

Embracing the Ghost in the Machine: A Review of *Thinking with AI: Machine Learning the Humanities* (2025)

Agustín Cosovschi <agustin_dot_cosovschi_at_inz_dot_si>, Inštitut za Novejšo Zgodovino, Ljubljana, Slovenija

DOI: pending

Abstract

Hannes Bajohr's *Thinking with AI: Machine Learning the Humanities* argues that humanists need to test their own concepts against the new phenomena that computer science and engineering throw at them.

The rise of Artificial Intelligence (AI) has led to powerful changes in all spheres of social life, from communication and culture to the realm of labor and economic relations. In the face of such transformations, scholars, artists, and cultural workers have often tended to adopt either an optimistic stance and attempted to integrate machine learning and generative AI into their work, or a radically critical stance that focuses primarily on problems arising from the training processes of such tools. They cite among other issues the lack of transparency in the use of data, the reproduction of social and cultural biases, and the hidden environmental (and unequally distributed) cost of heavy data processing.^[1] In this context, the recent publication of the collective volume *Thinking with AI: Machine Learning the Humanities* by German literary scholar Hannes Bajohr offers an innovative approach to such questions and an appealing call to rethink how we deal with AI and its uses in the field of humanities.

1

While most works draw on disciplines such as history, anthropology, and cultural theory to examine AI, this volume adopts a different angle. The volume can be seen as a response to Critical AI Studies as they have been practiced until now. In the words of Fabian Offert, "One of the reasons why critical AI studies have not had such a big impact is that their frameworks and concepts do not suffice anymore" [Bajohr 2025, 115]. *Thinking with AI* seeks to provide an alternative approach to a conversation between AI and the humanities, not by scrutinizing AI under the lens of the humanities, but rather by using insights and concepts coming from AI and data science to rethink the very nature of the humanities themselves.

2

The volume offers a series of contributions dealing with a wide range of topics from different disciplinary perspectives. A significant part of the contributions can be seen as advancing a dialogue between theories of AI and the philosophy of language. In a chapter titled "Two Autonomies of the Symbolic Order: Data and Language in Neural Nets," Leaf Weathersby criticizes what he calls the "humanistic view of language" [Weathersby 2025, 44] and its refusal to acknowledge true language capabilities in AI. The author argues instead for the notion that LLMs are true users of language, not mere devices of language generation, and puts forward the idea that the realm of the symbolic cannot be seen any more as a uniquely human feature. In a dense and challenging chapter about writing, Mercedes Bunz takes a somewhat more skeptical perspective: Drawing on Jacques Derrida and André Leroi-Gouran, the author posits a distinction between human writing as an exteriorization of thought and writing as the result of processes of calculation of meaning such as the ones that underlie generated writing. For her part, Audrey Borowski draws on Hans Blumentberg's thinking about myths, metaphors, and other forms of non-conceptual thinking to offer a critique of the absolutism of contemporary algorithmic systems, underlining the limits of computability and instrumental rationality.

3

Aesthetics and the theory of media are a major line cutting across the volume. In a chapter titled "A Theory of Vibe," Peli

4

Grietzner draws from information theory and machine learning to reflect upon our perception of moods and vibes in works of art and cultural artifacts. For his part, Hannes Bajohr compares different types of works of art produced with tools for machine learning and stresses the collapse of text and image that is at work in multimodal AI, suggesting that we stop thinking of computers as only working with syntax without engaging in semantics and that instead we recognize these mechanisms as a true intertwining of form and content. By doing so, he puts forward the somewhat bold hypothesis that multimodal AI does produce meaning, albeit in a different form from the human mind. Bajohr's contribution is also complemented by a chapter by Lev Manovich, who suggests seeing multimodal AI media creation as one more stage in a long line of previous forms of artistic creation that have built on media reuse, thus challenging the popular notion that AI creation is less valuable or original than non-generative art. For his part, Fabien Offert scrutinizes how foundation models express history and its meanings in their representations of the past, for instance producing images of the fascist past in black and white. He shows that models such as CLIP or DALL-E tie certain meanings with certain medias in their representation of historical events and thus play a significant role in shaping our perception of the past.

