

Review of: "Re: Teleology and the Meaning of Life"

Inman Harvey¹

¹ University of Sussex

Potential competing interests: No potential competing interests to declare.

Qeios review of paper submitted by Osamu Kiritani

First Impressions

This paper is abnormal. It is abnormal in content, in structure, in the way it is framed for the reader to interpret. It is not amenable to the normal conventions of review for academic papers – and indeed I guess this provocation may be central to the aims of the author.

On the face of it, the paper takes as a starting point a paper ('TML' – of which more below) published in 2012 by the same author Osamu Kiritani (OK). It then posts the response by some version (presumably recent) of the chatbot ChatGPT to the prompt of 'argue against' the provided text of TML.

Following this response – which I find impressively coherent for a chatbot – OK chooses to elicit counterarguments with the fresh prompt of 'argue against' the first response.

Then in a final brief paragraph OK states that he had no personal input to the paper in amending any chatbot response. He closes with the question: were the chatbot responses solely attributable to ChatGPT, or instead attributable to a collaboration between OK and ChatGPT? The human author OK does not offer an answer, but in footnote 3 presents 3 brief answers offered by ChatGPT.

How should one assess this?

At first sight OK might be accused of imitating those lazy students who once used to write their essays with cut-and-paste from Wikipedia and nowadays – to the despair of their teachers – use cut-and-paste from ChatGPT. But those lazy students try to disguise their lack of personal input, yet here OK makes this fact very explicit, indeed this fact is the focus of the paper. Why?

The abnormal style of the paper defies a conventional interpretation. One needs to stand back and reassess what the context is. One may recall Robert Rauschenberg's 1951 *White Paintings* (Russell, 2014) where the artist (or his assistant) covered the whole canvas with normal house white paint. Rather than seeking any meaning in the surface of a blank canvas that has no external references, one has to consider relationships between the canvas, its framing in the gallery space, the viewers – including oneself. You are invited or provoked into 'stepping outside the box'.

The canvas here is not blank, there are two external references: the seed paper TML, and the role of ChatGPT that is the

subject of much current debate.

TML

If one only considered that final unanswered question – ‘is any original contribution of this paper solely to be credited to ChatGPT, or instead to a collaboration between OK and ChatGPT?’ – then the actual content of TML, the 2012 paper by OK used to seed this project, would be basically irrelevant; the same issues would arise with any paper XYZ. One could discuss the role of prompts in ChatGPT, the enormous corpus that ChatGPT was trained on, the role of the text TML in the original prompt that OK provided, and the role of subsequent prompts. But since the paper does not pursue such threads, maybe we should indeed consider the content of OK’s TML, ‘*Teleology and the meaning of life*’.

OK asserts ‘The aim of this paper [TML] is to give a biological explanation to the meaning of life. I argue that the meaning of life is survival and reproduction, appealing to the teleological notion of function in philosophy of biology’. On reading this paper I am continually frustrated by the repeated use of the phrase ‘the meaning of life’ as if it has some unambiguous universally agreed sense – but it does not! Eg alternative possible senses include:

- The meaning(s) of the word ‘life’ in the english-speaking community – consult a dictionary!
- The distinction between living and non-living systems as classified by biologists
- Does some aspect A of the world of a living system B have meaning for B, and how could a third party C know that is the case? (with various versions of A, B, C)
- E.g. does a source of food A have meaning for B1 = a bacterium? Or for B2 = a human?
- E.g. does *its own life* have meaning for B1 = a bacterium? Or for B2 = a human?
- Etc etc

These all imply a great range of different *meanings*, and I struggle to decide which is intended. Just conceivably the author was referring to some version of the issue that I phrase differently as: what does it mean, what does it take for a system to have *motivations*, for things in their world to matter to them such that they direct their actions accordingly. What does it take for such a system to be an *Agent*, responsible for its own actions?

Evolved organisms have motivations, ChatGPT does not

I hope this rephrasing in terms of *motivations* does indeed cover the focus of TML, because if so then much of TML can be seen as fairly closely in agreement to my own views appearing in Harvey (2024 forthcoming). There I argue that there is a major issue in grounding any attribution of inherent *motivation* to a system – whether a biological organism or some AI system. If a system is seen to be performing action X, how can we tell whether it is doing so because it *wants* to do X? Who is *responsible* for that action: the system itself (our likely answer for a biological organism); or the system’s designer (a likely answer for an AI system); or nobody (perhaps for a dust-devil blowing in the wind)?

Just as Hume (1739) argues that one cannot derive an “ought” from an “is”, one cannot directly derive or ground a motivation solely from the physical description of a system – with just one exception that I am aware of. Where systems such as biological organisms are the outcome of aeons of Darwinian evolution, the current products of that evolution have

necessarily descended from many generations of *survivors*; without exception, all their ancestors survived long enough to pass on genetic material. In my view this gives ground for ascribing the basic motivation of a *survival instinct* to such products of evolution. And from this basic motivation we can derive further motivations such as to seek food, escape danger, etc.

So here there are clear parallels between my own arguments (Harvey 2024 forthcoming) and those of TML – assuming we can relate the phrase there “meaning of life” to my grounding of motivation. We agree on grounding in survival and reproduction in the context of evolution, and indeed I will agree with TML’s appeal to the teleological notion of function in the philosophy of biology. I go further to argue that AI systems such as ChatGPT, that (in their current form) have no such evolutionary history simply do not have *their own* inherent motivations. At most they are tools for performing in accordance with the motivations of their human overlords.

This neatly brings us back full circle to a key question central to the review of this paper: to what extent is ChatGPT to be considered the author? From my perspective, ChatGPT has no inherent motivations of its own, it is no more than a (very technically impressive) tool for carrying out what it is commanded to do by its human overlords. These overlords include the designers of ChatGPT, the human authors whose texts contributed to its training, the curators who selected these texts; and last but not least, the human (here OK) who chose the prompts. For what it is worth, when I ask ChatGPT whether it is the author of its statements, or whether it has motivations, the responses are consonant with my own views. ChatGPT is inherently incapable of authoring this paper, of initiating – unprompted — the prompts that triggered it.

Elegant and profound

As with Rauschenberg’s *White Paintings*, the contribution of OK the human author is deliberately minimal but nevertheless key. As with the *White Paintings* it is witty, elegant, provocative, it made me smile, it has profound depths. And if, as I hope, the author’s concerns about the ‘meaning of life’ (a phrase that I find very unclear) can be matched to my own concerns about motivations, then this text points out the unmotivated nature of ChatGPT, without explicitly expressing this. “Whereof one cannot speak, thereof one must remain silent” – as someone once spoke, ironically.

References

Harvey, Inman (2024 forthcoming). Motivations for AI, for Deep Learning, for ALife; Mortality and Existential Risk. *Artificial Life* (accepted for publication 18 Oct 2023)

Hume, D. (1739). *A Treatise of Human Nature*. Book III, Part i, Section I. London.

Kiritani, Osamu (2012). Teleology and the Meaning of Life. *Journal of Mind and Behavior* 33 (1-2):97-102.

Russell, D. (2014). Mindfulness in art: Robert Rauschenberg’s *White Painting*.

Hektoen International *A Journal of Medical Humanities*:2014 6(3).

<https://hekint.org/2017/01/23/mindfulness-in-art-robert-rauschenbergs-white-painting/>

