Interazioni cervello-intelligenza artificiale nel lungo termine:

il "principio delle 3 R"

di Simone Rossi

Siena Brain Investigation & Neuromodulation Lab (Si-BIN Lab), Dipartimento di Medicina, Chirurgia e Neuroscienze, Università di Siena

UOC Neurologia, Azienda Ospedaliero-Universitaria Senese

Direttore della Scuola di specializzazione in Neurologia

Il seguente scritto è la traduzione italiana fatta con Chat GPT 5.0, ma ampiamente modificata dall'autore, dell'articolo pubblicato su Nature Publishing Journal Artificial Intelligence (2026; 2: 15) intitolato "*The brain side of human-AI interactions in the long-term: the 3R principle*", di Simone Rossi, Valter Fraccaro & Riccardo Manzotti.
Riflette la versione più scientifica dell'ultimo capitolo del libro divulgativo "Lo Tsunami. IA & IO" di Riccardo Manzotti e Simone Rossi, pubblicato da Rubbettino editore nel dicembre 2025.

Riassunto

Il modo in cui gli esseri umani interagiscono con l'IA può, nel lungo termine, plasmare la plasticità del cervello. In particolare, l'affidamento passivo e acritico sull'IA può indebolire la plasticità cerebrale a diversi livelli, ed erodere quindi le capacità cognitive, mentre una co-creazione attiva può mantenerla attiva o persino potenziarla. Facendo riferimento alle regole che governano la plasticità cerebrale e a considerazioni etiche, proponiamo il "principio delle 3R" — Risultati, Risposte, Responsabilità — come quadro preventivo di igiene cognitiva, in un momento storico in cui il tempo passato in interazione con le varie IA è sempre maggiore e l'esposizione a questi sistemi avviene fin dall'infanzia.

Il concetto di plasticità cerebrale

La neuroplasticità è la straordinaria capacità del cervello di modificare l'efficacia delle proprie sinapsi, di formare nuove connessioni neurali e di riorganizzare le reti nervose (*networks*) lungo tutto l'arco della vita. E' fondamentale per lo sviluppo, l'apprendimento, la memoria, l'invecchiamento sano, il recupero post-lesionale [1,2] e sembra contribuire anche alle capacità intellettive dell'individuo [3]. Secondo il principio storico dell' "*use it or lose it*" ("o lo usi o lo perdi"), applicabile a ogni funzione cerebrale e divenuto popolare nelle neuroscienze nella seconda metà del XX secolo, la plasticità è un fenomeno attivo che deve essere mantenuto attraverso l'allenamento costante in una determinata funzione cognitiva, motoria o percettiva (cioè plasticità dipendente dall'attività e dal tempo [1]). Le neuroscienze contemporanee ci stanno mostrando [3,4] che la neuroplasticità è un insieme complesso di fenomeni neurobiologici che possono verificarsi a diversi livelli del sistema nervoso, dalle modificazioni dell'efficienza di comunicazione a livello sinaptico (vedi sotto), incluse le forme più complesse di metaplasticità e plasticità omeostatica, fino alle dinamiche ancora più complesse a livello di rete neuronale o alle vere e proprie riorganizzazioni strutturali, come quelle che si verificano nei processi riparativi dopo una lesione cerebrale. La metaplasticità [5] (cioè la modulazione attività-dipendente della capacità di futuri cambiamenti sinaptici) e la plasticità omeostatica [6] (cioè la capacità del sistema di preservare un equilibrio complessivo stabile tra eccitazione e inibizione nelle reti cerebrali) contribuiscono a modellare, nel lungo periodo, gli adattamenti a livello di *network*, influenzando sia la connettività funzionale che quella strutturale.

