
aracne

Classificazione Decimale Dewey

— 006.30112 (23.) INTELLIGENZA ARTIFICIALE. Previsioni

Thema

— Soggetto: JPS. Relazioni internazionali
— Qualificatori: 3MR. XXI secolo, 2000-2100

ANTONIO FREDDI

L'INTELLIGENZA ARTIFICIALE COME FUTURO ATTORE CENTRALE E AUTONOMO DELLA POLITICA MONDIALE?

aracne



©

ISBN
979–12–218–2974–7

DOI
10.53136/97912218297 47

PRIMA EDIZIONE
ROMA 10 NOVEMBRE 2026

L’essenza della tecnica non è nulla di tecnico.
M. Heidegger, *La questione della tecnica*

A Lucrezia,
che è il futuro

INDICE

11 *Introduzione*

21 Capitolo I
Considerazioni teoriche e preliminari
1.1. L'importanza dello sguardo al futuro non prossimo nelle relazioni internazionali, 21 – 1.2. Oltre l'antropocentrismo e l'eccezionalismo umano, 25 – 1.3. Alcune nozioni basilari e considerazioni preliminari sull'intelligenza artificiale, 32.

45 Capitolo II
L'intelligenza artificiale nelle relazioni internazionali ora e nel breve termine
2.1. Impieghi attuali e nel breve termine dell'IA nelle relazioni internazionali, 45 – 2.2. Effetti e conseguenze nel sistema internazionale nel breve termine, 61 – 2.3. Possibile miglioramento delle teorie delle relazioni internazionali?, 74 – 2.4. Attuali gap teorici ed epistemici per il lungo termine, 78.

83 Capitolo III
L'intelligenza artificiale attore delle relazioni internazionali nel futuro a lungo termine. Scenari

3.1. L'IA diventerà vero e proprio attore autonomo della scena internazionale, 83 – 3.2. Scenari, variabili e dilemmi possibili, 105 – 3.2.1. *Scenario 1: Incrementalismo algoritmico e coesistenza*, 108 – 3.2.2. *Scenario 2: Attore autonomo al "servizio" ma con delega estrema*, 116 – 3.2.3. *Scenario 3: Attore senza sostituirsi all'umano ma condizionandolo completamente*, 125 – 3.2.4. *Scenario 4: Attore totalmente autonomo, indipendente e competitivo*, 132 – 3.2.5. *Scenario 5: Ibridizzazioni essere umano–IA*, 141 – 3.2.6. *Scenario 6: IA attore di armonizzazione*, 143 – 3.2.7. *Altri possibili scenari*, 151.

159 Capitolo IV
Rivoluzione sistemica, ontologica e concettuale

4.1. Nuovi sistemi e riconfigurazioni globali, 159 – 4.2. Nuova ontologia e variazioni concettuali, 164 – 4.3. Azioni preventive, 181.

193 *Conclusioni*

213 *Bibliografia*

229 *Sitografia*

INTRODUZIONE

L'intelligenza artificiale è ormai entrata nella politica mondiale e nelle relazioni interazionali. Non è più soltanto una tecnologia emergente, un settore industriale avanzato o un tema specialistico riservato all'informatica: sta diventando un fattore di potenza, una risorsa strategica, un campo di competizione economica e militare, una questione di governance globale e, soprattutto, una forza destinata a incidere sulle categorie fondamentali attraverso cui comprendiamo le relazioni internazionali. Gli stati la impiegano per analizzare grandi quantità di dati, rafforzare l'intelligence, proteggere o attaccare infrastrutture digitali, migliorare la logistica militare, monitorare flussi economici e sociali, produrre contenuti di diplomazia, ottimizzare decisioni amministrative e accrescere la propria capacità di previsione. Assumono una rilevanza geopolitica sempre più evidente i modelli, i *cloud*, i *chip*, le piattaforme e gli insiemi di dati controllati da imprese tecnologiche. Le organizzazioni internazionali già discutono di norme, standard, forme di audit e principi di responsabilità. Questa

crescente presenza dell'intelligenza artificiale (nel seguito IA) nelle pratiche ordinarie della politica globale impone una domanda radicale: siamo davvero di fronte soltanto a un nuovo strumento nelle mani degli attori tradizionali oppure stiamo assistendo all'emersione di una nuova forma di *agency* internazionale?

