Cambridge Quarterly of Healthcare Ethics (2024), 33: 1, 76–88
doi:10.1017/S0963180123000464

CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Artificial Intelligence and Human Enhancement: Can AI Technologies Make Us More (Artificially) Intelligent?

Sven Nyholm

Ludwig-Maximilians Universität München, Munich, Germany
Email: s.nyholm@lmu.de

Abstract

This paper discusses two opposing views about the relation between artificial intelligence (AI) and human intelligence: on the one hand, a worry that heavy reliance on AI technologies might make people less intelligent and, on the other, a hope that AI technologies might serve as a form of cognitive enhancement. The worry relates to the notion that if we hand over too many intelligence-requiring tasks to AI technologies, we might end up with fewer opportunities to train our own intelligence. Concerning AI as a potential form of cognitive enhancement, the paper explores two possibilities: (1) AI as extending—and thereby enhancing—people's minds, and (2) AI as enabling people to behave in artificially intelligent ways. That is, using AI technologies might enable people to behave as if they have been cognitively enhanced. The paper considers such enhancements both on the level of individuals and on the level of groups.

Keywords: artificial intelligence; cognitive enhancement; human downgrading; human enhancement; human intelligence; intelligent behavior; the extended mind

Introduction

When people discuss the impressive progress that has been made with respect to artificial intelligence (AI), they frequently highlight a set of Go games in which the human world champion of Go, Lee Sedol, faced off against DeepMind's computer program AlphaGo. These games took place in March of 2016. Lee Sedol managed to win one out of five games. But he lost the other four. This was a sensation. Many people had thought that it would be extremely hard to create AI technologies that could defeat human Go masters in this sophisticated game.

Much discussion about this case has been about Lee Sedol and his reaction to what happened. He said that he would retire from competing in Go and that playing Go had become meaningless for him.[1] But here I want to consider another human being involved in this set of games: namely, the man who was carrying out the moves that were recommended by AlphaGo, and who was thereby going through the motions, so to speak, of beating Lee Sedol in Go. AlphaGo was a computer program, not a robot equipped with arms. So AlphaGo itself did not move the playing stones around on the board. Before the games, AlphaGo had been trained on data from thousands of human-played Go games, and it had also been trained by playing millions of games against itself. This allowed AlphaGo to develop novel Go-playing strategies, and what AlphaGo did was to recommend moves in the game. A human—the person I will focus on here—took these recommendations and carried out the moves in the game, by placing the stones on the recommended parts of the Go board. This person (someone working for DeepMind) thereby became able to perform in a manner that corresponded to beating the world champion of Go in four out of five games.

Now consider this question: Did this person in effect achieve the ability to play Go at a level at which not even the human world champion (Lee Sedol) was able to play? In other words, was the DeepMind employee exhibiting a form of human enhancement when he was acting on recommendations from the AI-driven AlphaGo computer program?

Of popular focus in discussions about AI progress is the large language model technology ChatGPT, which was released to the public in November of 2022. This is an AI tool that can produce text (including essays and poems) in response to prompts from human users. When I tried using ChatGPT, for example, I entered the prompt, "what would Martin Heidegger think about the ethics of AI?"—and ChatGPT immediately produced a fairly impressive short essay on the topic. In fact, I think ChatGPT did a better job than at least some—or perhaps even many—of the students in my classes would do. In all honesty, I think that the short essay that ChatGPT produced is better than what I would be able to come up with on the spot if I were asked to quickly write an essay on what Heidegger would say about AI ethics. We are faced with the questions: If I—or a student—began using ChatGPT, or some other large language model, to produce academic texts, would we thereby become enabled to generate better texts than if we did not use ChatGPT? Would this AI technology then serve as an enhancement of our writing abilities? Could it be understood as a form of human enhancement?

Whether AI can be seen as a form of human enhancement is explored below—I will offer reasons in favor of this idea, but I will also offer reasons to doubt it. In a recent paper with coauthors,[2] we argued that it is doubtful if human users deserve credit for outputs produced by AI technologies like ChatGPT or other forms of generative AI. It was our position that putting a prompt into such a technology, which then produces some impressive outcome, is not a sign that either we have any particular form of talent, or that we had made any special effort showing ourselves deserving of credit. That line of reasoning could suggest that AI technologies like ChatGPT do not enhance the capacities of the people who use them. (I will return to this issue later.) However, there might also be other ways of conceptualizing this topic that would allow us to say that we can enhance ourselves and our abilities by using AI technologies.