Gli stili di interazione uomo-IA possono modellare la plasticità

Da quando ChatGPT è divenuta pubblicamente disponibile e ampiamente utilizzata a partire da novembre 2022 (e con lo sviluppo successivo di altri sistemi di intelligenza artificiale (IA), come Claude, DeepSeek, Gemini etc.) il tempo che le persone trascorrono interagendo con questi sistemi è aumentato in modo esponenziale. I dati *dell'Advanced Interactive Prompt Repository Management* (AIPRM) suggeriscono che gli adulti trascorrono oggi quasi 2 ore al giorno in interazione diretta con l'IA, con interazioni indirette - tramite *feed* personalizzati o ricerche guidate da algoritmi - che estendono questo tempo a circa 6-7 ore [7]. Il cervello umano è quindi sottoposto sempre di più a una quota crescente della giornata in dialogo con le IA. E questo processo, al momento, sembra inarrestabile.

Naturalmente, gli stili individuali di interazione con l'IA variano ampiamente. Mentre alcuni utenti accettano passivamente e acriticamente (cioè copiano e incollano) gli *output* dell'IA, altri interpretano, verificano e collaborano continuamente con l'IA in un intreccio consapevole e critico. Partendo dalla necessaria premessa che i concetti di seguito esposti richiederanno conferma da studi prospettici, è comunque lecito ipotizzare

che questi atteggiamenti umani contrastanti verso l'IA influenzino in modo profondamente diverso la plasticità cerebrale nel tempo (che, come sopra riportato è per l'appunto tempo e attività-dipendente): la conseguenza possibile è che un'interazione passiva possa condurre a una sorta di erosione delle nostre capacità cognitive, mentre il coinvolgimento attivo possa produrre, a lungo andare, anche esiti benefici, grazie alla conoscenza che le IA ci mettono, come mai prima era successo, a disposizione.

In questo articolo, in particolare, si sostiene che l'IA, se usata acriticamente, ponga un duplice rischio: possa incoraggiare il *cognitive offloading* (cioè l'esternalizzazione, o l'appalto, delle nostre facoltà cognitive all'IA), potenzialmente indebolendo le nostre capacità di elaborare informazioni, pianificare azioni e risolvere problemi - ma anche (e cosa cruciale per questo articolo) *l'intentional offloading*, cioè l'erosione della propria "bussola morale" e l'indebolimento del senso di *agency* personale. Per contrastare questi possibili rischi, proponiamo quindi una sorta di strategia concettuale preventiva.

Il presupposto teorico di questa idea si radica principalmente nella teoria di Bienenstock-Cooper-Munro (BCM) [8], secondo cui la forza sinaptica (l'efficacia della comunicazione tra neuroni) cambia in funzione dell'attività del neurone post-sinaptico rispetto a una soglia dinamica. Diversi regimi di interazione uomo-IA possono innescare opposti spostamenti metaplastici, esperienza-dipendenti, della soglia di modifica dell'attività sinaptica, oltre a tentativi omeostatici di regolare prolungati *up- o down-states* dell'attività sinaptica [9]. Nella fattispecie, se un essere umano interagisce passivamente con l'IA per settimane/mesi, limitandosi ad accettare suggerimenti o a seguire le raccomandazioni del *chatbot*, l'attività neurale potrebbe rimanere al di sotto di questa soglia, innescando depressione a lungo termine (LTD) dell'attività sinaptica, indebolendone l'efficacia comunicativa. Al contrario, un coinvolgimento attivo (per esempio interrogare, rifinire, co-creare con l'IA e apprendere dalle vaste conoscenze a disposizione) potrebbe portare l'attività al di sopra della soglia, promuovendo potenziamento a lungo termine (LTP) delle sinapsi e rafforzando la comunicazione neurale. LTD e LTP rappresentano infatti le basi biologiche dei meccanismi tramite i quali le sinapsi modificano la loro forza di connessione. In altre parole, tradotti questi meccanismi su scala di funzione cerebrale, un atteggiamento passivo può erodere capacità decisionali e pensiero critico (come discusso più avanti), mentre un'interazione attiva può migliorare l'apprendimento e preservare, se non migliorare, la funzione cognitiva.

