Ontologie del Domani: “Social Ontology Faces the Future”

*Paolo Valore**

È possibile tracciare un parallelismo ontologico tra i processi decisionali di un’intelligenza artificiale generativa, come ChatGPT, e le delibere di un corpo istituzionale complesso quale un’Università? In che misura l’aggregazione massiva di Big Data condivide la struttura metafisica dei processi democratici? E in che modo la filosofia dovrà farsi carico di questi problemi? Questi sono alcuni degli interrogativi che hanno costituito l’impalcatura teorica di un importante convegno internazionale, dal titolo “Social Ontology Faces the Future II: Artificial & Collective Agents”, ospitato il 19 e 20 giugno 2025 presso il centro di ricerca ANAMED della Koç University a Istanbul. L’evento ha cercato di superare la frammentazione disciplinare per mettere a sistema le ricerche sugli agenti cognitivi artificiali con quelle dedicate all’agenzia collettiva. La premessa del *workshop* è che questi due ambiti, spesso trattati separatamente, possano offrire chiavi di lettura essenziali per comprendere la natura stessa dell’azione. L’iniziativa, supportata dalla *Society of Applied Philosophy*, dalla *International Social Ontology Society* e dal *College of Social Sciences and Humanities* della Koç University, ha richiamato studiosi provenienti da diverse istituzioni internazionali, quali l’Università di Oxford, la McGill University (Canada), la Jagiellonian University di Cracovia e

* paolo.valore@unimi.it

l'Università di Helsinki, ma anche centri di ricerca italiani come l'Istituto Universitario Salesiano di Venezia.

Andrew Fyfe (Boğazici University), ha aperto la prima sessione con una relazione densa e provocatoria intitolata "How to Achieve Temporally Extended Agency in Artificial Agents Using Intra-Personal Collective Agency". Fyfe ha invitato a ripensare cosa renda un sistema artificiale un vero agente cognitivo. La maggior parte delle discussioni odierne si concentra sulle caratteristiche sincroniche (in un singolo istante), mentre Fyfe ha argomentato che la vera sfida è l'agenzia diacronica: la capacità di agire come un'unità unificata nel tempo. Distinguendo tra modelli "sottili" (come il test di Turing) e "spessi" di agenzia, Fyfe ha proposto una soluzione ontologica affascinante: modellare l'agente artificiale come un "gruppo intra-personale". In questa visione, le diverse sezioni temporali dell'agente devono coordinarsi condividendo una prospettiva "del noi". La conclusione è stata critica verso gli attuali sistemi di intelligenza artificiale: mancando di questa infrastruttura deliberativa "del noi" per coordinarsi con i propri sé futuri, essi falliscono nel costituirsi come agenti cognitivi temporalmente estesi, rimanendo ontologicamente frammentati.

Strahinja Đorđević (Università di Belgrado) ha spostato il *focus* sulle dinamiche epistemiche con l'intervento "Group Agency, Artificial Intelligence, and the Epistemic Dimensions of Third-Degree Monitoring Responsibility". Il cuore del discorso ha riguardato il fenomeno dello "sconfinamento epistemico". Đorđević ha analizzato se le organizzazioni e le intelligenze artificiali, in quanto agenti autonomi ma privi di intenzionalità umana classica, possano essere considerati detentori di una "responsabilità di monitoraggio di terzo grado" (intesa come il dovere strutturale di agire da filtro sistemico contro la diffusione di giudizi incompetenti,

essendo il primo grado relativo a chi emette un giudizio, il secondo grado relativo all'ascoltatore e il terzo all'infrastruttura che gestisce lo scambio). Il problema sollevato è strutturale: sebbene questi agenti influenzino l'ambiente epistemico, il loro distacco dalla comprensione semantica profonda solleva dubbi sulla loro capacità di garantire la validità dei giudizi. L'intervento ha quindi oscillato tra l'analisi delle capacità operative e i limiti ontologici di entità che, pur agendo, potrebbero non possedere la competenza necessaria per essere custodi dell'informazione.