Among other things, I will suggest that if we can extend the so-called extended mind thesis to apply to the use of AI—so that we can be said to extend our minds by using AI technologies—we might say that it is possible to enhance ourselves and our abilities with AI technologies. I will discuss this thesis both on the individual level (Can individuals enhance their abilities by making use of AI technologies?) and on a group level (Can groups enhance themselves by making use of AI technologies?). Of particular interest will be whether using AI technologies can be interpreted as a form of cognitive enhancement. In other words, can we enhance our cognitive abilities by using AI technologies? A related consideration will also be the notion that AI technologies might cognitively enhance human users, not by boosting their natural intelligence, but by giving them a kind of artificial intelligence; that is, by enabling humans to behave as if they have enhanced natural intelligence.

The idea that AI can function as a form of cognitive extension has recently been discussed by some researchers, in particular José Hernández-Orallo and Karina Vold.[3] Moreover, the idea that AI can function as a form of human enhancement has also been discussed, but only to a limited extent. Specifically, the idea that AI-driven recommender systems could function as a form of moral enhancers has been addressed by authors such as Alberto Giubilini, Julian Savulescu, and Michał Klincewicz.[4] Relatedly, Alexandre Erler and Vincent Müller have argued that AI can function as an augmentation of human intelligence.[5] My aim here is to add to these deliberations by relating the perspective on AI as a form of enhancement to the discussion of AI as a form of cognitive extension. Additionally, the existing discussion is expanded by suggesting that AI might under certain circumstances give humans a form of artificial intelligence.

The paper is divided into the following sections. (1) I will first discuss human intelligence, and a worry promulgated by some members of the press that AI and other contemporary technologies might be making people less intelligent. (2) There then follows a consideration of the concept of AI, including why it might initially be thought to not be a form of human enhancement. (3) Next up is the idea of human enhancement—in particular, cognitive enhancement—along with some distinctions from the literature on enhancement and the extended mind thesis, which provide some interesting ideas about how AI might function as a form of cognitive enhancement. (4) I will next discuss possible limitations to the view of AI as a form of human cognitive enhancement, and its relation to John Searle's so-called Chinese room

https://doi.org/10.1017/S0963180123000464 Published online by Cambridge University Press

argument[6] against the idea that AI technologies have any form of understanding. This will help to illustrate the idea that perhaps we should think of AI technologies as enabling a form of group-level cognitive enhancement, rather than—or perhaps in addition to—enhancement of individuals. It will also motivate the idea that perhaps what AI might give us is a form of artificial intelligence, rather than a straightforward form of cognitive enhancement. Four different ways of thinking about whether AI technologies can enhance people and their abilities will be identified. (5) Lastly, the sections conclude by considering whether praiseworthiness (or lack thereof) can be used as a criterion for determining whether AI can serve as a human enhancer.

Human Intelligence and Cognitive Enhancement

In the spring of 2023, I was contacted by a German journalist from the BR (*Bayrischer Rundfunk)* who wanted to interview me and other researchers about the question of whether overreliance on AI might make people "dumber."[7] This seemed to be an idea that was making the rounds in various media circles in Europe around that time. There was also a show on Swedish television, "Idébyrån," in which a physics professor, a philosopher, and a psychologist had been invited to discuss the thesis that "You are getting dumber at the same time as machines are getting smarter."[8] Journalists were responding partly to general worries in society, but also to research suggesting that a steady increase in intelligence quotient (IQ) observed throughout the 20th century might recently have become reversed. Norwegian researchers had found that people had recently started scoring less well on IQ test.[9] One of the possible explanations offered was overreliance on modern technology.

Such discussions raise the question of what we understand by human intelligence and what is the relation between our intelligence and our use of modern technology. Traditionally, human beings' ability to create and use technologies has been seen as one of the things that show we are more intelligent than other animals. But now worry is mounting among some that using certain technologies, in particular AI, will make us less intelligent. Reflection is needed on whether we are enhancing ourselves by creating AI technologies or whether we are risking a form of "human downgrading," to use Tristan Harris and Aza Raskin's expression.[10] Relatedly, the philosopher Shannon Vallor has raised worries that many modern technologies might lead to moral de-skilling, reducing our skill at making moral decisions due to lack of practice.[11] The philosopher and legal academic John Danaher goes so far as to worry that transferring more agency into technologies may lead to "a crisis of moral patiency," whereby human beings are reduced from active agents to passive patients, including with respect to moral decision-making.[12]

Now, there are many different attempts to define natural intelligence. These range from what might be called behavioristic definitions to definitions referring specifically to psychological capacities. Notably, if intelligence is understood in terms of certain forms of behavior or capacities for complex behavior, then technologies—such as AI technologies—can potentially become intelligent. But if intelligence is defined in a way that refers to subjective or conscious states, then it seems much less plausible that it would be possible to create technologies with something that corresponds to natural intelligence of the sort found in humans.

