



Where and How AI Shapes our Thinking is the Key Point

Francesco Bianchini¹ 

Received: 12 March 2026 / Accepted: 14 March 2026
© The Author(s) 2026

Abstract

This paper examines how the progressive integration of AI into epistemic practices is reshaping human knowledge production and our broader interface with reality. Building on Bianchini (2025a) and engaging with Branda (2026), it argues that AI can no longer be regarded as an epistemically neutral tool, but actively influences human cognition in supportive, indifferent, or detrimental ways. The widespread adoption of Large Language Models across digital environments introduces risks of conceptual homogenization and threatens epistemic diversity and creative originality. Drawing on the System 0 framework proposed by Chiriatti et al. (2024) and its extension in Branda's concept of the Thinkframe, the paper explores how AI participates in cognitive processes both at the individual and collective levels. While these frameworks illuminate important dynamics of human-AI co-creation, they require further theoretical refinement, particularly regarding the specific role of AI within distributed cognitive systems and its implications for epistemic autonomy and authority.

Keywords AI · Extended cognition · System 0 · Thinkframe · AI-Extenders

Francesco Branda (2026) builds on my account of the progressive conceptual reconfiguration of AI (Bianchini 2025a), interpreting it as a salient moment within an ongoing process of theoretical and practical redefinition imposed by the rapid transformations of the field. Among the shifts identified in that analysis, Branda foregrounds in particular the abandonment of the view of AI as a domain producing epistemically neutral tools. Instead, AI emerges as a technological domain in which devices are engineered that actively shape human knowledge and cognition, whether

✉ Francesco Bianchini
francesco.bianchini@unibo.it

¹ University of Bologna, Bologna, Italy

in supportive, indifferent, or detrimental ways, thereby potentially affecting, and in some cases compromising, their epistemic autonomy.

While this often-overlooked dimension is undoubtedly significant, it is equally important to recognize that the progressive permeation of AI into our social and professional environments is reshaping, if not distorting, our epistemic perspective and, more broadly, our interface with reality. Until recently, this interface was more firmly grounded in the possibility of relying on identifiable control agencies and relatively stable validation mechanisms. By contrast, the pervasiveness of AI systems and the velocity with which their outputs are generated and disseminated make both *ex ante* and *ex post* forms of oversight increasingly difficult. This transformation is further intensified by the fact that AI-generated content, especially textual material, is progressively becoming the scaffolding of our epistemic practices, including core processes of knowledge production. To request information from a search engine today often means encountering a response that has already been synthesized or filtered by generative AI, leaving aside the fact that search engines themselves have long relied on AI techniques, particularly in the domain of Natural Language Processing. Similarly, asking an AI system to, for instance, brainstorm ideas, produce a summary, or generate bibliographic references does not merely accelerate pre-existing tasks; it reconfigures the structure of our engagement with information. What is at stake is not only a modification of how knowledge is produced, but a more profound shift in our relation to knowledge as such.

All of this has produced a radical transformation within a remarkably short time span. Until recently, the online production and circulation of texts primarily raised concerns about the accreditation of *plausible but false* knowledge: digitally disseminated content could achieve epistemic legitimacy despite being partially or entirely fabricated, generating complex and convincing texts that were only *apparently* informative because they were invented (for an illustrative case, see Bianchini 2025b). Today, the situation has shifted once again. The dominant stance can no longer be, nor should it be, limited to a merely critical assessment of the reliability of generative-AI outputs. Rather, it must also account for the growing role of *co-participation* and *co-creation* in knowledge-producing practices. Knowledge production increasingly depends on texts generated by, for instance, Large Language Models (LLMs), which are now embedded across the digital tools we routinely employ: search engines, writing and editing environments, presentation software, and similar platforms. These texts are typically subjected to a sequence of transformations that involve the repeated deployment of LLMs at different stages of the same epistemic workflow. Under these conditions, evaluative attention shifts from the final product, the isolated output, to the *process* through which that output is generated, revised, and integrated. This shift also reconfigures the associated epistemic risks and opportunities. The central problem is no longer exclusively that of discriminating true from false content, but increasingly that of safeguarding creative originality and epistemic diversity against a tendency toward homogenization. Such flattening effects are driven by the standardization implicit in the extensive use of synthetic data, such as summaries, brainstorming outputs, and research texts generated by LLMs, which risk converging on recurrent conceptual patterns, styles, and inferential pathways.

Branda's proposal to frame – by making explicit its original potential – the way in which AI influences the thinking of the human user, or of the human being who interacts with it in cooperative and co-creative tasks, initially refers to the proposal of Chiriatti et al. (2024) regarding System 0 thinking. Recalling Daniel Kahneman's (2011) well-known distinction between two types of human cognition, one more instinctive and therefore faster (System 1), and one more rational and deliberative and therefore slower (System 2), Chiriatti and colleagues propose considering the unity produced by the dynamic and continuous interaction between human cognition and generative AI as a new system (precisely System 0), independent from and original with respect to the other two. System 0 is still a psychological system, even if not, or not entirely, biological, because it is entrusted with cognitive tasks that are performed in an integrated and distributed manner by both AI and human agents. In this sense, the system extends human cognition with respect to the activities traditionally associated with both System 1 and System 2. Naturally, this development may have both positive effects, such as the enhancement of human cognitive capacities, and negative consequences, including uncertainty about the authorship of decisions and the increasing reliance on synthetic data.

