Journal of Posthumanism
Winter 2024
Volume: 4, No: 1, pp. 39–41
ISSN: 2634-3576 (Print) | ISSN 2634-3584 (Online)
journals.tplondon.com/jp

TRANSNATIONAL PRESS® LONDON

Received: 2 September 2023 Accepted: 31 March 2024
DOI: https://doi.org/10.33182/joph.v4i1.3155

Freed, S. (2019). AI and Human Thought and Emotion. CRC Press.

Amanda Hsu Yuk-kwan[1]

By the time Sam Freed's *AI and Human Thought and Emotion* came into print in 2019, OpenAI had already launched GPT-1 and GPT-2 following the Google Brain team's publication of "Attention Is All You Need" (2017). These releases heralded the Transformer era where large-scale language models such as GPT continue to achieve breakthroughs in natural language processing technologies. Sam Freed's book, however, commences with a provocative statement that the development of AI technology is not as trail-blazing as it may appear, due to the absence of conceptual revolution. Freed pinpoints that the concept of neural networks, which renders the AI unsupervised deep learning possible, can indeed date back to the 1940s and 1950s, if not earlier. However much ChatGPT's conversational prowess continues to hold us in awe, Freed argues that an anthropomorphisation of AI remains endemic in science fiction to date. Freed attributes the present non-existence of humanised artificial intelligence to AI architects' fixation with the ideal/rational AI model under the hegemony of scientism across multiple academic disciplines. Succeeding the counter-Enlightenment tradition, he contends that such religious belief in scientific rationalism has caused developers' reluctance to venture into the uncertain and uncharted realm of human subjectivity. Drawing heavily on Husserlian and Heideggerian phenomenology, Freed proposes that AI technical design could instead be grounded on the notion of human introspection at the stage of discovery, in order to approximate human beings' non-linear, nebulous thinking processes whilst engaging with the surroundings consciously. This conception of anthropic AI—contrary to the "western, modern, well-trained, and adult intelligence" (Freed, 2019, 99)—is what Freed aims to achieve, via criticising the rationalist and positivist traditions, seeking recourse from phenomenology, and by legitimising the application of introspection to the programming of artificial intelligence.

Freed ascribes scientism to the human pretence of being able to acquire a thorough understanding of every matter through rationalisation, in contradistinction to the anxiety of exposing the disorder of subjectivity. Freed provides genealogical accounts for the ideological positionalities informing contemporary AI technology developments, from post-French Revolution Comtean positivism, the post-Great War logical positivism arising from the Vienna Circle under the influence of Wittgenstein, to Rudolf Carnap's further development of the Anglo-American analytic philosophy. Against this historical backdrop, Freed positions the rise of behavourism in the early 20th century, after John B. Watson turned psychology into a scientific investigation of humans' external behaviour at the expense of subjective experiences. Freed underscores that, despite the later revolutionary emergence of cognitive psychology under Chomsky's influence which revived the interest in human

[1] Amanda Hsu Yuk-kwan, Ph.D. Candidate, Department of English, The Chinese University of Hong Kong.
E-mail: suis.amanda@gmail.com

mental processes such as language acquisition, the backbone of cognitive psychology remains the rationalist belief that the human mind is a machine. He further highlights that Herbert A. Simon—the pioneer of the mid-20th century cognitive revolution—inherited Watson's behaviorist negligence towards subjectivity and introspection. Freed does acknowledge Simon's significant role in the development of AI's heuristic abilities thanks to his more realistic understanding of human nature, which can be seen in his economic theories concerning the human tendency to compromise at the cost of the best available choice after considering various factors and limitations. Simon's lasting rationalist legacy within the AI community, however, perpetuates the exclusion of human subjectivity from AI design, hence the current dominance of an ideal/rational AI paradigm. In Freed's view, the ideal/rational AI is a failed simulation of human intelligence. This is underscored through the stark contrast between the involvement of massive data in the training of artificial intelligence for high accuracy and the fact that human thinking often operates through generalisations based on limited information. Over-emphasis upon logic, objectivity, correctness, and performance optimisation, according to Freed, deprives AI practitioners of insights about how human subjectivity and fallibility may contribute to the further growth of AI. This, as a result, stands to limit the potential of AI in performing tasks which involve human-machine reciprocity, or which require the machine to fully comprehend the needs of untrained human users, such as the elderly under the care of an AI system. In this regard, Freed turns away from the Anglo-American analytic philosophical traditions to Continental phenomenology to explore the possibility of building his anthropic AI.