Two examples of definitions of intelligence used by prominent contributors to discussions of AI are that intelligence amounts to the ability to take the right actions to promote one's goals at the right time (a definition used by Joanna Bryson[13]) and that intelligence roughly amounts to sophisticated behavior (a definition recently used by David Chalmers[14]). Another example is Peter Railton's view that intelligence should be understood as a capacity for problem solving.[15] Those three definitions do not essentially refer to conscious experience and thereby seem to allow technologies to potentially be intelligent. In contrast, if intelligence is understood in terms of the ability to understand semantic meaning, and one thinks that this requires subjective consciousness, one might think that creating intelligent machines would be very difficult, since this would require the ability to create machines with conscious states.[16]

In what follows, I will use a broad definition of human intelligence and assume that, on a general level, human intelligence is a set of cognitive capacities that helps us to expand our knowledge and

https://doi.org/10.1017/S0963180123000464 Published online by Cambridge University Press

understanding of the world, and that it also helps us to better achieve our goals and live in accordance with our values. Moreover, I will assume that intelligence is (i) a basic potential that human beings have —and that members of other species might also have, though of different kinds—and/or (ii) a more or less fully realized potential, whereby different people might be more or less intelligent in their thinking or behavior.

This way of thinking implies at least two things about the idea of cognitive enhancement. First, anything that helps to boost our capacities to acquire knowledge and understand the world, and that might enable us to act in ways that better achieve our goals in accordance with our values, could potentially be seen as a form of cognitive enhancement. Second, we can distinguish between enhancements of our basic potential for intelligence, on the one hand, and enhancements of the development of our basic potential for intelligence, on the other.

Normally, different tools and new inventions are seen as potential ways of enhancing our human abilities, so that most new technologies would potentially be forms of human enhancement, at least in a general sense. This would suggest that like any other tool or invention, AI should also potentially be a form of human enhancement of our abilities, including our abilities to use and develop our intelligence. Yet as we saw at the beginning of this section, there are those who worry that if we can create AI, then this might make us less intelligent—or perhaps less prone to use or develop our cognitive abilities. Let us therefore now consider what we should understand by the idea of AI and some different perspectives on how it relates to human intelligence.

Artificial Intelligence and Its Relation to Human Intelligence

Currently, the label “artificial intelligence” has many different applications. One reaction is to take an inclusive approach and say “Yes, let us use the label broadly.” Another response would be to say that “‘Artificial intelligence’ should be reserved for the use of machine learning techniques.” This second approach would exclude what is sometimes called “symbolic” or “rule-based” AI, also known as “good old-fashioned AI,” from qualifying as forms of AI. A way of sidestepping such discussions about what techniques should, or should not, be used for a technology to count as AI is to employ a Marvin Minsky[17]-inspired functional way of defining AI. Because the discussion here is whether AI could ever serve as a form of cognitive enhancement, I will be using that wider, functional type of approach.

According to this common way of explaining what AI is, technologies with artificial intelligence are those technologies that can perform, or take over, tasks that we humans need our natural intelligence to perform.[18] For example, since human intelligence is needed to write texts, drive cars, or make medical diagnoses, any technologies that can generate texts in response to prompts (such as large language model technologies), self-driving cars, or medical diagnosing systems possess artificial intelligence. The reason being that such technologies can take over tasks that we use our natural intelligence to perform.

A question is whether such technologies thereby have something that can be compared with human intelligence. The history of thinking about AI over approximately the last 70 years contains an array of different answers to this question. Alan Turing, writing as early as 1950, before the term “artificial intelligence” had been coined, discussed the question of whether machines can think.[19] He was responding, in part, to a paper by the neurologist Geoffrey Jefferson that had been published a year earlier in 1949.[20] Jefferson argued that machines cannot think. Turing famously suggested that a better question to ask is whether it is possible to build machines that behave as if they can think or that can convincingly imitate human thinking. According to this perspective, AI is an imitation of human intelligence. All that matters is whether AI can behave as if it has some form of intelligence.

The researchers who coined the term “artificial intelligence” in 1955 spoke, not about imitating human intelligence, but instead about simulating it. In an influential research proposal, John McCarthy and a group of colleagues wrote that many, or all, aspects of human intelligence can be described in such a precise way that it should be possible to create technologies that can simulate learning and other key aspects of human intelligence.[21]

https://doi.org/10.1017/S0963180123000464 Published online by Cambridge University Press

If we fast-forward to the 1990s, in the most widely used textbook on AI, the computer scientists Stuart Russell and Peter Norvig defined AI as the creation of artificial intelligent agents.[22] This means that AI is the creation of artificial agents that can "perceive" their environment and "act" so as to effectively achieve their goals in that environment. Given that Russell and Norvig understand human intelligence in terms of the capacity to effectively achieve goals, they appear to hold that it is unproblematic to say that AI agents can have intelligence like that of human beings. In contrast, the influential philosopher of information Luciano Floridi has suggested that what is distinctive about AI is that it is the creation of agents that are able to achieve goals without the need for any intelligence.[23] We human beings achieve our goals by using our human intelligence. In contrast, Floridi believes that AI agents act in the world and effectively achieve goals without possessing anything we should think of as being similar to human-like intelligence.