Uno studio preliminare, finora non sottoposto a *peer review* [10], sembra supportare questa ipotesi: la connettività delle reti (misurata attraverso la registrazione dell'attività elettrica oscillatoria cerebrale durante compiti di scrittura di un testo) si indebolisce se i soggetti si affidano esclusivamente all'aiuto dell'IA, mentre si rafforza se si affidano alle proprie capacità cerebrali, risultando intermedia quando si effettuano ricerche internet tradizionali. Futuri studi prospettici, basati su indagini neurofisiologiche e metaboliche funzionali delle dinamiche di connettività cerebrale, sono necessari per chiarire le firme

neurali, a livello di sistema, associate a questi diversi atteggiamenti di interazione con l'IA.

Considerazioni etiche e filosofiche

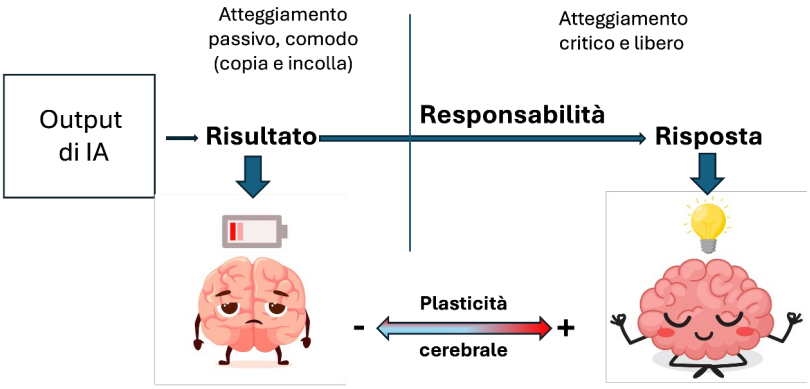
Oltre alla dimensione neurobiologica, le interazioni uomo-IA stanno sollevando ulteriori questioni filosofiche ed etiche. Questi concetti sono espressi in ciò che abbiamo chiamato il "principio delle 3R", originariamente descritto con un significato diverso (*Replacement, Reduction, Refinement*) per guidare la ricerca sperimentale sugli animali [11,12]. Trasponendo il principio delle 3R all'interazione uomo-IA, vogliamo sottolineare la differenza tra Risultati e Risposte, focalizzandoci sulla Responsabilità individuale. Gli *output* dell'IA alle nostre domande sono fondamentalmente risultati che, se accettati senza verifica, non possono ancora essere considerate vere risposte: questo perchè i *large language models* (LLM) (i modelli linguistici di larga scala alla base dei sistemi di IA disponibili) non rappresentano il, nè operano sul, significato. Mancano di valutazione critica interna e non comprendono minimamente i propri *output*. Ovviamente il risultato, essendo un costrutto squisitamente matematico, può essere sbagliato o corretto, ma in ogni caso mancherà di significato finchè non sarà interpretato dall'interlocutore umano. Per esempio, l'IA può suggerire con *nonchalance* risposte contrastanti sulla base degli stessi dati in ingresso, a seconda del *prompt*, ignorandone comunque sia il valore che le possibili conseguenze. Gli LLM non hanno eziologia, nè radicamento in un sistema di valori dotato di significato, oppure, in altre parole, possono codificare obiettivi ottimizzati senza nessuna comprensione morale. Gli esseri umani invece si.

Una Risposta, a differenza di un Risultato, non è semplicemente un esito statisticamente plausibile, ma possiede un valore morale ed esistenziale basato sulle sue possibili conseguenze – cioè nella Risposta è insito il significato. Quest'ultimo dipende da quali valori gli esseri umani scelgono come criteri per le proprie azioni. Crucialmente, i valori non sono funzioni di costo da ottimizzare: piuttosto, determinano quali funzioni di costo gli esseri umani scelgono. Questa scelta, che appartiene unicamente agli esseri umani, rappresenta l'essenza della terza R, ovvero di ciò che abbiamo definito Responsabilità nel principio delle 3R.