In chapter two, Freed's phenomenological approach to AI enters a thought-provoking dialogue with Hubert Dreyfus's critique of AI. Freed shares Dreyfus's view from *What Computers Can't Do* (1972) that current AI fails to replicate human intelligence, owing to its incapability of manifesting a Heideggerian "being-already-in-a-situation" (202). Freed's anthropic AI project can therefore be construed as an attempt to negotiate between the humanities and science disciplines via proposing to formalise what is traditionally considered unformalisable. He next recuperates Husserl's definition of human consciousness as a self-inspectable phenomenon, particularly regarding how human individuals subjectively experience the world from their first-person point of view. Drawing on Husserl's theorisation, Freed suggests that the introspection of such phenomena should entail distancing oneself from various ideological interpellations, despite Heidegger's argument in *Being and Time* (1962) that it is impossible to detach a human individual from his or her situatedness in the world given that one is "thrown" into it (174). Freed revisits Watson and Simon's expurgation of introspection from the realm of cognitive science, arguing that the kind of introspection they reject is the one that is mediated and reported by the psychologist; their acceptance of an untrained human subject's self-report of non-linear train of thoughts as raw data in fact bespeaks the accommodation of human subjectivity in scientific cognitive psychological studies. Freed thereupon resolves the conflict between the study of the subjective and that of the objective to lay the foundation for his anthropic AI programming.

In his delineation of how the human act of introspection could possibly be coded, Freed differentiates his initiative from the Blue Brain project and reiterates multiple times that his concern is mainly technological; in other words, the scientific pursuit of accuracy and truth is unrelated to his thesis. Unlike regular programmers who typically think in terms of source codes in the first place, Freed suggests that anthropic AI programmers see natural thinking processes as they are before determining what codes to use. Freed employs existing programming languages—such as SQL and Python—to brainstorm ways to introduce such concepts as suboptimal solutions and mistakes

whilst constructing his non-Boolean anthropic AI. For instance, he proposes to replace the conventional zero-or-one Boolean binarism with Lotfi Zadeh's fuzzy logic, by assigning numerical values to adverbs of degree, such as "very" and "somewhat," so that the algorithm can handle ambiguous ideas—such as whether 10 degrees Celsius is warm or cold, a query which straddles two sets of concepts. The goal is for the AI to beat its own learning path through fuzziness, confusion, or even mistakes, such that the algorithm can leverage its own knowledge for problem-solving or decision-making in a more human-like manner. Although Freed does not include a comprehensive pilot notebook detailing every single piece of code for producing the proposed anthropic algorithm, the coding examples and strategies he provides in chapters 10 to 12 are sufficient to serve the purpose of inviting the AI community to critically reflect upon their over-reliance on rational methods and their disregard for the significance of human subjective experiences within AI development.

The salient contributions of Freed's *AI and Human Thought and Emotion* are twofold. On the one hand, Freed's call for the inclusion of creative writers and literary experts in the traditional AI team echoes C. P. Snow's (1959) critique of the segregation of arts and science. Freed's bold anthropic AI proposal alerts his readers to the possibilities that the humanities discipline may bring to future AI advances. Notwithstanding Freed's attempt to demarcate technology from science to justify the futility of fully replicating the cognitive system of the human brain, his questioning of the conventional exclusion of human subjectivity from science and technology prompts us to rethink the role of humanity in the field during the Anthropocene where there are rising posthuman concerns. The ethical issues which may arise from the interactions between the anthropic AI and human beings, both of whom are constantly learning through trial and error, will certainly be the next question to be addressed.

References

Dreyfus, H. L. (1972). The limits of artificial intelligence. In *What computers can't do: A critique of artificial reason* (pp. 197–217). Harper & Row.

Heidegger, M. (1962). *Being and time* (J. Macquarrie & E. Robinson, Trans.). Harper. (Original work published 1927)

Snow, C. P. (1959). *The two cultures and the scientific revolution*. Cambridge University Press.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, & R. Garnett (Eds.), *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)* (pp. 6000–6010). Curran Associates Inc. https://dl.acm.org/doi/10.5555/3295222.3295349