Beyond these more or less problematic aspects, also highlighted by the authors, it seems difficult to dispute the idea that, from the perspective of the individual as a cognitive agent, AI actively participates in cognitive capacity and performance, not only as a support but also as a modifier. The extent to which this occurs has been the subject of extensive debate in the literature on the Extended Mind Thesis, ranging from discussions on the autonomy (or lack thereof) of AI extenders (Hernandez-Orallo and Vold, 2019) to debates on cognitive enhancement (Clark, 2025) or cognitive diminishment (Fasoli et al., 2025). However, whether in positive or negative terms, a decisive, rather than merely secondary, influence of AI on human thought appears to be acknowledged by all authors. Defining its exact nature, and therefore its limits and potential, now seems an urgent task, given the widespread diffusion and ease of use of AI tools. Understanding what AI extends cognitively, what it appropriates, and what it replaces are substantial problems, with broad epistemological and ethical implications, because they contribute to redefining the ontology of cognition itself. A direct example concerns the now common use of systems such as GPT to write texts entirely on the basis of ideas proposed by the human user. It is easy to see how this activity could undermine humans' ability to write independently, especially during phases of secondary education and learning. Yet, beyond the actual risks and potential harmful consequences of such practices, they also represent one of the many ways in which AI shapes, both positively and negatively, the way we think. Indeed, precisely this example further corroborates that claim, providing both a theoretical and an empirical justification for it.

From a broader perspective, Branda introduces the concept of the Thinkframe, which extends the discourse initiated by System 0, operating primarily at the individual level, to the more complex level of structures in which multiple individuals are connected in cognitive activities through active tools enabled by AI (usually machine learning) algorithms, such as social networks, collaborative platforms, and systems based on recommendation algorithms. The expanded and systemic nature of the Thinkframe is intended to enhance processes of co-creation through a more dis-

tributed flow of information. In contexts where multiple users are involved, this flow is also more likely to be subject to critical monitoring than in systems based on interaction with a single human user. In this sense, the Thinkframe could be interpreted as a way of addressing or partially overcoming some of the critical issues identified in the System 0 framework.

However, the concept is not without its weaknesses. For instance, it remains necessary to define more precisely what differentiates the Thinkframe from other systems that can be interpreted in terms of distributed cognition (Hutchins, 1995), and therefore to clarify the specific role played by AI within such systems. Certainly, the presence of numerous human agents, rather than a single user, can confer greater epistemic authority and reliability on the system as a whole, as Branda emphasizes. Yet if the role of AI is not clearly specified, as is instead partly the case in the System 0 perspective, the Thinkframe risks being an insufficient explanatory framework for the environments in which AI is currently integrated with many human activities. For this reason, identifying the precise nodes of integration, as well as the directions of exchange and interaction between AI systems and human cognitive agents, appears to be a necessary condition for employing the Thinkframe concept in an analytically effective way.

Acknowledgements No.

Author Contributions One author for the whole paper

Funding Open access funding provided by Alma Mater Studiorum - Università di Bologna within the CRUI-CARE Agreement.

Data Availability Not applicable.

Declarations

Competing interests No.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bianchini, F. (2025a). Generative artificial intelligence: A concept in progress. *Philosophy & Technology*, 38, Article 46. <https://doi.org/10.1007/s13347-025-00875-8>
- Bianchini, F. (2025b). Evaluating Intelligence and Knowledge in Large Language Models. *Topoi*, 44, 163–173. <https://doi.org/10.1007/s11245-024-10072-5>
- Branda, F. (2026). When artificial intelligence shapes the way we think. *Philos Technol*.

- Chiriatti, M., Ganapini, M., Panai, E., et al. (2024). The case for human–AI interaction as system 0 thinking. *Nature Human Behaviour*, 8, 1829–1830. <https://doi.org/10.1038/s41562-024-01995-5>
- Clark, A. (2025). Extending minds with generative AI. *Nature Communications*, 16, Article 4627. <https://doi.org/10.1038/s41467-025-59906-9>
- Fasoli, M., Cassinadi, G., & Ienca, M. (2025). The dark side of cognitive enhancement: A framework for the technologically induced cognitive diminishment. *Journal of Cognitive Enhancement*, 9, 346–359. <https://doi.org/10.1007/s41465-025-00331-7>
- Hernández-Orallo, J., & Vold, K. (2019). AI Extenders: The Ethical and Societal Implications of Humans Cognitively Extended by AI, in *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AIES '19)*. Association for Computing Machinery, New York, pp. 507–513. <https://doi.org/10.1145/3306618.3314238>
- Hutchins, E. (1995). *Cognition in the Wild*. MIT Press.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

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