These ideas reviewed above, about how AI relates to human intelligence, are compatible with the general view that AI technologies are technologies created for the purpose of taking over tasks that we use our human intelligence to perform. In what follows, I will remain, for the most part, neutral as to whether AI technologies imitate, simulate, have, or are able to act without the need for having anything like human intelligence.

A key question here, however, is whether handing over to AI technologies tasks that we use our intelligence to perform might potentially be a way for us to make ourselves less intelligent. One worry could be that we would primarily be leaving to ourselves tasks that do not require us to be very intelligent. Or could we perhaps be seen as extending our own intelligence out into these technologies as we start to increase our use of AI technologies?

It should be noted that at first it seems unlikely, if not impossible—at least in the short run—that handing over intelligence-requiring tasks to AI technologies would diminish our basic potential for intelligence. As noted above, we can think of intelligence as a basic capacity. And there would presumably need to be evolutionary changes to human brains in order for that basic potential to be eliminated.

However, we can also think of intelligence and its levels in terms of varying degrees of development of our basic capacity for intelligence. From that perspective, it is not altogether implausible to think that if we hand over too many tasks to AI systems, and we therefore have fewer occasions or incentives to develop our capacity for intelligence, then there is a risk that we might end up being less intelligent than we could otherwise be. If we assign to AI technologies many of our intelligence-requiring tasks, rather than engaging in these tasks ourselves, and instead rely heavily on AI technologies, the increased dependence could prevent some people from fully realizing their intellectual potential.

The question arises whether this way of thinking about the relation between our human intelligence and AI technologies is the only way to conceptualize the connection. Let us consider whether there are alternative ways of reasoning that would allow us to view using AI technologies as a form of enhancement of our own capabilities. What is needed in order for AI technologies to potentially be seen as a form of human enhancement of our capabilities might be a Gestalt shift in how we conceive of the relation between AI technologies and ourselves.

Might AI Technologies Enhance Human Intelligence by Extending Our Minds?

If we consider our own minds and our capacities as being wholly separate from AI technologies, and we hand over our intellectual tasks to them, there is concern, as discussed above, that our own capabilities may atrophy and weaken. However, viewing ourselves as separate from the technologies that we create and use is not the only way of conceiving of our relationship to our technologies.[24] Another way of viewing our relation to technologies is not one of separation, but rather that of merging, or forming units.[25]

One version of this is referred to as the "extended mind" thesis, which could be generalized to what might be called the "extended capacities" thesis. This idea—which stems from the work of Andy Clark and David Chalmers[26]—understands humans and some technologies as forming "coupled systems," which may function in analogous ways to how our internal capacities function.

Simply put, Clark and Chalmers suggest that when couplings of human agents and entities, such as computers or even low-tech artifacts like pen and paper, form systems that perform functions associated with human minds, extended minds are created. The mind of the human agent is then extended out into the technology (or broader environment) that enables the human–technology composite to perform the relevant functions. Clark and Chalmers formulate various criteria for when this can be said to happen.

On the one hand, Clark and Chalmers suggest that the following three conditions should hold in order to make sense of conceiving a human and a piece of technology as a "coupled cognitive system": (1) All system components should play active causal roles, whereby elements inside and outside of the person's body affect each other; (2) all system components help regulate the behavior of the agent in the way a mind is usually thought to do; and (3) if the external parts (e.g., a computer or the pen and paper) are removed, the overall system's behavioral capacities are diminished.

On the other hand, not all external parts of the environment or technologies we use are part of extended minds, in the view of Clark and Chalmers. The following three conditions should hold: (1) The external component is a constant, reliable part of the person's life; (2) the information or other inputs from the external part(s) should be easily available; and (3) the person needs to be disposed to automatically endorse the inputs from the external parts. (Although not addressed here, another test Clark and Chalmers suggest for whether something is part of a person's "extended mind" is moral in nature: Would it be an assault on a person or their mind if we remove, destroy, or otherwise interfere with the external part?)

Clark and Chalmers are not recognized for relating the AI–human relationship discussion to the literature on human enhancement; but others have done so, including Neil Levy[27] and John Danaher.[28] They both discuss a distinction between what they call "external" and "internal" forms of human enhancement. Traditionally, the ethics of human enhancement has focused on changes within the body or brain that are thought to function as enhancements. For example, people may take different sorts of drugs or place various forms of technology into their bodies or brains in order to enhance their capacities. However, according to Levy and Danaher, if we change people's environments or increase access to apps or other systems that offer recommendations that can improve their mental health or behavior, this can also improve people's thinking or behavior.[29] These changes would be examples of an enhancement.