The mechanisms of AI are also the object of several reflections dealing with the philosophy of science and the mind. Babette Babich has written a chapter that interrogates the ontological status of AI by raising the philosophical problem of other minds, that is, the notion that one cannot ascertain the existence of minds beyond one's own. The author draws on Friedrich Nietzsche's insights about the human inclination to project coherence onto disorder and the role of lying and fiction as tools in human perception, which closely relates to our own tendency to perceive artificial agents as if they were true subjects. Nietzsche makes another appearance in Christina Vagt's chapter, which also insists on the importance of forgetting and fiction for cognition. Taking as a starting point the phenomenon of catastrophic forgetting in AI, that is, when artificial neural networks are trained on new data and forget previously learned information and tasks, the author challenges the conflation of computer models with the workings of the brain that stands as the basis of modern neurocognitive approaches, and stresses the problems of thinking about the human mind uniquely through models that follow Boolean principles. Along a related line of inquiry, Orit Halpern's chapter offers an intellectual genealogy of financial derivatives that converges with the history of machine learning and neoliberal economics: By showing how the ideas of Milton Friedman and Friedrich Hayek blended with those of Donald Hebb and Frank Rosenblatt in a common understanding of intelligence as networked, she sheds light on the intellectual kinship that ties contemporary artificial intelligence with the financialization of global capitalism.

Thinking with AI stands out by virtue of its ambition to reframe our understanding of AI, underlining not only its extraordinary and disruptive character, but also and especially how it relates to other, earlier forms of knowledge production and cultural creation that we have come to normalize. Markus Krajewski's contribution fully partakes in that tendency and is perhaps one of the most original pieces of the volume. Krajewski's chapter, titled "Intellectual Furniture: Elements of a Deep History of Artificial Intelligence," delves into the history of assisted thinking and offers a history of diverse devices that have aided scholars in their intellectual endeavors throughout history. By examining cases of desks and cabinets with sophisticated systems of drawers that were meant to organize and guide the thinking of their user, such as Gottfried Wilhelm Leibniz's *scrinium literatum* or Niklas Luhmann's *Zettelkasten*, Krajewski seeks to show that current AI-assisted thinking represents not so much a rupture with respect to previous forms of work, but rather the extension of a long tradition of intellectual collaboration between the human and the nonhuman.

Bajohr's book has the virtue of gathering some of the brightest and most original authors in the contemporary humanities in a very coherent and stimulating collected volume. One of the weaknesses of the book is perhaps the limited novelty of some of the contributions: Certain chapters are essentially newer versions of previously published texts, with some of them even dating to several years before the volume. Bajohr's volume should perhaps be seen not so much as a fully original academic contribution, but rather as the act of taking a stance in the context of contemporary debates surrounding AI and the humanities; as manifesto and a plaidoyer for a different way of thinking about, and with, AI. All in all, it is an appealing invitation to temporarily suspend our distrust before the manifold ontological challenges posed by AI and instead embrace them, using them to reconsider what we know, and how we know it.

Notes

[1] For a classic critique of LLMs, see [Bender et al. 2021].

Works Cited

Bajohr 2025 Bajohr, H. (ed.) (2025) *Thinking with AI: Machine Learning the Humanities*. Open Humanities Press. Available at: <https://www.openhumanitiespress.org/books/titles/thinking-with-ai/>

Bender et al. 2021 Bender, E.M. et al. (2021) "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?  " in *FACCT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 610–623. Available at: <https://doi.org/10.1145/3442188.3445922> **Weathersby 2025** Weatherby, L. (2025) "Two Autonomies of the Symbolic Order: Data and Language in Neural Nets", in H. Bajohr (ed.) *Thinking with AI: Machine Learning the Humanities*. Open Humanities Press. Available at: <https://www.openhumanitiespress.org/books/titles/thinking-with-ai/>

Weathersby 2025 Weatherby, L. (2025) "Two Autonomies of the Symbolic Order: Data and Language in Neural Nets", in H. Bajohr (ed.) *Thinking with AI: Machine Learning the Humanities*. Open Humanities Press. Available at: <https://www.openhumanitiespress.org/books/titles/thinking-with-ai/>



This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License.