https://doi.org/10.1038/s44387-025-00063-1

The brain side of human-AI interactions in the long-term: the "3R principle"

Simone Rossi, Valter Fraccaro & Riccardo Manzotti

Neuroplasticity is shaped by how humans interact with AI. We argue that passive, uncritical, reliance on AI may weaken activity-dependent brain plasticity and erode cognition, whereas active co-creation can sustain or enhance it. Drawing on plasticity rules and ethical considerations, we propose the 3R principle—Results, Responses, Responsibility—as a preventive framework for cognitive hygiene, urging education towards AI use to preserve agency, meaning-making, and long-term brain health.

The concept of brain plasticity

Neuroplasticity is the remarkable brain's ability to modify synaptic efficacy or form new neural connections and reorganize networks throughout lifespan. It is foundational to development, learning, memory, healthy ageing, post-lesional recovery[1,2] and possibly even human intelligence capabilities[3]. According to the historical "use it or lose it" principle, applicable to every brain function, which became popular in neuroscience in the latter half of the 20th century, plasticity is an active phenomenon that needs to be maintained through consistent engagement in a given cognitive, motor or perceptual brain function (i.e., activity- and time-dependent plasticity[1]). Contemporary neuroscience is now showing[3,4] that neuroplasticity is a complex ensemble of neurobiological phenomena that may take place at several levels of the nervous system, spanning from modifications of communication efficiency at synaptic level (see below), including metaplasticity and homeostatic plasticity, up to network-level dynamics or structural reorganizations. Metaplasticity[5] (i.e., the activity-dependent modulation of the capacity for future synaptic change) and homeostatic plasticity[6] (i.e., the ability of the system to preserve a stable overall excitatory-inhibitory balance in brain networks) contribute to shape, in the long term, network level adaptations by impacting on functional and structural connectivity.

Styles of human AI-interaction may shape plasticity

Since ChatGPT became publicly available and widely used as of November 2022, the time people spend interacting with artificial intelligence (AI) has significantly increased. Data from the Advanced Interactive Prompt Repository Management (AIPRM) suggest that adults now spend nearly 2 h per day in direct AI interaction, with indirect interactions—via personalized feeds or algorithm-driven searches—extending this to about 6–7 h[7]. Therefore, the human brain is more and more subjected to an increasing amount of daytime in dialogue with AIs.

Of course, styles of interaction with AI vary widely. While some users passively and uncritically accept (i.e., copy and paste) AI outputs, others interpret, verify, and continuously collaborate with AI in a conscious and critical interplay. Based on the premise that the following concepts will require support from prospective studies, we hypothesize that these contrasting human attitudes to AI will differently impact brain plasticity over time, with the consequences that passive interaction may lead to cognitive erosion, while active engagement may produce beneficial outcomes. In particular, we argue that AI, if used uncritically, poses a dual risk: it may encourage *cognitive* offloading, potentially undermining our abilities to process information, plan actions, and solve problems -but also (and crucially for this comment) *intentional* offloading- i.e., eroding our moral compass and weakening personal agency. To counter these risks, we then propose a preventive framework.

This idea is rooted primarily in the Bienenstock–Cooper–Munro (BCM) theory[8], which posits that synaptic strength (the efficacy of communication between neurons) changes depending on the postsynaptic neuron activity relative to a dynamic threshold. Different regimens of human-AI interaction may trigger opposite metaplastic, experience-dependent shifts in the long-term depression (LTD)/long-term potentiation (LTP) threshold, as well as homeostatic attempts to regulate prolonged up or down-states of synaptic activity[9]. If a human interacts passively with AI over weeks/months, by just accepting suggestions or following chatbot's recommendations, neural activity could remain below this threshold, triggering LTD and weakening synaptic strength. In contrast, active engagement (e.g., questioning, refining, co-creating with AI) could raise activity above the threshold, promoting LTP and strengthening neural communication. In other words, a passive attitude may erode decision-making and critical thinking (see later in the article), whereas active interplay may enhance learning and preserve, if not improve, cognitive function.