Kayla Carnation, ingegnere specializzata in Computer Science presso l'Università della Pennsylvania con un *background* nella progettazione di sistemi per la difesa e la sicurezza USA, ha presentato "Epistemic Emergence in Human and AI Systems". Carnation ha offerto una raffinata analisi, comparando i fondamenti dell'intelligenza emergente nei collettivi umani rispetto ai sistemi artificiali. La tesi centrale è che vi sia una differenza radicale nella fondazione metafisica: l'intelligenza collettiva umana emerge da stati mentali e intenzionalità condivisa, mentre quella dell'IA si fonda su strutture computazionali e aggregazione statistica. La relatrice ha sostenuto che nessun agente cognitivo (né quello biologico isolato né quello sintetico) possiede oggi un'agenzia epistemica completa. La soluzione proposta è l'integrazione in strutture ibride, dove l'intenzionalità umana e la potenza di calcolo della macchina si fondono, colmando le rispettive lacune ontologiche.

Il tema della responsabilità è tornato prepotentemente nell'intervento di Pelin Kasar (Central European University), "The Responsibility Gap as a Problem of Tracing". Kasar ha affrontato i preoccupanti "vuoti di responsabilità" interpretandoli come fallimenti della strategia del "tracciamento". Utilizzando l'analogia dell'autista ubriaco

(responsabile non per l'azione al momento dell'incidente, ma per la scelta precedente di bere), Kasar ha mostrato come questa logica si inceppi con le IA. Nel caso di azioni guidate da pregiudizi impliciti o algoritmi opachi, spesso manca quel momento pregresso di controllo e conoscenza consapevole necessario per ancorare la responsabilità morale. L'intervento ha suggerito che trattare questi casi come "azioni non intenzionali" potrebbe aiutarci a mappare meglio lo statuto etico di questi nuovi agenti cognitivi.

Niël Conradie (RWTH Aachen University) ha proposto una difesa pragmatica con "Human-Machine Interaction and Collectives". Rifiutando una visione binaria dei vuoti di responsabilità tecnologici, Conradie ha introdotto il concetto di "spettro di interattività". La sua ipotesi ontologica è che, all'aumentare della profondità dell'interazione, l'uomo e la macchina cessino di essere entità discrete per formare un collettivo moralmente rilevante. Invece di cercare disperatamente un colpevole umano o di antropomorfizzare la macchina, dovremmo, secondo Conradie, considerare il collettivo ibrido come il vero agente cognitivo responsabile, risolvendo così il problema del vuoto di responsabilità senza forzature metafisiche.

A chiudere la prima giornata è stato Pekka Mäkelä (Università di Helsinki) con "Institutional Agents Facing Disruptive Technologies". Mäkelä ha operato una distinzione cruciale tra la "responsabilità dell'IA" (il tentativo di rendere la macchina un agente morale) e l'"AI responsabile". Criticando l'idea che si possa formalizzare la moralità all'interno del codice (riferendosi ai lavori di Alfred Mele sull'autonomia), ha difeso un approccio *istituzionale*. La risposta ai rischi delle tecnologie autonome non risiede nell'algoritmo, ma nella struttura sociale: servono regole costitutive che definiscano ruoli e compiti formali. È l'istituzione, in quanto agente

cognitivo collettivo, che deve assorbire e gestire la responsabilità attraverso un *design* normativo chiaro e vincolante.

La seconda giornata si è aperta con l'intervento congiunto di Irene Dominecale e Marco Emilio (Istituto Universitario Salesiano Venezia), "The Ambivalent Challenge of Tokenizing Collective Agency". I relatori hanno esplorato l'impatto delle tecnologie *blockchain* sull'ontologia sociale, adottando una prospettiva "non ideale". Il contributo ha analizzato come gli artefatti digitali (*token*, *smart contracts*) non siano semplici strumenti, ma strutturino nuove forme di agenzia collettiva. Dominecale ed Emilio hanno evidenziato come il *design* di questi artefatti influenzi le potenzialità operative disponibili per gli agenti cognitivi umani, sottolineando la necessità di gestire le asimmetrie di potere epistemico per generare fiducia in ambienti decentralizzati. La "tokenizzazione" appare quindi come un processo che ridefinisce i confini stessi dell'azione sociale.