The striking move that Levy makes—and that Danaher also discusses in response to Levy—is to relate the idea of external enhancements to the extended mind thesis. By doing so, the distinction between external and internal enhancements is weakened or perhaps even dissolved. In this way, some external enhancements could be seen as becoming parts of our extended minds. Levy writes:

> if the mind is not confined within the skull…[then] intervening in the mind is ubiquitous. It becomes difficult to defend the idea that there is a difference in principle between interventions which work by altering a person's environment and that work directly on her brain, insofar as the effect on cognition is the same; the mere fact that an intervention targets the brain directly no longer seems relevant.[30]

Danaher interpretates this view in his summation:

> if the [extended mind hypothesis] is true, then we are always enhancing the human mind through the use of technology.[31]

If the Levy and Danaher proposition is sound and worthy of serious consideration, then we could potentially apply this to AI technologies as well, with the resulting position that some AI technologies can work as a form of human enhancements. This means that when we use AI technologies (e.g., to write texts, to play Go games, or in other intellectual endeavors), we are extending our minds and giving ourselves new capabilities and thereby enhancing ourselves.

At this point, a host of new questions needs to be addressed: Are there researchers who already discuss the idea that AI technologies could potentially be seen as extensions of our minds or cognitive abilities? If so, do they argue that this is a form of human enhancement? Do AI technologies fulfill the above-listed

https://doi.org/10.1017/S0963180123000464 Published online by Cambridge University Press

criteria for technologies to qualify as extensions of our minds or capacities from Clark and Chalmers's earlier work? Alternatively, if there is only a partial match between how AI technologies relate to our minds and our capacities and the criteria Clark and Chalmers suggest, might AI technologies nevertheless be seen as some form of human enhancement technologies—for example, as a form of external enhancement? The last question will be treated in the next section. But we begin with the questions just formulated.

Karina Vold and her collaborators have suggested—though without discussing human enhancement explicitly—that a subset of AI technologies can indeed be seen as extensions of our minds and cognitive capacities. What is noteworthy about Vold's discussion, however, is that she makes a distinction between AI technologies that operate in a highly autonomous way (e.g., a fully automated self-driving car) and AI technologies that function more as support systems requiring extensive human input and engagement.[32] Importantly, Vold argues that the more independent some AI technology is in the way it operates, the less it makes sense to see it as an extension of a human mind. However, Vold argues, if an AI technology is operating in a less autonomous way—and human input and engagement are needed—then it does make sense to see it is as an extension of our minds and capacities. In other words, the kinds of criteria that Clark and Chalmers list relate less to fully autonomously operating AI technologies (e.g., fully automated self-driving cars) than to human engagement-requiring AI technologies (e.g., apps on phones that give recommendations based on continuous inputs from, or engagement by, the user).

Along similar lines, Chalmers in a talk given at a workshop about the philosophy of large language models discussed the question of whether large language models such as ChatGPT could be said to extend users' minds.[33] Chalmers asked whether letting such technologies write texts for us in response to prompts provided by us could be seen as a way of extending our minds into these AI technologies. Interestingly, Chalmers argued that a technology like ChatGPT might be operating too independently for it to qualify as an extension of the user's mind.

More specifically, the concern expressed by Vold and Chalmers is that for some forms of AI technologies, it is *not* the case that when humans use these AI technologies, (1) all system components play active causal roles, whereby elements inside and outside of the person's body affect each other; (2) all system components help regulate the behavior of the agent in the way a mind is usually thought to do; and (3) if the external parts are removed, the overall system's behavioral capacities are diminished. Nor may it be the case that (1) the external component is a constant, reliable part of the person's life; (2) the information or other inputs from the external part(s) are easily available; and (3) the person is disposed to automatically endorse the inputs from the external parts.

Thus, a consideration that must be taken into account in our discussion as to whether AI technologies (a) can be seen as a form of extenders of human minds and (b) therefore potentially be seen as forms of human enhancement, can be stated as follows: Some of the leading researchers who are most open to the idea of the possibility of extending human minds with the help of technologies—such as Vold and colleagues, or indeed Chalmers—think that the plausibility of the extended mind thesis as it relates to AI technologies will be limited to a subset of all AI technologies. It will not apply seamlessly to the whole set of AI technologies. Briefly put, the more autonomously the AI technologies are operating, the less it makes sense to see them as extensions of our minds.

