Qeios

Peer Review

Review of: "Superintelligence: Identification of Friend or Foe Future Landscape of Cooperation with Non-human Intelligence"

John Dorsch[1]

1. Center for Technology and Environmental Ethics, Prague, Czech Academy of Sciences, Prague, Czech Republic

Overall Assessment

After reviewing the abstract and introduction, I have concerns regarding whether this paper meets the standards of peer-reviewed research in a reputable journal focused on the philosophy of cognitive science or technology. The paper frequently makes broad claims that require substantiation but lacks supporting evidence. Additionally, it introduces technical concepts without sufficient clarification or explanation of their relevance, making it difficult for the reader to follow the author's reasoning. The paper would benefit from a clearer articulation of its main research questions and methodological approach. Moreover, it does not engage with key literature in the ethics of technology, which is essential given the subject matter. Finally, the writing includes linguistic inconsistencies and instances of opaque language that impede comprehension.

To enhance the paper’s clarity and rigor, I suggest that the authors reflect on the following foundational questions:

1. What is the central research question?
2. What is the thesis and main argument in response to this question?
3. How can the argument be structured in a clear and compelling way?

In addressing these questions, the following steps could help strengthen the paper:

a) Engage with existing literature that addresses the research question. This will ensure the argument is situated within the broader academic discourse and acknowledges prior contributions. It would be

 doi.org/10.32388/DLPR14

particularly beneficial to integrate relevant discussions from the field of technology ethics.

b) Clearly differentiate the paper's argument from existing positions in the literature. This will help establish the paper's original contribution and should be outlined in the introduction.

c) Present arguments in a structured, step-by-step manner. Consider breaking down the reasoning into logical steps, akin to a formalized argument, and seek feedback on potential weaknesses.

d) Consider the intended readership. If aiming for accessibility to a broader audience, concepts should be introduced with sufficient background information. If assuming prior expertise, the introduction should clarify this expectation and reference sources where readers can acquire necessary background knowledge.

Review of Specific Sections

Abstract (Score: 3/10)

The abstract lacks clarity, contains linguistic inconsistencies, and includes statements that require further support. Specific concerns include:

- The statement that *"non-human intelligence is universally acknowledged by all participants of discussion as a necessary element of any consciousness, regardless of its nature"* is problematic.
  - It is unclear who the *"participants of discussion"* are. What specific debate or academic discourse does this refer to?
  - The claim itself is inaccurate. There are philosophical positions, such as panpsychism, which argue that consciousness is fundamental and does not necessarily depend on intelligence. Similarly, some bioenactivist views hold that consciousness is intrinsic to life, even at a cellular level, without requiring intelligence in a conventional sense.
- There are linguistic inconsistencies, such as missing articles and unclear phrasing. For example:
  - *"The shared modus of intelligence evaluation [...] offers promising direction"* should be *"a promising direction"* or *"promising directions."*
- Some sentences are difficult to parse due to unnecessary complexity. For instance:
  - *"However, this approach's successful resolution of an objective basis for intelligence studies unveils inescapable challenges."*
    - This phrasing is vague and difficult to interpret. Consider rewording for clarity.

- The research questions should be more explicitly stated. It is somewhat clear that the paper intends to compare computational capability with intelligence, but the phrasing *"the potential for higher intelligence to exert adverse effects on less intelligent counterparts"* is unclear. What does it mean to compare something to *a potential outcome*?
- The relevance of certain statements is unclear. For example:
  - *"It is conceivable that pure intelligence, as a computational faculty, can serve as an effective utilitarian tool."*
  - Why is conceivability important in this context?

Introduction (Score: 4/10)

The introduction does not adequately introduce the necessary technical concepts, presents sweeping claims without justification, and does not engage with relevant literature on AI ethics, despite making ethical assertions.

- Some statements are unclear or oddly phrased. For example:
  - *"The obligatory development of a highly intelligent life is debatable."*
    - What does *"obligatory development"* mean in this context? If the intention is to question whether development is necessary for intelligence, this should be stated more explicitly.
  - *"There is still a possibility of arguing about primary non-human intelligence, which raised human intelligence and, consequently, neural network-based AI."*
    - The term *"primary"* is ambiguous. Additionally, the idea that intelligence *raises* intelligence and AI is unclear and should be further explained.
- The introduction has an inconsistent register. The initial paragraphs contain definitions and broad descriptions of living systems, followed by references to highly technical concepts without sufficient explanation. For example:
  - The mention of *"Solomonoff–Kolmogorov–Chaitin complexity"* assumes reader familiarity without providing context for its relevance to the discussion. Instead of reiterating fundamental biological facts, the paper would benefit from dedicating space to introducing and justifying the inclusion of technical notions.
- Some claims require further justification. For example:

- *"Until it becomes autonomous and self-serving on a sufficient scale, there is no Darwinian type of danger from AI for humankind."*
    - It is unclear what *"Darwinian danger"* refers to in this context. Many researchers in AI ethics argue that AI already poses significant risks, even without full autonomy.

- The use of technical terms such as *"autopoietic"*, *"allopoietic"*, and *"poietic"* requires clear definitions for readers unfamiliar with these concepts.
- The assertion that *"any highly developed intelligence is inherently moral, and higher levels of development translate into a more benevolent stance"* is particularly problematic.
    - There is no clear philosophical or empirical basis for this claim. Does the author suggest that intelligence and benevolence are intrinsically linked? If so, this argument should be explicitly developed and defended.

Conclusion

The paper presents an interesting and ambitious discussion, but its current form requires significant refinement. The main areas for improvement include:

1. Clearly stating the central research question and structuring the argument logically.
2. Engaging with relevant philosophical and ethical literature, particularly in AI ethics.
3. Providing clear explanations for technical concepts rather than assuming familiarity.
4. Avoiding sweeping or controversial claims without proper justification.
5. Improving clarity and precision in language to enhance readability.

I encourage the authors to refine their argumentation, engage more deeply with existing research, and clarify their key claims. Doing so will significantly strengthen the paper's contribution to the discourse on intelligence and AI.

Declarations

Potential competing interests: No potential competing interests to declare.