Research Article

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

Sanford Drob[1]

1. Fielding Graduate University, Santa Barbara, United States

It is argued that with the rapid development and proliferation of AI, the field of psychology must be prepared to contend with transformations in relationships, sexuality, work, the arts, the self, and even death that will threaten to unravel the activities, relationships, creative ventures, institutions, and very personhood that comprise our humanity. As a profession, it is our responsibility as psychologists to prepare for these radical transformations, participate fully in the cultural and political discourse on AI, and sound the alarm regarding the dire psychological consequences of unregulated AI development.

In the 1956 American science-fiction film, *Forbidden Planet*, we learn of a civilization on the distant planet Altair that had perished 200,000 years ago. The inhabitants of this planet, the Krell, had developed an ultimate machine that enabled them to enormously augment their intelligence and to effectively materialize anything through the exercise of will, thought, or imagination. As the film progresses, we come to realize that it was precisely this machine that destroyed the Krell and their civilization, as it provided their subconscious with unlimited power and resulted in "monsters from the Id." A scientist, Morbius, who was the last surviving crewmember of an ill-fated expedition to Altair from Earth and who had found the Krell's machinery fully operational, utilizes their machine to achieve super intelligence. In the process, Morbius unleashes these same monsters from the Id, which, one by one, kill the members of a rescue party sent from Earth. Morbius finally comes to realize that it is his own subconscious, augmented by the Krell technology, that is causing the destruction. He surrenders his obsession with Krell science, and after departing from Altair, the captain of the rescue mission orders the destruction of the planet, and with it, the Krell scientific knowledge and technology. At the close of the movie, we learn that in a million years humanity will reach the scientific and technological sophistication of the Krell.

The writers of "Forbidden Planet" did not, in 1956, fathom that less than 70 years later, within the lifetime of some who saw this film when it first opened in theaters, we would be on the verge of achieving a technology that could deliver a machine of superintelligence with the potential to absolutely transform, if not destroy, humanity.

Like many others, I have concerns that the development of AI does in fact pose a threat to humanity, but my concerns do not stem so much from the possibility that AI will take over, run amok, and destroy the human species (although this possibility can hardly be ruled out), but rather from the likelihood that AI (like the Internet) will magnify humanity's destructive instincts and 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.

It is for this reason 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.

In this presentation, I will, in broad strokes, outline some of the challenges to psychology in both of the above senses of the term, and I will discuss some of the philosophical issues raised by these challenges and how these issues relate to the practice of our profession. I will organize my discussion around themes that are psychologically relevant. While focusing on near-term issues, I will also touch upon some longer-term AI possibilities that, while not immediately on the horizon, could easily be realized well within the lifetime of individuals who are now living.

I believe that it is inarguable that artificial intelligence is on the brink of changing our relationships, sexuality, work, creativity, and the fundamental nature of our individual and collective human identities. It has already begun to dissolve the fundamental distinctions between truth and falsehood, reality and fantasy, life and death, and man and machine. The evidence for these assertions is available in the AI technologies that are already developed and being distributed, that are currently in development, or that are clearly within reach of our current scientific and technological capacities.

Artificial Truth

The artificialization of truth is rapidly accelerated by the capacity of AI to create "deep fakes," including convincing photos of people and things that do not exist, depictions of events that never occurred, videos of identifiable individuals speaking in their own voices reciting words that were never theirs, and

AI "hallucinations," the tendency of chatbots to occasionally produce authoritative sounding but nonetheless fictionalized facts. A recent article in *The New York Times* invited readers to determine which of ten faces were photographs of actual persons and which were AI-generated, and noted that in most instances human observers more often judged the AI photographs (as opposed to the actual ones) to be real, and that those whose judgments were more confident were most likely to be wrong.[1] (I myself, confident in my discernment, obtained a score of 20% on this test).

AI Relationships

AI technology now has the capacity to create computer images of people who interact with us, come to know us, and express "feelings" about us in an increasingly convincing manner. What began as an application that could provide "company" for isolates and the elderly is rapidly expanding into the business of AI friends, boyfriends, and girlfriends. A brief documentary that appeared in the *New York Times* is a witness to the emotional connection and vicissitudes of women in China who become involved with such AI boyfriends.[2] Some AI applications provide the user with the opportunity to design their friends' gender, age, appearance, and personality. In due course, perhaps within the next several years, these applications will take advantage of holographic images and/or virtual reality experiences that will enable the user to experience their AI friends as fully and even tactilely present. Already, there are AI robots being marketed as sexual objects/companions, and we may only be several years away from applications that enable us to interact with holographic AI lovers. When this occurs, pornography will be transformed into virtual sexuality. One can imagine a time when individuals prefer to have virtual sexual experiences with fantasized lovers who have the precise appearance and perform the precise actions that correspond to their sexual fantasies. Perhaps real-world encounters will be limited to professional pornographers or those who remain interested in procreation, and the rest of us will be having sex with virtual avatars. A similar pattern will likely extend to conversations and friendships. One can imagine a time in the not-too-distant future where the majority of the population spends hours each day in virtual reality, just as many today spend hours daily on their smartphones.