Preliminary, so far unpeer-reviewed study[10], seems to support this hypothesis: networks' connectivity (measured through the recording of brain oscillatory electrical activity during writing tasks) is weakened if subjects rely exclusively on AI help, while it is strengthened if they rely on their own brain capacities, and is intermediate when traditional internet searches are performed. Future prospective studies based on neurophysiological and metabolic functional investigations of brain dynamics and connectivity are warranted to disclose system-level neural signatures of these different AI-interaction attitudes.

Ethical and philosophical considerations

Beyond the neurobiological dimension, human-AI interactions may rise additional philosophical and ethical questions. We have encapsulated these concepts in what we call "the 3R Principle", originally described with a

different meaning (Replacement, Reduction, Refinement) to guide animal research[11,12]. By translating the 3R principle to human-AI interaction, we would like to stress the difference between *Results* and *Responses* by focusing on *Responsibility*. AI outputs are basically results, if accepted without verification, rather than genuine responses: this because large language models (LLMs) neither represent nor operate on meaning. They lack internal critical evaluation and do not understand their outputs. Thus, as results, they may be wrong or right, but it is clear that they lack meaning until interpreted by humans. For instance, AI can equally suggest both a positive or negative course of action based on the same data, depending on the prompt, ignoring their value or consequences. LLMs have no aetiology, no grounding in a meaningful value system or, in other words, they can encode optimized objectives without moral understanding. Humans do.

A response, unlike a result, is not simply a statistically viable outcome but has a moral and existential value based on its consequences—i.e., a response has meaning. The latter depends on which value humans chose to as criteria for their actions. Crucially, values are not cost functions to be optimized: rather, they determine which cost functions humans choose. This choice, the uniquely pertains to humans, is the essence of what we referred to as Responsibility in the 3R principle.

Any cognitive decision process involves existential considerations that extend beyond purely rational processes. For example, an agent may need to choose between increasing Gross Domestic Product and reducing environmental impact, which cannot be resolved by rational calculation alone. Responsibility entails selecting which outcome is preferable. While rational decisions rely on deliberate evaluation and consequence assessment, responsible choices are not fully determined by reason alone and place the burden of responsibility on the agent.

Nobel laureate Herbert Simon distinguished between rational deliberations and pay-off choices[13]. Simon recognized a pre-rational, axiological, and existential foundation for responsibility. He noted: "There is no satisfactory theory today that explains how values are formed, where they come from, or how they are modified in the decision-making process"[14]. LLM and AI provide no existential contribution unless they mimic human values. In fact, an active aetiological attitude requires at least intentionality and embodiment, which are missing from current LLMs that focuses on generation of stochastically consistent content. Clearly, the brain does the extra job of choosing values and therefore assuming the responsibility of his decisions. In summary, assuming responsibility has associated costs and requires additional effort beyond standard cognitive deliberation that, if neglected, will reduce such an aspect in human beings.

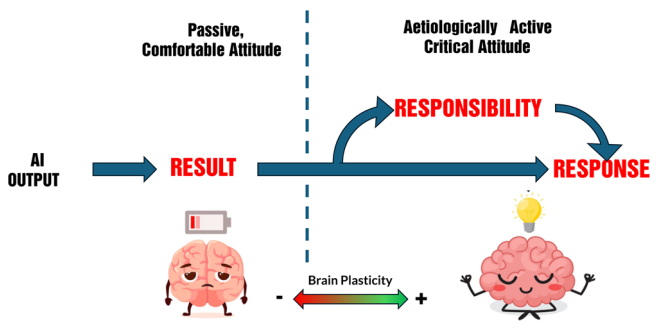

Fig. 1 | The 3 R principle: IA provides results that can become true responses only by means of the responsible evaluation of humans. A passive attitude reduce activity-dependent brain plasticity in the long term. An aetiologically active attitude promote brain plasticity. Credits for the Fig. CC BY-NC-ND.

