

Ripensare l'intelligenza artificiale: dall'autonomia alla simbiosi

Donato Malerba

Dipartimento di Informatica, Università degli Studi di Bari Aldo Moro

Abstract. In un momento storico in cui l'Intelligenza Artificiale (IA) sta ridefinendo in profondità il nostro rapporto con la tecnologia, si afferma con crescente rilevanza il paradigma dell'IA simbiotica: un modello fondato su forme di cooperazione autentiche, affidabili e adattive tra esseri umani e sistemi artificiali. Questo approccio supera la concezione dell'IA come semplice strumento autonomo e si avvicina a un'idea di agentività condivisa. Una tale visione, capace di rispondere in modo costruttivo a molte delle preoccupazioni odierne sull'IA, trova una concreta attuazione nello Spoke 6 del progetto FAIR (Future Artificial Intelligence Research), interamente dedicato all'intelligenza artificiale simbiotica.

Keywords. IA autonoma, IA simbiotica

Introduzione

Per decenni, lo sviluppo dell'Intelligenza Artificiale (IA) è stato guidato dall'ambizione di creare agenti autonomi, capaci di risolvere problemi in modo indipendente, senza necessità di interazione o supervisione umana. Non a caso, il paradigma emergente dell'IA agentica si riferisce a sistemi autonomi progettati per perseguire obiettivi complessi con un intervento umano minimo (Acharya et al., 2025). L'IA autonoma (Fig. 1) ha prodotto risultati straordinari, e i veicoli a guida autonoma rappresentano una delle applicazioni più visibili con sistemi in grado di percepire l'ambiente circostante e di navigare senza input umano (Van Brummelen et al., 2018). Questo orientamento verso l'autonomia è rafforzato anche dai modelli prevalenti dell'IA: l'apprendimento supervisionato mira a generalizzare da esempi, senza necessità di ulteriori istruzioni, il reinforcement learning si fonda sull'autoapprendimento attraverso l'interazione con l'ambiente, e le architetture agent-based sono concepite per operare in modo indipendente in contesti complessi. La visione cartesiana della mente come calcolo logico, separata dal corpo e dal contesto, ha fornito una giustificazione epistemologica al paradigma dell'IA autonoma. A ciò si aggiungono motivazioni industriali altrettanto forti, grazie alla promessa di risparmi operativi e di soluzioni che sono scalabili e replicabili su larga scala.

1. IA Autonoma: i limiti

Tuttavia, oggi è sempre più evidente che l'IA autonoma presenta limiti strutturali, teorici ed etici. Anzitutto, l'autonomia totale si rivela spesso un'illusione: anche i sistemi più avanzati dipendono da dati etichettati da esseri umani, e da scelte che riflettono valori,

priorità e vincoli umani. L'intervento umano è dunque implicito, sebbene spesso invisibile, in ogni fase dello sviluppo e dell'impiego di questi sistemi. Inoltre, la rincorsa all'autonomia ha spesso trascurato il potenziale trasformativo della collaborazione tra esseri umani e sistemi intelligenti. Questa viene ostacolata dall'opacità dei modelli appresi dai dati, che rende difficile comprenderne il funzionamento interno e verificarne le decisioni o attribuire responsabilità in caso di errore. L'esclusione degli esseri umani dal processo decisionale impedisce un loro intervento in situazioni di rischio con eventuale correzione delle decisioni prese dalla macchina. L'IA autonoma è spesso fragile quando si confronta con situazioni nuove o non previste in fase di addestramento. Peraltro, anche quando l'IA funziona correttamente dal punto di vista tecnico, può produrre risultati in contrasto con norme sociali, principi etici o bisogni concreti delle persone. Infine, all'IA autonoma manca una competenza essenziale nelle interazioni tra esseri umani: la capacità di negoziare significati, di comprendere ambiguità o di interpretare situazioni aperte e sfumate.