In terms of the hypothesis we are exploring as to whether AI could serve as a form of human enhancement in general and a form of cognitive enhancement in particular, one partial conclusion might be that AI technologies are a form of "internal" enhancement (i.e., internal to our minds) when, and only when, they function as mind extenders, requiring extensive human engagement, and are not operating excessively independently. Another possibility is that more independently operating AI technologies may serve as a form of "external" enhancements (i.e., they might enhance our abilities without becoming internal to us or our minds). These options can be explored further by returning to the opening examples.

Artificial Human Intelligence

Let us return to the man who was moving the stones around on the board when AlphaGo faced off against Lee Sedol in the famous set of Go games. Let us suppose that this man did not understand the

https://doi.org/10.1017/S0963180123000464 Published online by Cambridge University Press

strategies behind the suggestions that AlphaGo made as to what moves he should make. Let us also suppose that this man had only had a vague idea, at best, of how Go is played. Yet, with the recommendations from AlphaGo, he was able to exhibit motions of behaving in a way that from the outside could appear as if he were a cognitively enhanced person.

Suppose next that, instead of it being obvious and known to everyone that he was moving the stones around based on AlphaGo's recommendations, onlookers were unaware that all his moves were actually generated by an AI technology. Imagine in this version of the example that AlphaGo was making recommendations to this man via voice commands, which were played to him in a small earbud that was hard to spot from the outside. In this fictional account, it could look even more as if this man had been cognitively enhanced in some way. In such a scenario, for example, Lee Sedol (supposing that he also did not know what was going on) might think that he, the world champion, was facing another player with greater Go-playing abilities than he had.

In my fictional version of the case, and indeed even in the real-world account, there are some parallels between the man moving the stones around on the Go board based on the directions from AlphaGo and the man in John Searle's well-known "Chinese room" argument.[34] As many readers may know, Searle imagined a man in a room with an instruction book for how to combine different Chinese characters. In this story, messages written in Chinese are given to the man via a mailbox; the man looks at the instruction book to decide what Chinese characters to select; he puts his responses on paper and then outputs them via the mailbox.

To someone outside of the room, it might appear as if whoever is in the room is a native Chinese speaker. But this man, Searle affirms, clearly does not understand Chinese. And the parallel to AI is supposed to be that even if a computer passes the Turing test—that is, it can produce outputs that can be mistaken for human outputs—this does not show that the computer understands its outputs. Similarly, the Go-playing man in my fictional version of the case might look as if he knows how to play Go at a higher level than the world champion Lee Sedol. Hence, he might look as if he is cognitively enhanced; but this man is only able to play Go at this exalted level by matching his behavior with instructions from AI.

Bringing large language models like ChatGPT into this comparison of examples—suppose that rather than having an instruction book, the man in the Chinese room example has something similar to ChatGPT that he uses to produce the messages in response to the inputs that come via the mailbox. But now instead of looking up what to do in the instruction book, the man inputs these Chinese messages into an app or computer program that works similarly to ChatGPT—he enters "prompts"—and then the ChatGPT-like program outputs messages. The man prints these messages using a printer in the room and then delivers them to the outside via the mailbox. Again, it might look like there is a Chinese speaker in the room to people on the outside. But again there is somebody in the room who does not understand Chinese, but who may appear to have the ability to communicate intelligently in Chinese. Is this man in the ChatGPT version of the Chinese room scenario cognitively enhanced? Relatedly, is the man in my fictional version of the Go game cognitively enhanced?

If, in comparison, the man moving Go stones and the man in the Chinese room qualify as having extended their minds with the help of the AI technologies instructing them what to do, then by the logic in the reasoning from Levy and Danaher, these two people would seem to have been internally cognitively enhanced. That is, their minds have been extended out into these AI technologies, and their minds now have abilities they previously did not possess.

However, if we follow Vold or Chalmers, and we think that these AI technologies are too independent from the people performing actions in the examples with respect to how autonomously they operate, then the man moving Go stones and the man writing Chinese symbols should not be thought to have extended minds. Accordingly, there would be no internal enhancement of their minds, from the Levy and Danaher perspective.

Might there be another way of thinking about the man moving Go stones and the man in the Chinese room so that they could be said to be cognitively enhanced? We could say that these people have been made, in certain respects, artificially intelligent. That is, the Go-playing man who makes brilliant moves based on recommendations from AlphaGo and the man in the Chinese room who is using a large

https://doi.org/10.1017/S0963180123000464 Published online by Cambridge University Press

language model to communicate in Chinese have in certain respects become artificially intelligent in their behavior. At least, this appears to be a way of interpreting what is happening if we follow either Turing's, or indeed, Floridi's way of thinking about what AI is.