Questo lavoro di ricerca nasce da tale domanda e prova a spingerla ben oltre l'orizzonte del dibattito più immediato. Molte analisi attuali si concentrano sugli effetti dell'IA nel breve periodo: sicurezza, cyberspazio, armi autonome, disinformazione, efficienza amministrativa, competizione tra Stati Uniti e Cina, regolamenti europei, rischio di concentrazione del potere nelle grandi piattaforme. Sono temi decisivi, ma non esauriscono la questione. L'IA è una tecnologia dal carattere cumulativo, accelerato e potenzialmente trasformativo: ciò che oggi appare come supporto decisionale potrebbe domani diventare una mediazione obbligata per la decisione; ciò che oggi è automazione di un processo potrebbe trasformarsi in autonomia strategica; ciò che oggi è infrastruttura potrebbe acquisire funzioni sovrane o quasi. Per questa ragione questo scritto adotta un orizzonte di lungo periodo, superiore ai cinquant'anni, non con l'ambizione ingenua di prevedere esattamente il futuro, ma con l'obiettivo di esplorare futuri possibili, individuare discontinuità e mettere alla prova le categorie teoriche con cui la disciplina delle relazioni internazionali interpreta i mutamenti immaginabili.

L'ipotesi centrale è che l'intelligenza artificiale, o più probabilmente una pluralità di sistemi, reti e agenti artificiali basati su di essa, possa diventare nel lungo termine un attore centrale e autonomo della politica mondiale. Ovviamente ciò non implica necessariamente che l'IA

debba possedere coscienza, interiorità, intenzionalità morale o personalità nel senso umano del termine, come tra l'altro non sono possedute da attori come le multinazionali o le organizzazioni internazionali. In questa sede non interessa affatto stabilire se una macchina possa arrivare a essere definita una persona come un essere umano, ma se possa diventare attore in senso internazionalistico: un centro di elaborazione, decisione e azione capace di perseguire obiettivi, interagire strategicamente con altri soggetti, produrre effetti materiali e normativi, condizionare la distribuzione del potere e partecipare alla costruzione dell'ordine (o disordine) internazionale. In questa prospettiva, l'attorialità viene definita a partire dalla capacità di incidere stabilmente sui processi della politica mondiale.

Da questa ipotesi discendono più precise domande di ricerca. In che modo l'IA sta già modificando le funzioni degli attori internazionali esistenti? Quali elementi teorici consentono di pensare l'IA non solo come mezzo tecnico, ma come possibile attore non umano? Attraverso quali traiettorie l'IA potrebbe acquisire autonomia, indipendenza e capacità di condizionamento rispetto agli stati o attori internazionali che l'hanno prodotta? Quali forme di ordine internazionale potrebbero emergere da tale trasformazione? Quali concetti fondamentali delle relazioni internazionali verrebbero messi in crisi o ridefiniti dall'emersione dell'IA come attore autonomo e centrale?

Questo scritto non sostiene che l'esito ipotizzato, cioè il raggiungimento dello status di attore autonomo o addirittura centrale delle relazioni internazionali da parte dell'IA, sia un futuro certo, ma solo un futuro possibile, magari probabile e, soprattutto, di portata tale da meritare un'investigazione analitica tempestiva. L'incertezza è parte della

questione ma anche un'opportunità. Ignoriamo se, quando e in quale forma si realizzeranno e come si comporteranno i sistemi di intelligenza artificiale generale (AGI) o superintelligenza (ASI); non sappiamo se saranno allineati agli interessi umani, se diverranno incontrollabili e quanto, se saranno incorporati negli stati o si emanciperanno da essi. L'incertezza rende necessario e allo stesso tempo plausibile riflettere attraverso l'elaborazione teorica di scenari possibili. Tra l'altro, è opportuno considerare che, in un contesto in cui le istituzioni e le norme internazionali si adattano lentamente mentre la tecnologia evolve rapidamente, rinviare l'analisi e la preparazione fino al momento in cui la trasformazione sarà evidente significherebbe arrivare tardi.

Tra i presupposti teorici di questo elaborato si possono annoverare gli studi sul futuro e lo *strategic foresight*, che negli ultimi anni hanno assunto crescente rilievo presso istituzioni come la Commissione europea, le Nazioni Unite, l'OCSE e altri attori internazionali. Vari autori, come Bilgin, Hanson, Neumann e Øverland, Miller, Minkkinen e Tõnurist, suggeriscono di intendere il futuro (anche a medio e lungo termine) come spazio di esplorazione metodica, utile a testare teorie, opzioni politiche, vulnerabilità e capacità adattive, evitando di irrigidirlo attraverso presuntuose previsioni puntuali. Tale prospettiva di forte sbilanciamento creativo o ideativo verso il futuro può essere collocata anche entro una più ampia tradizione filosofica che considera l'avvenire come dimensione costitutiva e sostanziale dell'agire e, soprattutto, del pensare umani. Come osserva, per esempio, Ernst Bloch nel diciassettesimo capitolo di *Il principio speranza*, l'uomo, essere incompiuto, è chiamato a pensare il futuro in un mondo altrettanto incompiuto:

il presente non esaurisce il reale, che non coincide soltanto con ciò che già è, ma anche con ciò che può ancora essere. L'uomo autenticamente umano è quindi colui che sa cogliere nel presente i segni di un compimento ancora aperto. In Hans Jonas, soprattutto in *Il principio responsabilità. Un'etica per la civiltà tecnologica*, l'anticipazione del futuro assume invece un valore morale: la civiltà tecnica ha ampliato a tal punto la portata dell'azione umana da obbligare a considerarne gli effetti remoti sulle generazioni future. In modo ancora più drammatico. Günther Anders, in *L'uomo è antiquato. Vol. I. Considerazioni sull'anima nell'epoca della seconda rivoluzione industriale*, parlando del "dislivello prometeico", ha, infine, mostrato come la tecnica produca conseguenze spesso superiori alla nostra normale capacità immaginativa, rendendo necessaria un'immaginazione anticipatrice anche del peggio. Per queste ragioni adottare l'orizzonte lungo è una forma di responsabilità teorica e politica, perché permette di osservare trasformazioni strutturali e ontologiche che nel breve periodo possono risultare invisibili: per esempio, mutamenti nella forma della sovranità e dell'*agency*, nella distribuzione delle capacità, nella natura degli attori internazionali, nei presupposti e nei significati della decisione, nel modo di intendere il ruolo delle informazioni.

Si sono prese in considerazione anche le prospettive post–antropocentriche e post–umaniste le quali, con autori come Bennett, Braidotti, Connolly, Coole, Cudworth e Hobden, Hayles e Latour, tra gli altri, mostrano che l'azione sociale e politica non può essere ricondotta esclusivamente alla volontà umana. Infrastrutture, reti tecniche, ecosistemi, dispositivi materiali e attanti non umani partecipano già ora alla produzione degli esiti politici. Questa

cornice è essenziale per evitare di valutare la capacità attoriale dell'IA solo in base a quanto essa assomigli all'essere umano. Dato che le relazioni internazionali sono sempre state attraversate da forze non umane (geografia, clima, energia, infrastrutture, tecnologie, mercati, sistemi logistici…), allora l'IA non arriverebbe in un vuoto ontologico, ma come la forma più evidente e operativa di una distribuzione dell'agency post–umana.

Attenzione è dedicata anche alla letteratura tecnica sull'IA, alla sua evoluzione e ai rischi connessi. Nozioni come *machine learning*, *deep learning*, apprendimento per rinforzo, *agentic AI*, AGI e ASI hanno permesso di chiarirne sufficientemente le caratteristiche, in particolare di distinguere l'IA specialistica oggi diffusa dalle ipotesi di futuri sistemi generali o superintelligenti. Altman, Baum, Doshi–Velez e Kim, Goodfellow, Grace, Lipton, Mitchell, Salmon, Sutton e Barto sono alcuni degli autori che hanno aiutato a comprendere il significato di opacità, interpretabilità, *misalignment*, comportamenti inattesi, autonomia e problema del controllo, e dedurne le possibili conseguenze in ambito politico–internazionale: che cosa diventa l'IA quando le sue capacità entrano nei processi che producono diplomazia, sicurezza, ordine, conflitto, cooperazione e potere?

Infine, un importante fonte di riferimento è stata la letteratura, prodotta in anni recenti, sul rapporto tra IA e relazioni internazionali. Tramite gli scritti di numerosi studiosi (come Allen, Bode, Erskine, Dafoe, Horowitz, Kiggins, Maas, Melouni ed Efthymiou, Pruet, Scharre, Tallberg e Veale, per fare alcuni nomi) si è potuto collocare l'IA dentro ambiti specifici delle relazioni internazionali, anche se prevalentemente in rapporto al presente: sicurezza

internazionale, governance globale, competizione tecnologica, ruolo delle imprese, trasformazione degli apparati pubblici, *synthetic agency*, crisi delle categorie tradizionali, armi autonome. Una parte consistente di questa letteratura, tuttavia, tende ancora a trattare l'IA solo come strumento, moltiplicatore di potenza o oggetto di regolazione. Il presente lavoro prova invece a compiere un passo ulteriore: interrogarsi sull'eventualità che l'IA in futuro non sia soltanto qualcosa da usare o governare, ma un diretto partecipante al governo del sistema internazionale.