Alper Güngör (McGill University) ha portato la discussione sul terreno dell'estetica con "AI art, collaboration, and appreciation". Güngör ha decostruito la nozione di autorialità nell'era generativa. Se l'IA manca degli stati mentali per essere un autore, e l'utente spesso fa troppo poco per esserlo appieno, chi è l'artista? Güngör ha proposto il concetto di "co-creazione", simile alla dinamica presente nei videogiochi. Ha suggerito che, anche senza una comunicazione diretta delle intenzioni, esiste una collaborazione ontologicamente rilevante tra *designer*, utente e sistema. In questa visione, anche la curatela del materiale di *training* diventa parte della storia causale dell'opera, essenziale per il suo apprezzamento estetico.

Clara Reidl-Reidenstein (Università di Oxford) ha presentato una provocatoria relazione dal titolo "Citizens of where? The

impossible contradiction of claiming jurisdiction over LLMs". La relatrice ha evidenziato un paradosso imminente: gli Stati sentiranno la necessità funzionale di rivendicare giurisdizione sui *Large Language Models* per proteggere i cittadini dai cattivi consigli (equiparabili a negligenza professionale). Tuttavia, le attuali teorie dello Stato si basano sulla territorialità e sull'incarnazione fisica degli agenti. Gli LLM, definiti come "agenti minimi" (intenzionali ma non morali) e privi di localizzazione fisica univoca, sfuggono a queste categorie. Reidl-Reidenstein ha concluso che o ripensiamo l'attribuzione di responsabilità o dobbiamo rivedere la condizione di *embodiment* (incarnazione) necessaria per la sovranità statale.

Un tono decisamente più scettico è stato adottato da Krzysztof Posłajko (Università Jagiellonian di Cracovia) in "Functionalism, alien minds, paperclip maximisers, and corporate responsibility ascriptions". Posłajko ha mostrato un *lato oscuro* del funzionalismo: se definiamo la mente in base al ruolo funzionale, le corporazioni potrebbero qualificarsi come agenti cognitivi, ma probabilmente risulterebbero essere "menti aliene". Utilizzando l'analogia con i cosiddetti *massimizzatori di graffette* (intelligenze artificiali che distruggono il mondo pur di massimizzare la produzione di graffette), ha suggerito che le corporazioni potrebbero avere strutture motivazionali rigide e "psicopatiche", prive della capacità di provare rimorso. Questo scenario ontologico renderebbe problematica, se non impossibile, l'attribuzione di una genuina responsabilità morale, spingendoci verso una visione finzionalista.

Il convegno si è concluso con la *keynote lecture* di Christian List (LMU Munich), intitolata "Can artificial agents have free will?". In un garbato ma fermo controcanto rispetto allo scetticismo emerso nelle sessioni precedenti (laddove

Posłajko aveva evocato lo spettro di “menti aliene” incapaci di agire morale e Kasar aveva lamentato l’impossibilità di tracciare una responsabilità in “azioni non intenzionali”), List ha proposto un ribaltamento di prospettiva. Adottando una visione naturalistica robusta, ha argomentato che il libero arbitrio si fonda su tre capacità funzionali: l’agenzia intenzionale, la possibilità di scegliere tra opzioni alternative e il controllo causale sulle proprie azioni. Se accettiamo che queste proprietà possano essere realizzate a un livello “agenziale” superiore, indipendentemente dal substrato fisico (neuroni o silicio), allora non vi sono ostacoli ontologici insormontabili per attribuire il libero arbitrio anche agli agenti artificiali avanzati o agli agenti di gruppo. Una conclusione che ha aperto molti interrogativi e che sembra confermare che la soluzione ai dilemmi etici dell’IA potrebbe richiedere non meno, ma *più* metafisica.