Work

The possibility that AI programs will replace us at work is becoming ever more likely. Such replacement is even possible with regard to professions such as law and medicine, though until the thorny problem of "hallucinatory facts" is resolved, complete reliance on AI in the professions will be held at bay. While current versions of AI psychotherapists may appear to be superficial and unimpressive, we must remember that we are only at the beginning of harnessing this technology. Imagine in the not-too-distant future, AI psychotherapy programs, trained on a multitude of transcripts generated by master psychotherapists, whose responses reflect such training, and which turn out to be superior to those that any novice individual psychotherapist can generate. Consider also the possibility that artificial intelligence will perform psychological evaluations and write forensic psychological reports that are more comprehensive and less biased than those produced by flesh-and-blood professionals.

Creativity

Large Language Models are not only producing well-written (if at times faulty) responses to virtually any informational question but are capable of responding to questions requiring mathematical, logical, and practical reasoning about the real world. AI models are also said to generate responses to complex ethical dilemmas that are at a level of 90% agreement with those of ethicists. Chatbots are becoming increasingly capable of writing essays, poetry, and even novel-length fiction. DALL E-2 and other AI art generators are able to produce images in a range of artistic styles in response to brief but specific verbal prompts. As these image applications improve, the work available to graphic designers will shrink, if not disappear. AI will also challenge fine artists who may ask themselves why they should continue when a painting that would require several weeks of intense work can be generated by an AI imagery app in a matter of seconds. While the aesthetic value of AI-generated poetry, fiction, and painting is not yet generally at a level comparable to human creativity, an AI-generated painting has won first prize in an important, (albeit digital) art competition,[3] and it is only a matter of time before the creative products of artificial intelligence are indistinguishable from, and in some ways superior to, those of individual human beings. Significantly, the problem of "hallucinations" is not relevant to AI poetry, painting, and music, as "facts" are not at stake in these endeavors.

It is arguable that all art and literature involves the imitation, processing, and transformation of what has been done in the past, and machine applications are able to imitate, process, and transform past materials at a rate and comprehensiveness that is far beyond the capabilities of the human brain. AI music generation applications are already creating recordings in the style and/or voice of such artists as Drake, Frank Sinatra, and the Beatles, and, unless limited by legislation prohibiting this, it will soon be possible with a few strokes on the keypad to produce entirely new and convincing Beethoven symphonies and Beatles albums. While it is perhaps too soon to say how such products will be accepted by the public, the initial response to such simulated music has been positive. Like artists, composers will be disheartened by the fact that material comparable, and perhaps superior to their own, can be produced by machines in a mere fraction of the time that it takes them to work their craft.

It will be interesting to see when the first AI-generated professional papers on subjects in literature, art, criticism, and philosophy are accepted by peer-reviewed journals. Though the problem of hallucinatory quotes and citations is relevant in this case, this is an area to be watched closely.

Nick Bostrom, a philosopher and futurist, has said that artificial intelligence is the last invention that human beings will ever have to make,[4] suggesting that AI will take over from here.

Death

In one of the early *Superman* movies, the hero retreats to his Fortress of Solitude to consult a holographic version of his deceased father. The technology to create such conversations with the dead is already with us. Extensive videotaping of individuals while they are alive, perhaps augmented by AI processing of writings and interviews with those who know them, will enable the dead to "live on," converse intelligently, and dispense advice and wisdom to the living. We already have applications that enable us to converse with such living persons as Miley Cyrus, and soon our children will be able to do so with their deceased parents. Holographic projections of the dead will enable us to invite them for conversations at our dinner table, and or to interact with them extensively in virtual reality. In this way, AI will pierce the boundary between the dead and the living and between life and death.

AI programs that provide the dying with a virtual experience of religious ecstasy and transport them into a virtual heaven will soon be well within the realm of possibility.