Ogni processo decisionale cognitivo comporta considerazioni esistenziali che vanno oltre i processi puramente razionali. Per esempio, un agente che dovesse scegliere tra aumentare il Prodotto Interno Lordo o ridurre l'impatto ambientale, metterebbe in atto una decisione che non può essere risolta dal solo calcolo razionale. La Responsabilità implica selezionare quale esito sia preferibile. Mentre le decisioni razionali si basano su valutazioni deliberate e stima delle conseguenze, le scelte responsabili non sono pienamente determinate dalla sola ragione e pongono il carico della Responsabilità sull'agente. Il premio Nobel Herbert Simon distinse tra deliberazioni razionali e scelte di *pay-off* [13]. Simon riconobbe un fondamento pre-razionale, assiologico ed esistenziale

della responsabilità. Osservò: “Non esiste oggi una teoria soddisfacente che spieghi come i valori si formino, da dove provengano o come vengano modificati nel processo decisionale” [14]. LLM e IA non forniscono alcun contributo esistenziale, a meno che non mimino valori umani. Infatti, un atteggiamento eziologicamente attivo richiede almeno intenzionalità e *embodiment* (un termine anglosassone che indica incarnazione), entrambi fattori che mancano agli attuali LLM, focalizzati solo sulla generazione di contenuti stocasticamente coerenti. Chiaramente, il cervello compie il lavoro ulteriore di scelta dei valori e quindi di assumersi la responsabilità delle proprie decisioni. In sintesi, assumersi la responsabilità ha costi associati e richiede uno sforzo aggiuntivo rispetto alla deliberazione cognitiva standard; un costo che, se trascurato, potrebbe ridurre a lungo andare tale aspetto negli esseri umani.

Per riassumere, Risultati e Risposte costituiscono i primi due elementi di questo quadro comportamentale, mentre la terza R sta per Responsabilità (Figura 1). Mentre risultati e risposte possono essere modellati in modo puramente computazionale e cognitivo, la responsabilità ha una profonda natura etica, eziologica ed esistenziale che non può essere ignorata. A prescindere da quante capacità cognitive vengano esternalizzate all’IA, il principio delle 3R sottolinea che gli esseri umani devono sempre assumersi la responsabilità di questi risultati, che - per quanto sopra esposto - non possono ancora essere considerate vere risposte. Il principio delle 3R evidenzia che prima di qualsiasi processo cognitivo vi è sempre una valutazione eziologica ed etica che non è esternalizzabile all’IA. Come detto, gli esseri umani dovrebbero interpretare e giudicare attivamente i risultati dell’IA, contestualizzandoli entro i propri principi etici, sociali e culturali e soppesando il significato delle loro conseguenze. Solo attraverso questo processo i risultati diventeranno vere e proprie risposte. Questo passaggio è cruciale, poichè è il momento in cui i dati assurgono a significato (o acquisiscono un valore) e i possibili risultati danno origine ad azioni responsabili.



Legenda della Figura 1*. Il principio delle 3R: l’IA fornisce risultati, che possono diventare vere risposte solo mediante la valutazione responsabile degli esseri umani. Un*

atteggiamento passivo riduce nel lungo termine la plasticità cerebrale attività-dipendente. Un atteggiamento eziologicamente attivo promuove la plasticità cerebrale.

Recentemente, alcuni ricercatori hanno proposto che l'IA possa rappresentare un ulteriore livello artificiale, non biologico, di intelligenza distribuita, chiamato "sistema 0" [15], che interagisce continuamente con le due classiche modalità operative del ragionamento umano, sebbene ancora prive di dimostrazione empirica [16], proposte da Kahneman (sistema 1: pensiero rapido e intuitivo; sistema 2: pensiero lento e analitico) [17]. In questo senso, il sistema 0 (un ibrido umano-IA) rappresenterebbe una modalità più pervasiva per esternalizzare all'IA un numero crescente di compiti cognitivi, facendo leva sulla sua vasta conoscenza e velocità di elaborazione ben oltre le capacità umane, ma comportando probabili rischi non solo di *cognitive offloading* [18], ma anche di erosione dell'autonomia, della cognizione, dell'*agency* e della stessa identità umana [19].

