# Should You Merge with AI?

If you upgrade your brain, you risk losing yourself.

I’m suggesting an alteration to the subtitle you had b/c you may not be able to upgrade consciousness -- you may lose it if you try.

By Susan Schneider

May 10th, 2019

As artificial intelligence creates large-scale unemployment, some professionals are attempting to maintain intellectual parity by adding microchips to their brains. Even aside from career worries, it’s not difficult to understand the appeal of merging with AI. After all, if enhancement leads to superintelligence and extreme longevity, isn’t it better than the alternative — the inevitable degeneration of the brain and body?

At The Center for Mind Design in Manhattan, customers will soon be able to choose from a variety of brain enhancements: “Human Calculator” promises to give you savant-level mathematical abilities; “Zen Garden” can make you calmer and more efficient. It is also rumored that if clinical trials go as planned, customers will soon be able to purchase an enhancement bundle called “Merge” — a series of enhancements allowing customers to gradually augment and transfer all of their mental functions to the cloud over a period of five years.

Unfortunately, these brain chips may fail to do their job for two philosophical reasons.

The first involves the nature of consciousness. Notice that as you read this, it feels like something to be you—you are having conscious experience. You are feeling bodily sensations, hearing background noise, seeing the words on the page, and so on. Without consciousness, experience, positive or negative, simply wouldn’t exist. [removing ‘like this’ because no experiences would exist, and adding a few embellishments]

Many philosophers view the nature of consciousness as a mystery. They believe that we don’t fully understand why all the information processing in the brain feels like something. They also believe that we still don’t understand whether consciousness is unique to our biological substrate, or if other substrates — like silicon or graphene microchips — are also capable of generating conscious experiences.

For the sake of argument, let’s assume microchips are the wrong substrate for consciousness. In this case, if you replaced one or more parts of your brain with microchips, you would diminish or end your life as a conscious being. If this is true, then consciousness, as glorious as it is, may be the very thing that limits our intelligence augmentation. If microchips are the wrong stuff, then AIs themselves wouldn’t have this design ceiling on intelligence augmentation — but they would be incapable of consciousness. [I made microchips plural above, rather than just saying ‘chip’ b/c you’d need more than one if you replaced multiple parts]

You might object, saying that we can still enhance parts of the brain not responsible for consciousness. It is true that much of what the brain does is nonconscious computation, but neuroscientists suspect that our working memory and attentional systems are part of the neural basis of consciousness. These systems are notoriously slow, processing only about four manageable chunks of information at a time. If replacing parts of these systems with AI

components produces a loss of consciousness, we may be stuck with our pre-existing bandwidth limitations. This may amount to a massive bottleneck on the brain’s capacity to attend to and synthesize data piping in through chips used in areas of the brain that are not responsible for consciousness.

But let’s suppose that microchips turn out to be the right stuff. There is still a second problem, one that involves the nature of the self. Imagine that, longing for superintelligence, you consider buying “Merge.” In order to understand whether you should embark upon this journey, you must first understand what and who you are. But what is a self or person? What allows a self to continue existing over time? Like consciousness, the nature of the self is a matter of intense philosophical controversy. And, given your conception of a self or person, would you continue to exist after adding Merge — or would you have ceased to exist, having been replaced by someone else? If the latter, why try Merge in the first place?

Even if your hypothetical merger with AI brings benefits like superhuman intelligence and radical life extension, it must not involve the elimination of any of what philosophers call “essential properties” — the things that make you you. Even if you would like to become superintelligent, knowingly trading away one or more of your essential properties would be tantamount to suicide — that is, to your intentionally causing yourself to cease to exist. So before you attempt to redesign your mind, you’d better know what your essential properties are.

Unfortunately, there’s no clear answer about what your essential properties might be. Many philosophers sympathize with the “psychological continuity view,” which says that our memories and personality dispositions make us who we are. But this means that if we change our memories or personality in radical ways, the continuity could be broken. Another leading view is that your brain is essential to you, even if there are radical breaks in continuity. But on this view, enhancements like Merge are unsafe, as you are replacing parts of your brain with AI components.

Advocates of a mind-machine merger tend to reject the view that the mind is the brain, however. They believe that the mind is like a software program: just as you can upload and download a computer file, your mind can add new lines of code and even be uploaded onto the cloud. According to this view, the underlying substrate that runs your “self program” doesn’t really matter — it could be a biological brain or a silicon computer. [changing to computer here b/c a single microchip wouldn’t be sufficient.]

However, this view doesn’t hold up under scrutiny. A program is a list of instructions in a programming language that tell the computer what tasks to do, and a line of code is like a mathematical equation. It is highly abstract, in contrast with the concrete physical world. Equations and programs are what philosophers call “abstract entities" — things not situated in space or time. But minds and selves are spatial beings and causal agents; our minds have thoughts that cause us to act in the concrete world. And moments pass for us — we are temporal beings.

Perhaps advocates of the software view really mean that the mind or self is just the thing running the program — but this just takes us right back to where we started: what is this thing, this self? Why be confident that it survives enhancements like Merge, or even less radical enhancements, like Zen Garden and Human Calculator? Both of these enhancements still involve major augmentations of certain cognitive capacities. Such changes may, for all we know, alter one’s personality and brain function in significant ways.

This leads me to suspect there may be a second kind of design ceiling on radical brain enhancement. The first design ceiling arises if microchips fail to underlie conscious experience — let’s call this the “consciousness ceiling.” This second ceiling, in contrast, involves the survival of the self. This “self ceiling” is a point beyond which the person who attempts to enhance is no longer the same individual as before, for the procedure causes that individual who sought enhancement to cease to exist. Because the nature of the self is so controversial, we don’t know if there’s a self ceiling. Nor do we know how high, or low, a self ceiling would be situated.

These potential ceilings suggest that we should approach the idea of merging with AI with a good deal of humility. Technological prowess is not enough. To flourish, we must appreciate the philosophical issues lying beneath the algorithms.

The Times is committed to publishing [a diversity of letters](https://www.nytimes.com/2019/01/31/opinion/letters/letters-to-editor-new-york-times-women.html) to the editor. We’d like to hear what you think about this or any of our articles. Here are some [tips](https://help.nytimes.com/hc/en-us/articles/115014925288-How-to-submit-a-letter-to-the-editor). And here’s our email: [letters@nytimes.com](mailto:letters@nytimes.com).

Follow The New York Times Opinion section on [Facebook](https://www.facebook.com/nytopinion), [Twitter (@NYTopinion)](http://twitter.com/NYTOpinion) and [Instagram](https://www.instagram.com/nytopinion/).