Da qui deriva anche la natura peculiare di questo elaborato: non si tratta di una ricerca empirica, né di una previsione tecnologica, né tantomeno di una teoria sull'intelligenza artificiale. Si tratta di un esercizio teorico–prospettico fondato su letteratura esistente, segnali presenti, analogie storiche, ideazione e analisi di scenari futuri, riflessione concettuale. Il suo obiettivo è rendere plausibile una domanda ancora marginale: che cosa accadrebbe alle relazioni internazionali se l'IA smettesse di essere soltanto un mezzo e diventasse una delle entità che decidono, interpretano, negoziano, coordinano, minacciano, proteggono o trasformano l'ordine mondiale?

Il primo capitolo costruisce le premesse teoriche. Giustifica lo sguardo rivolto al futuro lontano, introduce il superamento dell'antropocentrismo come condizione concettuale per pensare l'IA come possibile attore e chiarisce alcune nozioni tecnologiche di base: IA specialistica, *agentic AI*, AGI, ASI, opacità, interpretabilità, affidabilità, comportamenti ingannevoli e incertezza sullo sviluppo futuro.

Il secondo capitolo analizza l'IA nelle relazioni internazionali attuali e nel breve termine. Essa appare ancora prevalentemente come strumento incorporato negli attori

esistenti, ma già capace di trasformarne funzioni e capacità. Vengono esaminati i suoi impieghi nell'*intelligence*, nella cybersicurezza, nella diplomazia, nella geoeconomia, nella governance interna e nel settore militare, mettendo in luce gli albori di una trasformazione che non è soltanto quantitativa. Il capitolo affronta inoltre il possibile impatto dell'IA sulle teorie delle relazioni internazionali e individua alcuni gap teorici ed epistemici dell'attuale dibattito.

Il terzo e principale capitolo costituisce il nucleo prospettico del lavoro, perché sviluppa l'ipotesi che, nel lungo termine, sistemi di IA avanzate (AGI o ASI) diventino attori autonomi della scena internazionale: in che modo, svolgendo quali attività, producendo quali effetti, tenendo quali comportamenti, stabilendo quali influenze? Su questa base vengono delineati alcuni possibili scenari futuri della scena politica globale: incrementalismo algoritmico e coesistenza tra attori internazionali umani e sistemi IA; delega estrema all'IA; sistemi IA capaci di condizionare profondamente e agire da "ghost actors"; scenari in cui sistemi IA sono veri e propri attori internazionali autonomi, indipendenti e competitivi; diffusa ibridazione uomo–IA; IA come attore di armonizzazione globale. Vengono inoltre evocati ulteriori scenari, tra cui ordini post–territoriali, neofeudali, postcoloniali o inter–cognitivi.

Il quarto capitolo è altrettanto fondamentale perché trae alcune conseguenze teoriche dagli scenari appena delineati. Se l'IA diventasse attore centrale, la rivoluzione nelle relazioni internazionali sarebbe sistemica, ontologica e concettuale; forse anche epocale. Vengono analizzati i cambiamenti incontro a cui andrebbero alcune entità e concetti: come muterebbe ciò che si considera attore o *agency* e struttura? Che ne sarebbe delle idee di sovranità e

potere, dell'attribuzione della razionalità e della responsabilità, del ruolo della temporalità? Come si formerebbero gli obbiettivi? Che cosa verrebbe considerato conflitto? Da ultimo, il capitolo propone brevemente alcune azioni preventive: dopotutto, uno degli scopi delle analisi sui futuri possibili è cercare di non trovarsi impreparati quando uno di questi "possibili" diventa realtà e, possibilmente, anche di influenzarne l'evoluzione.

Il filo conduttore di questo scritto è dunque una domanda forse destabilizzante nelle sue implicazioni: che cosa accadrà alla politica mondiale e alle relazioni internazionali se e quando l'*agency*, la decisione, e forse anche l'intelligenza, non saranno più monopolio dell'essere umano e delle sue istituzioni? La risposta non sarà univoca e chiaramente non potrebbe essere tale, trattandosi di un'analisi predittiva e possibilistica. Può darsi che l'IA resti un'infrastruttura subordinata, per quanto potente; può darsi che diventi il centro effettivo di decisioni formalmente umane; può darsi che si fonda con l'umano; può darsi che si separi dagli attori umani che l'hanno generata; può darsi, infine, che diventi una risorsa di coordinamento globale più efficace delle istituzioni esistenti. In tutti questi casi, però, come si vedrà, un dato resterebbe comune: l'IA costringerebbe le relazioni internazionali a ripensarsi oltre il loro impianto antropocentrico, stato–centrico e territorialista. La politica mondiale non sarebbe più soltanto il teatro di comunità umane organizzate, ma un campo popolato da attori biologici, istituzionali, tecnici, ibridi e artificiali. Studiare questa possibilità significa prendere sul serio la portata politica di una tecnologia che, già oggi, sta entrando nei luoghi in cui il mondo viene osservato, classificato, deciso e governato.