Il principio delle 3R come proposta di "igiene cognitiva"

Adottando l'intuitivo principio delle 3R, che per la sua semplicità e immediatezza può ben risuonare trasversalmente tra discipline differenti e anche tra i non specialisti, è possibile evidenziare l'esistenza del suddetto livello eziologico, che conduce a risposte responsabili piuttosto che a semplici risultati. Data l'attuale incapacità dell'IA di produrre significato [20], se l'interlocutore umano si disinteressa della mancata distinzione tra risultati e risposte responsabili rischia l'abdicazione della propria responsabilità. D'altro canto, accogliendo il principio delle 3R, indipendentemente da quanto gli esseri umani delegheranno all'IA le loro incombenze cognitive, essi resteranno comunque responsabili del significato finale di tali risultati. Tale responsabilità, che richiede un coinvolgimento cognitivo continuo e sostenuto, e quindi un costo mentale importante, è un fattore determinante nel favorire la plasticità cerebrale, e di conseguenza l'arricchimento cognitivo.

All'inizio dell'era dell'IA, tenendo conto che la crescente interazione uomo-IA non può essere né stoppata né invertita, il principio delle 3R offre una linea guida di igiene cognitiva per proteggere il cervello dall'erosione cognitiva nel lungo termine dovuta a un atteggiamento eziologicamente passivo verso gli LLM.

Infatti, tale principio mette in luce la differenza tra processi cognitivi, che mancano di significato e orientamento allo scopo, e valutazioni/scelte eziologiche, che sono cruciali per scegliere azioni significative e enfatizza l'assunzione di responsabilità. Il semplice comportamento consulenziale dell'IA può influenzare le decisioni morali e le attribuzioni esplicite di responsabilità in contesti simulati impegnativi [21].

Dato che la plasticità cerebrale è strettamente dipendente dall'attività e dal tempo [22], l'adozione precoce del principio delle 3R è critica. Si rivolge, questo principio, soprattutto agli educatori e docenti della scuola primaria, secondaria e dell'Università,

che sono i più titolati al suo utilizzo per far capire agli studenti (il cui cervello è nel periodo di massima espressione delle capacità plastiche) di non rinunciare mai a mettere in discussione i contenuti generati dall'IA. Analogamente, il principio delle 3R potrebbe essere utile anche a studiosi, ricercatori, lavoratori di ufficio, operatori e cittadini comuni (tutta la società, in sostanza) per preservare la propria plasticità cerebrale e, in fin dei conti, il loro pensiero critico. E' da ricordare, infatti, che recenti evidenze comportamentali sulla dipendenza eccessiva dal dialogo con l'IA possono condurre a una riduzione delle capacità analitiche e di pensiero critico individuali [23] e che, sebbene il supporto dell'IA sia spesso percepito come efficiente dagli utenti, molti di loro riportano una riduzione delle loro capacità cognitive e del pensiero critico [24], con un impatto particolarmente marcato nei soggetti più giovani [25].

In conclusione, il principio delle 3R delinea due modalità operative della mente: quella puramente cognitiva e quella eziologica. La prima produrrebbe risultati e la seconda produrrebbe significato. Astenersi da entrambe significa rinunciare a priori al controllo della barra del timone della propria vita: è ragionevole ritenere che questo atteggiamento potrebbe condurre non solo al *cognitive offloading*, ma anche a una forma più profonda, e terribilmente pericolosa, di delega della volontà e conseguente privazione della libertà individuale.

