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

Tal Shafir[1]

1 University of Haifa

The author of this article describes some current and potential future effects of the development of artificial intelligence (AI) on various aspects of our lives and raises the concern that this development will destroy humanity. The author further asserts, "that the emergence of artificial intelligence and its blindingly rapid development poses the single greatest challenge to "psychology", a challenge both to the "psychology" of each and every one of us, and also to psychology as a profession dedicated to the understanding and healing of the human psyche".

While the author's predictions are interesting and intriguing, in my mind, he fails to explain how or why these predictions are troublesome and constitute a real threat to our psychology or to humanity. For example:

Artificial Truth: While the increasing number of deep fakes certainly raises concerns, tools and algorithms for detecting deep fakes and differentiating between real and fake photos, videos, or sounds (e.g., speech) are constantly being developed, and some exist already today. There is a high probability that improved AI "deep fakes" will be countered by the ongoing development of increasingly better and more sophisticated tools for detecting and/or preventing the possibility of creating "deep fakes", similar to the race today between hackers and those who secure digital information. Eventually, we will probably just learn to live with deep fakes and protect ourselves from their effects by constantly being aware of the possibility that any information we consume might be fake and always taking appropriate precautions when consuming information.

AI Relationships: The author is afraid that many of our relationships will be with robots or avatars instead of with real people. However, he fails to explain why having relationships with non-human beings that might satisfy our psychological needs is a bad thing that endangers our psyche and humanity. Many people today have close relationships with pets, who help them alleviate their loneliness, but to the best of my knowledge, no one has claimed that such relationships constitute a threat to humanity or to the psychology of those pet owners. It is not enough to just describe the futuristic relationships with non-human beings; the author should go one step further, describe the potential psychological implications of such relationships, and clarify why such potential implications are alarming.

Work: Following the industrial revolution, many manual professions ceased to exist or almost disappeared, and technological advancement brought demand for new professions that never existed before, such as computer programming or hardware engineering. A similar process might follow the "AI revolution." The fact that AI might replace some existing professions such as medicine, law, or psychotherapy and make them obsolete doesn't necessarily mean

that people will stop working for their living or that humanity will be destroyed.

Creativity: While AI can probably create music, literature, and visual art in different styles faster and often better than human beings, it can only do so based on the data that it was "fed," i.e., based on previous creations of human artists. Thus, AI essentially copies, albeit in a very complex and sophisticated way, the creations of previous human artists. In that sense, AI is more like a craftsman than a real artist and doesn't really endanger true creativity. It is also possible that with time, AI will become just a tool, and creativity will be expressed in the way people use this tool to create new forms of art.

Death: It is hard to imagine that people will converse with a hologram of a dead person without knowing it is a hologram. Such conversations will be similar to conversations people sometimes conduct in their imagination with their dead relatives, while knowing the person is dead, but imagining what s/he would have said in a specific situation. Therefore, such conversations will not really pierce the boundary between life and death, as the author suggests.

AI and the end of "humanity": The author raises the concern that futuristic AI-generated beings (robots, holograms, etc.) might not really feel and will not have a real soul, but we, human beings, are not well equipped to distinguish a simulated soul from a real soul. Although I can intuitively feel the anxiety that such a situation might cause, when thinking about this logically, the author doesn't really explain the potential philosophical, ethical, or psychological implications of not distinguishing a simulated from a real soul, and why such a situation endangers our psychology and/or humanity. Based on neuroscientific findings, we know today that our brain doesn't really differentiate between its perception of the external environment vs. the internal environment (i.e., our thoughts, imagination, and beliefs), in terms of their effects on the brain. Thus, from the point of view of our brain, the external real situation, i.e., whether AI has a real or simulated soul, doesn't really matter: as long as we perceive AI as having real feelings, our experience of these feelings will be the same, whether they are indeed real or simulated. Thus, playing the devil's advocate, I would ask: if we cannot differentiate between a real and simulated soul, and if we perceive a simulated soul as real, what difference does the real situation make, and what's bad about always perceiving AI as having real feelings and a real soul?

In sum, although the author describes some interesting futuristic AI effects, his failure to relate to possible counterarguments, or to delve more deeply into the potential ethical or psychological drawbacks resulting from those futuristic effects, makes the article somewhat one-sided and not entirely convincing. Moreover, adding some suggestions for how to answer the questions he raises or how to solve the concerns he brings up will add another layer to the discussion of this topic and will make the article more whole and more interesting.