AGENTI DIGITALI: QUALE COOPERAZIONE CON L'UMANO?

https://www.quotidianosanita.it/lettere-al-direttore/agenti-digitali-quale-cooperazione-con-lumano/

Lucio Romano *Centro Interuniversitario di Ricerca Bioetica (CIRB)*

02 Luglio 2026

Gentile Direttore, nei processi decisionali si rileva un sempre maggiore sviluppo e ricorso agli *Agenti digitali* (*Agenti AI*), sistemi in grado di raggiungere un obiettivo specifico con una supervisione limitata. Completano attività, pianificano e memorizzano. Progettati per operare con un certo grado di autonomia, combinando molteplici funzionalità, incluse capacità cognitive avanzate.

Se i sistemi di IA Generativa generano contenuti rispondendo (*output*) a una domanda (*prompt*), gli *Agenti AI* agiscono: prenotano, negoziano, filtrano informazioni, redigono atti, eseguono codici, selezionano candidati, moderano contenuti, gestiscono comunicazioni.

Come pubblicato su Nature Biotechnology l'ambito della ricerca biomedica, ad esempio, è il settore dove il ricorso agli *Agenti AI* è sempre più diffuso.

Si configurano come esperti computazionali capaci di affiancare e in taluni casi competere in attività ad elevata intensità cognitiva e operativa quali la revisione della letteratura scientifica, la formulazione di ipotesi, l'analisi dei dati e l'interpretazione dei modelli. Non secondaria l'applicazione nella ricerca biomedica, l'analisi dei dati clinici e l'identificazione di biomarcatori. In particolare, il ricorso agli *Agenti AI* contribuisce a scalare il processo decisionale clinico e l'esecuzione oltre la capacità umana, rendendoli particolarmente utili per richieste di "cura" orientate alla personalizzazione, alla precisione e alla predittività.

A fronte di sempre più diverse utilizzazioni degli *Agenti AI*, si pongono diversi interrogativi. Basti pensare a quanto, con l'aumentare del coinvolgimento degli *Agenti AI* nei processi decisionali davvero rilevanti (es. assunzione del personale, cure ai pazienti, erogazione di sovvenzioni, etc.), non sia possibile pensare il loro agire indipendentemente dalle questioni epistemologiche, etiche e giuridiche che esso solleva.

Quanto possiamo davvero sapere (*explicability*) dentro questi sistemi? Chi risponde moralmente delle loro conseguenze e come si costruisce un'*accountability* credibile? Che cosa comporta per attività che abbiamo sempre ritenuto di esclusiva pertinenza umana?

Trattare di *Agenti AI* significa, necessariamente, affrontare la questione della *agency* (*agentività*). Termine, ormai largamente citato, da intendere come capacità di agire con intenzionalità, consapevolezza, finalizzazione sulla base di una volontà. Con produzione di effetti sulla base di stati interni piuttosto che come mera reazione meccanica a stimoli esterni. Possiamo dire che rappresenta un concetto cardine, non senza controversie. Al riguardo si individuano due impostazioni contrapposte: la prima colloca l'*agency* nella dimensione esclusivamente umana; la seconda sostiene una concezione estensiva, suscettibile di includere anche enti di natura non umana come i sistemi artificiali.

Per questo motivo è rilevante richiamare “la visione alternativa” sviluppata da Luciano Floridi secondo cui “l’IA andrebbe compresa correttamente come una nuova forma di *agency priva di intelligenza*”.

Si rappresentano due orientamenti dell’*agenc*y. Un primo orientamento è volto “ad ampliare l’attuale concezione di intelligenza per includere anche le sue forme artificiali (tesi della *realizzabilità artificiale dell’intelligenza*, o RAI)”. Un secondo, invece, significa “estendere la comprensione dell’*agency* per includere una pluralità di forme, comprese quelle artificiali, che non presuppongono cognizione, intelligenza, intenzionalità o stati mentali (tesi della *molteplice realizzabilità dell’agency*, o MRA)”.

Supportando la tesi della MRA, si descrivono varie forme di *agency*: naturale, biologica, animale, sociale, artefattuale, umana individuale, umana sociale e, appunto, artificiale. Con quest’ultima, ed ecco l’aspetto interessante, “l’IA è compresa più correttamente come una nuova forma di *agency* che opera attraverso l’elaborazione computazionale e dei dati, che eccelle in compiti specifici ma manca dell’intelligenza, dell’intenzionalità e dell’autodeterminazione che incontriamo nei sistemi biologici”.

Ricorrendo alla tesi della molteplice realizzabilità dell’*agency* erimuovendo così la ricorrente interpretazione antropomorfica dell’IA, l’*agency artificiale* si rappresenta come un’efficace integrazione alle capacità proprie degli “agenti umani”.