Recall that Turing's approach is to say that machines can think if they are able to imitate human thinking.[35] Remember too, that Floridi says that AI technologies are agents that are able to act in efficient ways without the need for any natural intelligence to guide their actions.[36] We might think that the people in the cited examples pass these tests. The man who is playing Go based on secret recommendations from AlphaGo behaves as if he is an extremely intelligent human being. In fact, intelligent enough to be able to beat Lee Sedol at Go. The man is also acting as a very effective agent, and he is doing so without the need for the impressive natural human intelligence that Lee Sedol is displaying when he is making his moves in the game. Similarly, the person in the Chinese room behaves as though he is able to communicate with people on the outside without possessing any actual intelligence with respect to the Chinese language.

Thus, when it comes to individual human beings, we have two alternative ways in which AI technologies might be thought to cognitively enhance us or, in other words, improve our intelligence. One way in which this can happen is if the AI technologies work as cognitive extenders, which extend our minds and give our minds new or improved abilities. The alternative is if the AI technologies give us a form of AI of the Turing or Floridi kinds. That is, the AI technologies might enable us to act as if we had an impressive level of natural human intelligence and/or they might enable us to become highly efficient agents without any need of improving our natural human intelligence.

In these instances, we are speaking about possibilities that relate to individuals. What can be said about the relationship to groups? Could groups of humans be (cognitively) mutually enhanced by making use of AI technologies? Here, too, we could imagine two possible ways of thinking that relate roughly to the two possibilities described above. The first might be to think that human groups sometimes interact in a sufficiently organized way that they could be said to have a corporate mind—an idea explored, for example, in the work of Christian List and Philip Pettit.[37] On such a view, perhaps AI technologies could be used to extend and, in effect, enhance these group-level minds. Another approach would be to think that the group might be able to use AI technology so that it behaves as if it has an enhanced level of intelligence on a group level. The group, in other words, might be given a form of artificial intelligence in the Turing sense and/or the Floridi sense.

The second possibility fits, to some extent, with a response that has sometimes been given to Searle's Chinese room thought experiment and his claim that the person in the room does not understand Chinese. Specifically, that the system that is formed in the room by the person and the instructions understands Chinese on a system level. In the same way, in my version of the example where the instruction book is substituted for a ChatGPT-like large language model technology, a system that understands Chinese is made up of the AI technology and the person in the room. Moreover, a larger system could be envisioned that includes the person in the room with the large language model, along with those outside of the room. In this way, a group-level community could have been formed whose members can communicate with each other in a language not possible before. The overall system has thereby, it might be argued, in a certain sense, become artificially intelligent in its behavior—at least in the Turing or Floridi senses.[38]

Accordingly, when it comes to whether AI technologies can be thought of as human enhancers, there are at least four possibilities to consider:

1: AI technologies (at least if they are not operating too autonomously from human beings) may extend individual people's minds and thereby enhance the cognitive capacities of these individuals.
2: AI technologies (including ones that operate highly autonomously) may enable people to behave as if they have enhanced cognitive abilities and thereby make these individuals artificially intelligent in certain respects.
3: AI technologies may extend the corporate minds of organized groups of people and thereby cognitively enhance these group agents.

https://doi.org/10.1017/S0963180123000464 Published online by Cambridge University Press

4: AI technologies may enable groups to behave (including allowing its members to interact) as if the groups or their members have enhanced cognitive abilities and thereby make these groups artificially intelligent in certain respects.

Concluding Reflections

The previous section articulated four possible ways in which human beings might be considered as being cognitively enhanced by AI technologies, including autonomously operating AI technologies. How plausible we think that those four different views are may depend on whether we apply criteria that simply consider whether the AI technologies allow us to act in good ways, that is to say, in ways that produce good effects or that are in themselves good—or criteria that consider whether we would be praiseworthy or could justifiably take credit for how we behave as a result of using these putative cognitive enhancement technologies. That is, the second criterion could be that if we can justifiably or plausibly be said to be praiseworthy for some behavior, this would be a clear indication of some sort of improvement of our capacities—given that praiseworthiness is often thought to track, at least in part, whether one is performing in a way that involves good uses of one's capacities.

In this way, a first simple test for whether something would qualify as an internal or external form of cognitive enhancement might just be whether it enables us to act in ways that are good or have good effects. On such a view, our capacities have been enhanced as long as there is anything that could be seen as a cognitive process—either inside or outside of our minds—that creates improvements in our behavior or with respect to what we can achieve. This would allow us, for example, to say that we have been cognitively enhanced if AI technologies give us a certain degree of AI in the Turing or Floridi senses, that is, if these technologies enable us to act as if we have improved cognitive capacities. However, this might seem like too behavioristic a test for whether we have been cognitively enhanced as a result of using certain AI technologies.

