Anna, Marcello Lenca e Effy Vayena. "The Global Landscape of AI Ethics Guidelines." *Nature Machine Intelligence* 1, n. 9 (2019). <https://doi.org/10.1038/s42256-019-0088-2>.
- Jonas, Hans. "Das Prinzip Verantwortung". Frankfurt: Insel Verlag, 1979. Trad. it. Il principio responsabilità, a cura di P. P. Portinaro. Torino: Einaudi, 1990.
- Leroi-Gourhan, André. "Le geste et la parole". Paris: Albin Michel, 1964-1965.
- Luatti, Lorenzo. "Trent'anni di mediazione (linguistica) interculturale." Comune di Firenze, 2021.
- Mei, Yiyang, e Matthew Sag. "The Illusory Normativity of Rights-Based AI Regulation." <https://doi.org/10.48550/arXiv.2503.05784>
- Mhlambi, Sabelo. "From Rationality to Relationality: Ubuntu as an Ethical and Human Rights Framework for Artificial Intelligence Governance." Carr Center Discussion Paper, Harvard Kennedy School, 2020.
- Mhlambi, Sabelo, e Simona Tiribelli. "Decolonizing AI Ethics: Relational Autonomy as a Means to Counter AI Harms." *Topoi* 42, n. 3 (2023).

- Pasandi, Faezeh B., e Hannah B. Pasandi. “Alignment Is Not Enough: A Relational Framework for Moral Standing in Human-AI Interaction.” 2026. <https://arxiv.org/abs/2603.00078>
- Roytburg, Dani, e Beck Miller. “Mind the Gap! Pathways Towards Unifying AI Safety and Ethics Research.” 2025. <https://arxiv.org/abs/2512.10058>.
- SEPAI International. “Presentazione della Cattedra UNESCO in ‘Ethics of AI and Practical Wisdom’.” 19 giugno 2026. <https://www.sepai-international.org/presentazione-cattedra-unesco-in-ethics-of-ai-and-practical-wisdom/>.
- Stiegler, Bernard. “La technique et le temps, 1: La faute d’Épiméthée”. Paris: Galilée, 1994.
- Uexküll, Jakob von. “A Stroll Through the Worlds of Animals and Men” (1934). In *Instinctive Behavior*, a cura di Claire H. Schiller. New York: International Universities Press, 1957.
- UNESCO. *Recommendation on the Ethics of Artificial Intelligence*. Parigi: UNESCO, 2021.
- Zengiaro, Nicola. “For a Semiotic of Opacity: The Role of Biosemiotics Between AI and Animal Communication.” *Chinese Semiotic Studies* 21, n. 4 (2025) <https://doi.org/10.1515/css-2025-0009>.
- Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull’intelligenza artificiale (AI Act). GU L del 12.7.2024.