A partire dai criteri di interattività, autonomia e adattabilità, ecco alcuni esempi che specificano le differenze tra le diverse *agency*.

“L’*agency naturale* opera in base a leggi fisiche e costanti, producendo pattern coerenti ma privi di autonomia e adattabilità. L’*agency biologica*, sia individuale che sociale, funziona attraverso la programmazione genetica, l’elaborazione sensoriale e la regolazione comportamentale, consentendo comportamenti autonomi e adattativi che sono molto meno prevedibili, anche entro i confini peculiari della specie. L’*agency umana* combina in modo unico i meccanismi biologici con l’elaborazione culturale e simbolica, dando luogo a una flessibilità e a una creatività senza pari. Mentre gli *agenti artificiali* apprendono attraverso processi di ottimizzazione algoritmica”.

Si delinea, ed ecco il passaggio significativo, una progressione “da forme più semplici a forme più complesse di *agency* che non implica una rigida gerarchia lineare, ma piuttosto un modello ramificato, con ogni forma di *agency* che mostra caratteristiche peculiari e vantaggiose in contesti specifici”.

In questo contesto l’IA rappresenta una forma di *agency* a sé stante, definita da obiettivi programmabili, capacità di adattamento basate sui dati e funzioni distribuite. Diversamente da quella umana, è certamente priva di coscienza, intenzionalità ma la sua precisione, scalabilità e riproducibilità le permettono di eccellere in domini specifici e ben delimitati.

Allora, quale rapporto tra *agency umana* e *agency artificiale*? In sintesi, si ribadisce che mentre “l’*agency artificiale* rappresenta una nuova forma di *agency* che emerge dall’interazione tra obiettivi programmati e comportamenti appresi, l’*agency umana* è l’unica che modella tutte le altre forme di *agency*”.

Sulla base di tali argomentazioni si possono richiamare alcuni principi chiave dell'etica dell'IA: complementarità, cooperazione e governance.

"Il futuro dell'*agency artificiale*", sottolinea Floridi, "non risiede nella capacità di replicare o superare una qualunque forma di intelligenza biologica, bensì nello sviluppo delle sue specifiche potenzialità, assicurando al tempo stesso un allineamento coerente con i valori umani e con le esigenze sociali e ambientali. Ciò richiede un equilibrio attento tra innovazione tecnologica e riflessione etica, sostenuto da un solido impianto teorico per comprendere l'evoluzione dell'IA e da strumenti concreti di *governance* per orientare la regolazione".

Mentre il dibattito pubblico si concentra sull'efficienza degli *Agenti AI*, resta sullo sfondo una ulteriore e insidiosa questione: il loro impatto sui meccanismi della democrazia. Perché non si tratta soltanto di automazioni ma di trasformazioni progressive di come si formano decisioni che riguardano la collettività.

Il primo rischio è quello della **opacità decisionale**. Quando, ad esempio, un *Agente AI* definisce l'accesso a servizi pubblici, l'allocazione di risorse, l'accesso alle cura o persino l'orientamento delle informazioni, il cittadino perde la possibilità di comprendere o contestare i criteri sottostanti. La trasparenza, condizione minima della responsabilità democratica, si assottiglia.

Il secondo rischio riguarda la **concentrazione del potere computazionale**. Gli *Agenti AI* più sofisticati sono sviluppati e controllati da un numero ristretto di "attori economici e tecnologici". Questo squilibrio rischia di trasferire capacità decisionali strategiche fuori dal perimetro della sovranità democratica, in assenza di adeguati strumenti di governance pubblica. Come richiama Leone XIV nella Lettera enciclica *Magnifica humanitas*, "quando un potere di tale portata si concentra in poche mani, tende a farsi opaco e a sfuggire al controllo pubblico, e cresce il rischio di uno sviluppo distorto che genera nuove dipendenze, esclusioni, manipolazioni e disuguaglianze".

Vi è poi il tema della **manipolazione del discorso pubblico. S**istemi capaci di generare e diffondere contenuti su larga scala possono alterare la qualità del confronto democratico, amplificando polarizzazione o disinformazione in modo difficilmente tracciabile.

Infine, la **responsabilità giuridica**resta un nodo aperto. Se un *Agente AI* partecipa a una decisione pubblica errata o dannosa, individuare chi ne risponde (lo sviluppatore, l'ente che lo impiega, l'utilizzatore finale, …) è tutt'altro che scontato.

Il compito che ci attende, dunque, è fondare condizioni perché gli *Agenti AI* siano comprensibili, governabili e coerenti con i valori che vogliamo continuare a riconoscere come umani e democratici.

Lucio Romano

Centro Interuniversitario di Ricerca Bioetica (CIRB)

02 Luglio 2026