What about religion? ChatGPT writes sermons! Belief in God? The AI revolution has prompted many (including Elon Musk) to conclude that we must already be living in a simulation, in which case our creator and "God" is itself a computer in a deeper level of "The Matrix," a god who/that we are worshiping when we pray to "our father, our king" in church, mosque, and synagogue, what I have described as the "videogame god."[5]

Life in Digital Reality

The philosopher Ned Bostrom has suggested that if at some point in the near or distant future AI technology reaches a point where we are able to digitalize our minds and bodies, and, in effect, download ourselves into an interactive, virtual reality, then there is no assurance that we are not already residing within such a "matrix."[6] Musk has gone so far as to suggest that the vast majority of conscious entities in the universe are digital, and it is thus highly unlikely that we are the natural, biological beings that we believe ourselves to be.[7] He has argued that we ought to *hope* that we are digitalized beings, because if we are not, this suggests that humanity destroyed itself before it was capable of realizing matrix-like applications of artificial intelligence.

I have dealt with the philosophical implications of these possibilities elsewhere,[8] but here I wish to point out that even without fully digitalizing our brains, it will soon be possible for us to spend considerable time, perhaps even the better part of our days, in virtual reality, where we not only have virtual interactions and relationships, but consult with virtual psychotherapists, palm readers, astrologers, etc. Such digitally controlled dream realities will challenge the primacy of the "real world" in a way that dwarfs the similar challenge that we now face with the relatively primitive technology of our current versions of social media.

An important question raised by these possibilities and likelihoods is whether artificial intelligence will *enter our world* in the form of physical, robotic androids, or, what will be more cost-efficient and thus likely, *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.

We already have digital phenotyping as a vehicle for biologically based psychiatric diagnosis, and it will not be long before AI-generated brain stimulation is used as a treatment for psychiatric illness and psychological distress.

The Challenge to Humanity

All of these eventualities pose challenges to psychology that dwarf those created by technological changes in the past, such as the advent of film, radio, television, tape recordings, and social media. As has often been stated, the rapid development of AI poses unfathomable challenges to humanity, challenges that will threaten the very concept and experience of being human.

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?

While many have called for a halt or a slowdown in the development of AI, economic, political, and national defense considerations make it unlikely that we will be able to halt the progression of superintelligence and stop short of the disaster that in *Forbidden Planet* destroyed the Krell civilization on Altair. In contrast to the long-term concerns expressed in that film, as early as 1966, the futurists Roger A. McGowan and Frederick I Ordwat III argued that with

the development of superintelligent computers in the ensuing decades, nations would be forced to cede full control to them or run the risk of being outmaneuvered and subjugated by other nations that had done so.[9]

Psychologists are facing the task of assisting both individuals and society in adapting to what can only be described as a radical posthuman world. I believe that this is the single greatest challenge for our field, and that financial, intellectual, and ethical resources must be channeled in its direction if psychologists and "psychology" are to survive into the posthuman era.

AI and the End of "Humanity"

AI can potentially destroy humanity, but not necessarily by eliminating us physically or taking over the world, but by rendering us impotent with respect to the acts and institutions that make us human. Already, AI programs are creating works of art in seconds that even discerning viewers are unable to distinguish from art created by people. AI is becoming more and more adept at writing essays, poetry, and other narratives, and it's only a matter of time before it will be producing fiction that is indistinguishable from the human product. When will the first AI-written book top *The New York Times* Bestseller list?

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?

The development of artificial intelligence raises a whole host of ethical conundrums that we were only now beginning to fathom. One critical question in this entire morass of change is how to understand and characterize the "intelligence" that is emerging so rapidly within AI. Is this intelligence human-like, or is it a mere "model" or simulacrum? Are the AI friends and lovers which will inevitably become increasingly indistinguishable from actual human beings,

Are our digital friends sentient loci of feelings and values, or are they mere "zombies" that may fool us into believing they are sentient, but which actually have nothing of human experience behind their imagistic and verbal presentations? Do they "feel," have desires and values, or do they, like AI-generated photographs of faces, only fool us into believing that they do?

Are AI "persons" to be accorded a status commensurate with their apparent intelligence and treated with the ethical and axiological respect that we should treat biological human beings? Will the answer to this question change depending on whether AI humanoids enter our world in the form of robots, or we enter the digital world in the form of digital avatars?

The view that AI is essentially soulless, without any possibility of becoming sentient or having genuine feelings and values, has not received the consideration it deserves. There is an assumption that if enough information is processed at a speed that duplicates the cognitive and other functional aspects of mind, consciousness will arise spontaneously, but this is speculative. The hardware that supports artificial intelligence is completely different from that of biological brains, and it may well be that consciousness is uniquely produced through an integration of processes at the biological, chemical, and even quantum level, processes that are simply not present in computers.