Thus, *Results* and *Responses* form the first two elements of this behavioural framework, while the third R stands for *Responsibility* (Fig. 1). While results and responses can be modelled purely in computational and cognitive ways, responsibility has an ethical, aetiological and existential nature that cannot be ignored. Regardless of how much cognitive capacities are externalized to AI, the 3 R principle underscores that humans must always take responsibility for these results, which aren't yet real responses. The 3R principle highlights that before any cognitive process, there is always an aetiological and ethical evaluation that is not externalizable to AI. As said, humans should actively interpret and judge AI results, contextualizing them within their ethical, social and cultural principles and weighting up the meaning of their consequences. Only though this process do results become responses. This step is crucial, as it is the moment when data becomes meaning (or acquires a value), and possible results give rise to responsible actions.

Recently, some researchers have proposed that AI may represent an additional artificial, non-biological layer of distributed intelligence, called system 0[15], that interacts continuously with the two classical, although still lacking of empirical demonstration[16], Kahneman's operational modalities of the human brain reasoning (system 1: fast, intuitive thinking and system 2: slow, analytical thinking)[17]. Thus, system 0 would represent a more pervasive way to outsource more cognitive tasks to AI, leveraging on its vast knowledge and processing speed beyond human capabilities, but carrying probable risks not only for cognitive offloading[18], but even of erosion of human autonomy, cognition, agency and identity, in ways that directly implicate human responsibility[19].

The 3R principle as a cognitive hygiene suggestion

By adopting the intuitive 3R principle, that could well resonate either across disciplines or among lay people due to its simplicity and immediacy, it is possible to highlight the existence of the aforementioned aetiological level, leading to responsible responses rather than to simple results. Given AI's current inability to produce meaning[20], failure to distinguish between results and responsible responses leads to abdication of responsibility. On the other hand, embracing the 3 R principle, no matter how much humans will delegate cognitive chores to AI, they will still be responsible for the meaning of such results. Such responsibility demands sustained cognitive engagement, a mental cost that may foster brain plasticity.

At the onset of the AI-age, taking into account that the growing human-AI interplay cannot be reversed, the 3R principle offers a cognitive-hygiene guideline to protect the brain from cognitive erosion in the long-term due to an aetiologically passive attitude towards LLMs.

Indeed, such a principle highlights the difference between cognitive processes, which lack meaning and goal-directedness, and aetiological evaluation and choices, which are crucial to choose meaningful actions and take responsibility for such actions. The mere AI advisory behaviour can influence moral decisions and explicit responsibility attributions in simulated high-stakes contexts[21].

Given that brain plasticity is activity- and time-dependent[22], early adoption of the 3R principle is critical. Educators in primary, middle, high school and University can use this framework to encourage students not to relinquish in questioning AI-generated content, from a perspective that is irreducible to pure cognitive processes. Likewise, the 3R principle might also be useful for scholars, researchers, white and blue collars and laypersons to preserve their brain plasticity and critical thinking. Indeed, recent behavioural findings on over-reliance on AI dialogue suggest a reduction of analytical and critical-thinking skills[23] and that, while AI support is often perceived as efficient, many users report reduced cognitive effort and critical-thinking practices[24], with younger individuals being especially affected[25].

In conclusion, the 3R principle delineates two modes of mind: the purely cognitive and the aetiological one. The former would produce results, and the latter would produce meaning. Surrendering both means relinquishing control of the bar: this will likely lead not only to cognitive offloading but also to a deeper and more dangerous form of offloading the will.

Data availability

No datasets were generated or analysed during the current study.