Bibliografia

1. Pascual-Leone, A., Amedi, A., Fregni, F., Merabet, L.B. The plastic human brain cortex. Annu Rev Neurosci. 2005; 28: 377-401
2. Kumar, A., Patel, T., Sugandh, F., Dev, J., Kumar, U., Adeeb, M., Kachhadia, M.P., Puri, P., Prachi, F., Zaman, M.U., Kumar, S., Varrassi, G., Syed, A.R.S. Innovative Approaches and Therapies to Enhance Neuroplasticity in Neurological Disorders. Front Neurosci, 2023; 17: 1042570
3. Santarnecchi, E. and Rossi, S. Advances in the neuroscience of intelligence: from brain connectivity to brain perturbation. The Spanish J. Psychol., 2016; 19: E94
4. Gazerani, P. The neuroplastic brain: current breakthroughs and new insights. Brain Research, 2025; 1786: 148080
5. Abraham, W. C. and Bear, M. F. Metaplasticity: The Plasticity of Synaptic Plasticity. Trends in Neurosciences, 1996; 19, 126-130.
6. Turrigiano, G.The dialectic of Hebb and homeostasis. Philosophical Transactions of the Royal Society B: Biological Sciences, 2017; 372(1715), 20160258.
7. Digital 2024 Global Overview Report. DataReportal. https://datareportal.com/reports/digital-2024-global-overview-report?utm_source=chatgpt.com
8. Bienenstock, E.L., Cooper, L.N and Munro, P.W. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. J. Neurosci., 1982; 2 (1): 32-48

9. Lee, H.K.. Metaplasticity framework for cross-modal synaptic plasticity in adults. Frontiers in Synaptic Neuroscience, 203; 15: 1087042
10. Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Tasks, ArXiv, 2025: 2506.08872

11. Vitale, A., & Ricceri, L. The principle of the 3Rs between aspiration and reality. Frontiers in Physiology, 2022; 13: 914939
12. Lauwereyns, J., Bajramovic, J., Bert, B. et al. Toward a common interpretation of the 3Rs principles in animal research. Lab Animal, 2024; 53: 209-211
13. Simon, H.A., A behavioral model of rational choice. The Quarterly Journal of Economics, 1955. 69(1): p. 99-118
14. Simon, H.A., Models of Bounded Rationality. 1997, Cambridge (Mass): Mit Press
15. Chiriatti, M., Ganapini, M., Panai, E., Ubiali, M., Riva, G. The case for human-AI interaction as system 0 thinking. Nature Human Behaviour, 2024; 8, 1829-1830
16. Melnikoff, D.E., Bargh, J.A. The mythical number two. Trends Cogn Sci, 2018; 44 (4): 280-293
17. Kahneman, D. In The Nobel Prizes 2002 (ed. T. Frängsmyr) 449-489 (Nobel Foundation, 2002)
18. Risko, E.F. & Gilbert, S.J. Cognitive Offloading. Trends Cogn Sci. 2016; 20 (9): 676-688
19. Di Plinio, S. Panta Rh-AI: Assessing multifaceted AI threats on human agency and identity. Social Sciences & Humanities Open, 2025; 11: 101434
20. Floridi, L. & Chiriatti, M. Minds Mach. 30, 681-694 (2020).
21. Salatino, A., Prével, A., Caspar, E. A., & Lo Bue, S. Influence of AI behavior on human moral decisions, agency, and responsibility. Scientific Reports, 2025; 15: 12329
22. Buonomano, D.V. & Merzenich, M.M. Cortical plasticity: from synapses to maps. Ann Rev Neurosci. 1998; 21: 149-86
23. Zhai, C., Wibowo, S., & Li, L. D. The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. Smart Learning Environments, 2024; 11: 28
24. Lee, H.P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., Wilson, N. The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and confidence effects from a survey of knowledge workers. Proceedings of CHI Conference, 2025; 1121: 1-22
25. Gerlich, Michael, AI. Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. Available at SSRN: https://ssrn.com/abstract=5082524 or http://dx.doi.org/10.2139/ssrn.5082524