When the medieval Kabbalists thought to create a "golem," an artificial human being, they envisioned it comprised of the *Sefirot*, the 10 value archetypes, including wisdom, knowledge, love, compassion, and beauty, archetypes they believed to be both the elements of the human soul and the constituents of the natural world. As Noam Chomsky has suggested, the "golem" called "AI" is an automaton, a probabilistic machine, a simulacrum of humanity devoid of values, wisdom, and, hence, devoid of soul.[10]

Still, there is a very difficult paradox with regard to AI, values, and wisdom. Years ago, I received a telephone call from the Lubavitcher Rebbe's secretary, asking me to respond to the question of whether a woman should reject an otherwise suitable match because he had confessed to having genital herpes. The rebbe knew that I had conducted a study on the psychology of genital herpes, but he wasn't interested in my opinion. He wanted me to ask three other researchers who had published in refereed journals on the topic of herpes. The Rebbe's secretary suggested that the Rebbe would follow the advice of two of the three as he would regard them as impartial. He wanted me to ask the researchers what advice they would give to their own daughters on this question. The rebbe was interested not in my or his own judgment on the question, but rather in the majority judgment of uninvolved experts knowledgeable on the subject. This is precisely the kind of practical reasoning that AI is especially adapted to. Only, instead of asking three researchers, it could comb the entire literature on the question and arrive at a majority opinion. And it could be argued that this would be the best available wisdom on the question. Indeed, it would be the rebbe's methodology writ large. The paradox here is that AI could provide us with wisdom without having any feelings or values itself.

What is frightening is that we will soon have a world comprised of a series of AI machines interacting with us and one another, writing plays and poetry, painting pictures, forging sculptures, composing music, solving mathematical problems, and making scientific discoveries, and then writing articles and

reviews about all of these things—and even creating images of human beings that act as if they are falling in love with us and one another, without an ounce of Soul in any of it. Human beings will see "soul" in all of this machinery 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. Perhaps humanity will even think that it can download itself into a matrix-like simulation and exist eternally, cut off from the original natural world (Elon Musk thinks we are already there!). Human beings are not well equipped to distinguish simulated soul from real soul. We are terrible lie detectors, and we have, from time immemorial, been vulnerable to hucksters, scammers, and seducers. We have certainly not evolved to distinguish digital golems from biological human beings, and it is very possible that we are rapidly being taken in by an image, as if we believed that the people we see on screen in motion pictures are really there in the room, interacting with one another and having consciousness and feelings... But we are not just being fooled; we are basically handing over the reins of the world to this huckster of our own creation, a huckster that will inevitably simulate the best within us but the worst as well, and the worst will be amplified beyond anything we can now imagine. We may not be far from living on a "forbidden planet" and, as psychologists, it is our responsibility to both prepare for this radical transformation and deconstruction of our humanity and sound the alarm.

Footnotes

[1] Test Yourself: Which Faces Were Made by A.I.? *New York Time*s, January 19, 2024.

[2] "My A.I Lover: Three women reflect on the complexities of their relationships with their A.I. companions. By Chouwa Lang. *New York Times*, May 23, 2023.

[3] AI won an art contest, and artists are furious | CNN Business. https://www.cnn.com/2022/09/03/tech/ai-art-fair-winner-controversy/index.html. Downloaded, January 21, 2024.

[4] Nick Bostrom: What happens when our computers get smarter than we are? (singularityweblog.com)

https://www.singularityweblog.com/nick-bostrom-ted/#:~:text=Artificial%20intelligence%20is%20getting%20smarter%20by%20leaps%20and,invention%20that%20humanity%20will%20ever%20need% (Downloaded January 21, 2024).

[5] Sanford L. Drob (2023) Are you praying to a videogame God? Some theological and philosophical implications of the simulation hypothesis, *International Journal of Philosophy and Theology*, 84:1, 77-91, DOI: 10.1080/21692327.2023.2182822.

[6] Nick Bostrom, 2003. Are You Living in a Computer Simulation? *Philosophical Quarterly* 53 (211): 243–255. Available online: http://www.simulation-argument.com/ (accessed on January 21, 2024).

[7] Musk, Elon. 2016. Full Interview. *Code Conference 2016.* Available online: https://www.youtube.com/watch?v=wsixsRI-Sz4 (accessed August 3, 2022).

[8] Sanford L. Drob (2023) Are you praying to a videogame God? Ibid.

[9] Roger A. MacGowan and Friedrick I. Ordway III, *Intelligent Life in the Universe*. Englewood Cliffs, NJ: Prentice-Hall, 1966, 233-5.

[10] "Noam Chomsky: The False Promise of ChatGPT." *New York Time*s, March 8, 2023.

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.