Simone Rossi[1] ✉, **Valter Fraccaro**[2] **& Riccardo Manzotti**[3]
[1]Siena Brain Investigation & Neuromodulation Lab (Si-BIN Lab), Department of Medicine, Surgery and Neuroscience, University of Siena, Siena, Italy. [2]Siena Artificial Intelligence Hub (SAI Hub), Siena, Italy. [3]Theoretical Philosophy, Department of Business, Law, Economics and Consumer Behaviour, IULM University, Milan, Italy.
✉e-mail: Simone.rossi@unisi.it

Received: 27 June 2025; Accepted: 10 December 2025;
Published online: 27 January 2026

References

1. Pascual-Leone, A., Amedi, A., Fregni, F. & Merabet, L. B. The plastic human brain cortex. *Annu. Rev. Neurosci.* **28**, 377–401 (2005).
2. Kumar, A. et al. Innovative approaches and therapies to enhance neuroplasticity in neurological disorders. *Front. Neurosci.* **17**, 1042570 (2023).
3. Santarnecchi, E. & Rossi, S. Advances in the neuroscience of intelligence: from brain connectivity to brain perturbation. *Span. J. Psychol.* **19**, E94 (2016).
4. Gazerani, P. The neuroplastic brain: current breakthroughs and new insights. *Brain Res.* **1786**, 148080 (2025).
5. Abraham, W. C. & Bear, M. F. Metaplasticity: the plasticity of synaptic plasticity. *Trends Neurosci.* **19**, 126–130 (1996).
6. Turrigiano, G. The dialectic of Hebb and homeostasis. *Philos. Trans. R. Soc. B Biol. Sci.* **372**, 20160258 (2017).
7. Digital 2024 Global Overview Report. *DataReportal*. https://datareportal.com/reports/digital-2024-global-overview-report?utm_source=chatgpt.com (2024).
8. Bienenstock, E. L., Cooper, L. N. & Munro, P. W. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J. Neurosci.* **2**, 32–48 (1982).
9. Lee, H. K. Metaplasticity framework for cross-modal synaptic plasticity in adults. *Front. Synaptic Neurosci.* **15**, 1087042 (2023).
10. Kosmyna, N. et al. Your brain on ChatGPT: accumulation of cognitive debt when using an AI assistant for essay writing tasks. Preprint at *ArXiv* https://doi.org/10.48550/arXiv.2506.08872 (2025).
11. Vitale, A. & Ricceri, L. The principle of the 3Rs between aspiration and reality. *Front. Physiol.* **13**, 914939 (2022).
12. Lauwereyns, J. et al. Toward a common interpretation of the 3Rs principles in animal research. *Lab Anim.* **53**, 209–211 (2024).
13. Simon, H. A. A behavioral model of rational choice. *Q. J. Econ.* **69**, 99–118 (1955).
14. Simon, H. A. *Models of Bounded Rationality* (MIT Press, 1997).
15. Chiriatti, M., Ganapini, M., Panai, E., Ubiali, M. & Riva, G. The case for human-AI interaction as system 0 thinking. *Nat. Hum. Behav.* **8**, 1829–1830 (2024).
16. Melnikoff, D. E. & Bargh, J. A. The mythical number two. *Trends Cogn. Sci.* **44**, 280–293 (2018).
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.* **20**, 676–688 (2016).
19. Di Plinio, S. Panta Rh-AI: assessing multifaceted AI threats on human agency and identity. *Soc. Sci. Humanit. Open* **11**, 101434 (2025).
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. *Sci. Rep.* **15**, 12329 (2025).
22. Buonomano, D. V. & Merzenich, M. M. Cortical plasticity: from synapses to maps. *Ann. Rev. Neurosci.* **21**, 149–186 (1998).
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 Learn. Environ.* **11**, 28 (2024).
24. Lee, H. P. et al. The impact of generative AI on critical thinking: self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In *Proc. 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)* 1–22 (Association for Computing Machinery, 2025).
25. Gerlich, M. AI tools in society: impacts on cognitive offloading and the future of critical thinking. https://ssrn.com/abstract=5082524 or https://doi.org/10.2139/ssrn.5082524 (2025).

Acknowledgements

The work received no funding.

Author contributions

S.R., V.F. and R.M. equally contributed to structuring, reviewing and writing the work.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Simone Rossi.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.