

Review of: "“The Forbidden Planet”: AI and Psychology: Prepare and Sound the Alarm!"

Shoji Nagataki¹

¹ Chukyo University

Potential competing interests: No potential competing interests to declare.

This article is excellent in that it presents a clear and comprehensive picture of the future that the development of AI will bring to mankind, with references to science fiction movies and so on. The various examples the author mentions are also very interesting and contribute greatly to the success of this article as reading material.

Let the reviewer's view be expressed along with the author's examples. As the author says, delegating to AI, AI-equipped humanoids, and avatars the roles that have been played by psychotherapists, lovers (or sex partners and sex workers), etc., can cause serious problems in certain cases. This is because their activities involve a psychological and physical penetration of the human being's innermost being. The technological issue here is the extent to which the technology to realize VR with not only visual perception but also tactile and olfactory sensations can be achieved. Of course, the development of such technology will involve not only technological issues but also philosophical and ethical issues.

Furthermore, AI will be able to easily reproduce the intellectual and aesthetic heritage that humans have historically created, such as literature, philosophy, art, criticism, crafts, etc., by converting them into data. We will no longer be able to distinguish between machine-created art, literature, and philosophy and their human equivalents. Another possibility is to collaborate with AI. In Japan, the most recent Akutagawa Prize-winning author confessed that she used ChatGPT for some of her award-winning works. But here, too, there is another problem: humans must resist the erosion of their own creativity by AI. Perhaps the future is bright for AI, but as the author thinks, the future for humans might be in stark contrast.

In the following, the reviewer would like to point out some questions in line with the author's description.

The author writes:

I have concerns that the development of AI does in fact pose a threat to humanity, but my concerns... stem... from ... the now almost certain likelihood that AI will unravel the activities, relationships, creative ventures, and institutions, and very personhood that from the dawn of civilization have constituted our humanity.

In this context, what does the author mean when he says that AI will unravel “very personhood”? The reviewer would ask the author to develop this issue in some more detail. The reviewer would like to know what philosophical implications the author puts into the word “personhood.”

For example, the author provides the following example.

We enter digitalized worlds and interact with digitally created human facsimiles, and/or digital avatars of actual human beings, living and dead. Especially in the latter case, whereby individuals spend considerable time in a virtual second life, with virtual AI friends, lovers, and avatars, the nature of the self stands to be radically transformed.

The author writes, "AI will unravel very personhood," and "(by AI) the nature of the self stands to be radically transformed." I think that these two sentences mean almost the same thing.

Why does a virtual second life come to radically transform the nature of the self? Is it different from immersing oneself in a certain hobby? For example, people enter into fantasy worlds by reading fantasy novels - I did so when I was in junior high school. But everyday life is different from that. The second life is a virtual one, and most rational people spend a certain amount of time there before returning to their daily lives. Virtual life gives us a cathartic effect. Why does the author think it can be more than that? One problem is that people may show a strong addiction to virtual technology, which could be solved by developing a medical treatment for addiction, as is the case with alcohol and drugs. The only concern if the reviewer takes the author's viewpoint is that if the tactile sensations can be fully realized virtually, addiction may become uncontrollable. In Japan, there is a phrase that tactile art drives people crazy. In other words, collectors of touchable works of art, such as tea utensils, may tend to lose their sense of money when they come into contact with a work of art that they love.

The author also states something to the effect that AI technology will bring the dead back to life. The author says, "AI will pierce the boundary between the dead and the living and between life and death." However, the reviewer is skeptical about this. The author's point is only that the dead will feel as if they are alive for the people who still live. Such experiences have been held in the past by human beings through dreams, hallucinations, and religions. Here the problem is "my death." No matter how much AI technology advances, we may never be able to overcome this first-person problem. The reviewer would like to ask the author's view on this issue.

Toward the end of this article, the author expresses his pessimistic opinion as follows.

Human beings will see "soul" in all of this machinery [AI activities] and be moved by the AI poetry, drama, music, and art—and perhaps the poetry, music, and art will be "moving"—but we will be fooled, just as many are fooled by psychopathic lovers who have no feelings.

As history teaches us, human beings are creatures who find a soul in a carved wooden idol or even in a piece of stone or wood. And humans have been betrayed by those fakes many times. In other words, the human imagination is such a rich one. On the other hand, I do not believe that AIs become necessarily psychopathic lovers. Such optimism is one way to prepare for the future. For, despite the warnings of people like the author, technological progress will not stop, even if it slows down temporarily.

The critical issue for the reviewer is to conceive of an extended society that includes “AI as moral agents” as our members, in preparation for the forthcoming future that the author has so richly depicted.

*The following sentence appears twice in this article, what is the author's intention?

By creating and subjecting itself to the idea of “artificial intelligence” and the temptations of AI simulacra, is humanity walking down the path to its own demise?