

Catalogue Description of *Future Minds: Enhancing and Transcending the Brain* (in press)

Susan Schneider

US Library of Congress, Institute for Advanced Study and The University of Connecticut
SchneiderWebsite.com

Homo sapiens may not be the most intelligent species on Earth for that much longer. The world chess, *Go*, and *Jeopardy!* champions are now all AIs. Indeed, AI is projected to outmode many human professions within the next few decades. And given the rapid-fire pace of its development, it may soon advance to artificial general intelligence (AGI)—intelligence that, like human intelligence, can combine insights from different topic areas and display flexibility and common sense. Yet it is naive to think that the future will be like traditional science fiction, with humans being surrounded by ultra-sophisticated AIs, while themselves remaining unenhanced by AI technologies. In reality, AI will not just transform the world, it will transform the human brain. Neural lace, the artificial hippocampus, brain chips to treat mood disorders – these are just some of the mind-altering technologies already under development.

As exciting as mind design may seem, Schneider argues that we are still lacking sufficient grasp of the nature of the mind, consciousness and the self. If we are not careful to understand these issues, we may experience one or more *perverse realizations* of AI technology -- situations in which AI fails to make life easier but instead leads to our own suffering or demise, or to the exploitation of other conscious beings. The perverse realizations generally fall into one of the following types: (i), overlooked situations involving the creation of conscious machines and (ii), scenarios that concern radical brain enhancement. For instance, radical enhancements, like uploading and replacing certain parts of the brain with microchips, will be infeasible if AI components do not support consciousness, and even if they do, the unbridled use of such technologies may not even be compatible with the persistence of the person. Yet these matters involve classic philosophical puzzles that have no easy solutions.

In considering how to avoid these scenarios, Schneider dismantles various misunderstandings about the nature of self and mind. These range from academic claims that the mind is a software program to the sentiment, common among tech gurus, transhumanists and media reports, that humans should merge with AI. In each case, the problem is the same: we haven't thought through the underlying metaphysical claims. Drawing from the literature on the metaphysics of the self, her earlier research in cognitive science on the computational nature of thought, her new tests for machine consciousness and her AI project with NASA, the author offers a path forward.