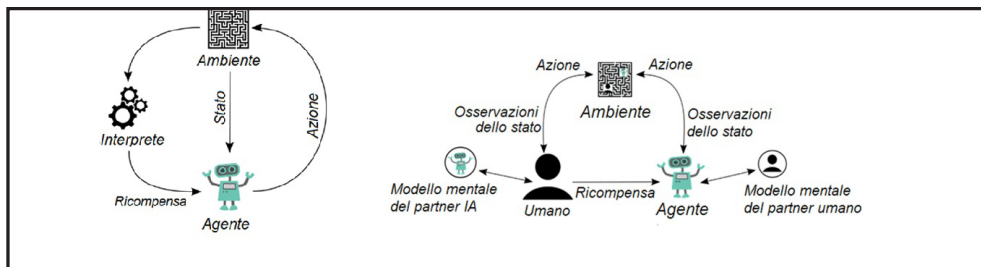


Fig. 1

A sinistra: l'IA autonoma, in cui un agente artificiale opera da solo per raggiungere il proprio obiettivo in modo efficace e accurato. L'agente percepisce lo stato dell'ambiente, agisce in autonomia e riceve una ricompensa in base alle sue prestazioni. A destra: l'IA simbiotica, in cui l'agente artificiale collabora con esseri umani, condividendo con loro un obiettivo. Grazie alla capacità di ragionare sulle azioni e sugli stati mentali delle persone, l'agente supporta decisioni condivise. I partner umani, a loro volta, comprendono il processo decisionale dell'IA e ne interpretano scopi e vincoli.

2. Un'alternativa: l'intelligenza artificiale simbiotica

Il paradigma dell'IA simbiotica (Fig. 1) propone un cambio di prospettiva radicale: anziché sostituire l'essere umano, la macchina ne potenzia le capacità cognitive, lo assiste in modo adattivo. Si tratta di costruire squadre miste di intelligenze umane e artificiali, in cui entrambe apprendono e si adattano reciprocamente. L'obiettivo non è solo raggiungere un risultato, ma farlo in modo tale che tutte le parti coinvolte comprendano le ragioni e i vincoli delle scelte compiute. Ciò significa che le sfide non si limitano alla risoluzione del compito, ma si estendono alla qualità dell'interazione e alla comprensione reciproca tra i membri della squadra eterogenea. Affinché tale collaborazione risulti efficace, è essenziale che l'agente artificiale sia in grado di comprendere e ragionare sulle azioni umane, tenendo conto degli stati mentali, delle intenzioni e delle conoscenze pregresse delle persone, adottando, cioè, una forma di *theory of mind* computazionale. Al contempo, è necessario che l'essere umano possa accedere e interpretare in modo trasparente il processo decisionale dell'IA, comprendendone motivazioni, vincoli e finalità. Solo attraverso questa bidirezio-

nalità cognitiva – la capacità dell’IA di adattarsi al modello mentale dell’utente e viceversa – è possibile costruire forme di cooperazione autentiche, affidabili e adattive, superando la visione dell’IA come semplice strumento e avvicinandosi a un modello di agentività condivisa.

Un esempio illuminante proviene dal mondo degli scacchi. Nella variante chiamata *freestyle chess*, il giocatore può consultare l’IA durante la partita. Sorprendentemente, si è osservato che le squadre vincenti non sono né gli esseri umani né le macchine da sole, ma la combinazione delle due (De Cremer, Kasparov 2021).

Il crescente impatto dell’IA nella vita quotidiana sta spostando il focus della ricerca dalle capacità tecniche delle macchine alla qualità della relazione uomo-IA. In questo scenario, il paradigma dell’IA simbiotica non rappresenta solo una prospettiva futura, ma una risposta necessaria alle dinamiche psicologiche, sociali ed etiche emerse nell’evoluzione dell’interazione con le tecnologie intelligenti. Come evidenziato in recenti studi (Prescott, 2024), le persone tendono a proiettare caratteristiche umane sugli agenti artificiali – un fenomeno noto come *antropomorfismo* – dando vita a interazioni che assomigliano sempre più a relazioni sociali, con effetti tangibili sul benessere individuale. Tutto ciò ha senso solo se l’IA non è concepita come un sostituto dell’essere umano, ma come un partner in grado di co-evolvere con lui, adattandosi ai suoi bisogni, emozioni e valori. In questo senso si inquadra anche una visione costruttivista del rapporto “simbiotico” tra esseri umani e sistemi artificiali (Carnevale et al., 2024).

3. Il progetto FAIR e lo Spoke 6: Symbiotic AI

Nell’ambito del progetto PNRR FAIR – Future Artificial Intelligence Research (<https://fondazione-fair.it/>), lo Spoke 6 affronta il tema della IA simbiotica, individuando cinque sfide fondamentali da affrontare per la realizzazione di sistemi davvero cooperativi tra esseri umani e macchine:

1. Design umano-centrico. I sistemi devono essere progettati per interagire con le persone, non solo per risolvere compiti. Questo implica la considerazione di emozioni, intenzioni, bisogni e modelli mentali degli utenti. È necessaria una progettazione orientata all’esperienza, che integri principi dell’interazione persona-macchina, dell’IA emotiva e dei modelli cognitivi del comportamento.
2. Comprensione dell’essere umano. Un sistema simbiotico deve saper interpretare chi ha di fronte, comprendendo testi, gesti, emozioni e intenzioni. L’IA deve acquisire competenze percettive e interpretative su base multimodale – testo, audio, video – per adattarsi in modo flessibile a contesti reali. Tecnologie chiave includono il natural language processing, il deep learning e l’analisi di dati eterogenei.
3. Input umano come risorsa. L’interazione non è un ostacolo, ma un’opportunità. L’IA deve essere in grado di integrare input provenienti dagli esseri umani, anche se incompleti o incerti. Servono approcci ibridi in cui i modelli statistici si combinano con rappresentazioni simboliche, grafi della conoscenza, inferenza e semantica.
4. Fiducia e trasparenza. La cooperazione richiede fiducia. Le decisioni dell’IA devono essere

comprensibili, spiegabili, accettabili e sostenibili. Oltre allo sviluppo di tecniche di XAI (eXplainable AI), è fondamentale integrare valori etici, principi giuridici e considerazioni ambientali e sociali. L'IA simbiotica deve essere progettata per l'empowerment, non per la sostituzione dell'essere umano.

5. Infrastrutture e scalabilità. Infine, la sfida tecnologica. Come sviluppare sistemi di IA simbiotica che siano efficienti, scalabili e distribuiti? Occorrono soluzioni capaci di operare su dati grandi ed eterogenei, anche in ambienti decentralizzati (edge computing), senza compromettere le capacità collaborative e adattive.

Con il coinvolgimento di sei università, un ente di ricerca nazionale e tredici imprese, lo Spoke 6 di FAIR ha avviato numerose ricerche dedicate all'IA simbiotica. Sono stati finanziati diciannove progetti su temi strettamente correlati, accomunati da una visione condivisa: non un'IA che lavori al posto dell'essere umano, ma insieme a lui. Una co-evoluzione possibile, secondo il paradigma della simbiosi.

4. Conclusioni

La sfida dell'IA, oggi, non è più imitare l'intelligenza umana, ma imparare a convivere con essa. Occorre costruire sistemi che crescano insieme agli esseri umani, anziché svilupparsi in completa autonomia. L'IA del futuro deve essere simbiotica, altrimenti rischia di risultare disallineata rispetto ai valori, ai bisogni e alle aspettative delle persone. Questa prospettiva rappresenta una risposta concreta a molte delle attuali preoccupazioni legate all'IA: può attenuare il timore della sostituzione occupazionale, ridurre bias e discriminazioni, rafforzare trasparenza e controllo, migliorare la tutela della privacy e della sicurezza, e infine promuovere un'innovazione realmente responsabile. Per realizzare tutto questo, è necessario affrontare numerose sfide di ricerca, come quelle già intraprese dallo Spoke 6, interamente dedicato all'IA simbiotica all'interno del progetto FAIR.

Riferimenti bibliografici

Deepak Bhaskar Acharya, Karthigeyan Kuppan, B. Divya (2025), Agentic AI: Autonomous Intelligence for Complex Goals—A Comprehensive Survey, *IEEE Access* (vol. 13), pp. 18912-18936.

Antonio Carnevale, Antonio Lombardi, Francesca A. Lisi (2024), A human-centred approach to symbiotic AI: Questioning the ethical and conceptual foundation, *Intelligenza Artificiale* (vol. 18, n.1), pp. 9-20

David De Cremer, Garry Kasparov (2021), AI should augment human intelligence, not replace it, *Harvard Business Review* (18 marzo).

Tony Prescott (2024), *The psychology of artificial intelligence*, Routledge, London.

Jessica Van Brummelen, Marie O'Brien, Dominique Gruyer, Homayoun Najjaran (2018), Autonomous vehicle perception: The technology of today and tomorrow, *Transportation Research Part C: Emerging Technologies* (vol. 89, aprile 2018), pp. 384-406.

Autore



Donato Malerba donato.malerba@uniba.it

Professore ordinario di Sistemi di elaborazione delle Informazioni presso l'Università di Bari. Direttore del Dipartimento di Informatica (2015-2022) e del CINI National Lab on Big Data (2014-2021). Componente del Consiglio Direttivo della Big Data Value Association e della PPP Big Data dell'UE (2015-2016). Responsabile scientifico dello Spoke 6 – Symbiotic AI – del progetto FAIR (Future Artificial Intelligence Research). I suoi interessi di ricerca riguardano la data science e l'intelligenza artificiale.