The other test—namely, the test that considers whether our resulting behavior is praiseworthy or does us credit—might be a better test. Here, it might be less plausible to think that we are truly cognitively enhanced if AI technologies give us a certain amount of artificial intelligence in the Turing or Floridi senses. That is, if technologies give us recommendations, perform cognitive tasks for us, or in other ways lead us to behave in a way that could give the impression that we have become more intelligent, this might be seen as not making us praiseworthy or not justifying our taking credit for whatever we achieve. According to this way of thinking, it is only if (1) AI technologies could be seen as being parts of our extended minds—and if (2) our level of praiseworthiness can depend on what goes on in our extended minds—that we can be seen as truly being cognitively enhanced by certain AI technologies.

I take it that there is a reason that a person who gains a measure of artificial intelligence by their use of AI technology is not praiseworthy in the way that another person would be if they had had a boost of their natural intelligence. The difference lies in the perception that, as Hannah Maslen and colleagues argue, being praiseworthy for some behavior or outcome typically requires that we have put in some special effort, that we have shown ourselves to have particular talents, that we have made some significant sacrifice, or that we have otherwise shown ourselves as having some level of excellence that sets us apart.[39] Relying on AI technologies that tell us what to do—or that help us produce impressive outputs—seems insufficient to qualify as ways of having made any extra effort, shown any particular talent, made any significant sacrifice, or otherwise displayed some special form of excellence.[40]

Returning to the man who was able to go through the motions of behaving as if he had a superior level of Go skill that enabled him to beat the world champion, his moves depended on recommendations made to him by the AlphaGo computer program. This being the case, the AI technology could be viewed as giving this man a form of artificial intelligence that allowed him to win against Lee Sedol. However talented the man may have been on his own, his victory depended on the strategies given to him by AlphaGo and is, therefore, not sufficient justification for viewing his triumph as particularly praiseworthy or as deserving credit.

Similarly, if a student—or a researcher—presents work as that individual's own product when it was really, mainly or perhaps wholly, generated by a large language model such as ChatGPT, they would clearly not deserve praise in the same way as if they had produced work themselves of equivalent quality. If it is not known that the work was produced by putting a prompt into a large language model, one might get the impression that, to their credit, the person has an impressive capacity deserving of praise. But when the origin of how the work was produced was learned, that laudatory impression would disappear. Therefore, it could be argued, cognitive enhancement involves more than a person's access to a technology such as ChatGPT—even though at first appearance, it may appear to be so when they present a text of apparent quality and claim authorship.

On a group level, it can perhaps seem much more plausible that the group cognitive abilities might be boosted if AI technologies enable the collective members to behave as if the group has become more cognitively capable. This is because any group-level agency—as Christian List notes[41]—can be seen as something akin to a form of artificial intelligence. That is so, List argues, independently of whether or not a well-organized group uses any modern forms of AI within their organization. This is because any group-level agency is a form of apparently intelligent behavior that is the result of something other than the natural intelligence of any particular individual in the group. Hence, on a group level, according to this way of thinking, there is no clear distinction between natural intelligence and artificial intelligence. This is a fascinating idea. However, what exact conclusions this idea from List could enable us to draw concerning AI and the possibility of cognitive enhancement on the level of groups is a larger topic than I have space to discuss in any detail here.

I conclude by returning to the level of the individual. Regarding individuals and cognitive enhancement via AI, I surmise that it is easier to become, in a certain sense, artificially intelligent by using AI technologies (i.e., become able to act as if one has some special form of intelligence) than it is to truly be cognitively enhanced by using AI technologies. For it to be truly convincing that AI technologies function as forms of cognitive enhancements for human users, it needs to be the case that the humans in question can be seen as extending their minds out into these AI technologies. Moreover, I think it needs to be the case that we can justifiably take credit for what we can do with our extended minds.

Technologies that fail to extend our minds but that enable us to act as if we have improved cognitive abilities, and thereby make us, in a certain respect, artificially intelligent, can be seen as a form of quasi-enhancements. We might call them "artificial cognitive enhancements." But it is less plausible to think of them as truly constituting a form of human enhancement.[42]

Competing interest. The author declares no competing interests.

https://doi.org/10.1017/S0963180123000464 Published online by Cambridge University Press

Notes

1. "Go master quits because 'AI cannot be defeated'", *BBC News* 2019 Nov 27; available at https://www.bbc.com/news/technology-50573071 (last accessed 28 June 2023).
2. Porsdam Mann S, Earp BD, Nyholm S, Danaher J, Møller N, Bowman-Smart H, et al. Generative AI entails a credit-blame asymmetry. *Nature Machine Learning* 2023;**5**:472–75.
3. Hernandez-Orallo J, Vold, K. AI extenders. In: *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*; 2019: 507–13. The idea that technologies can be extensions of human beings is not altogether new. An older example of this idea can, for instance, be found in Marshall McLuhan's work from the sixties. See McLuhan M: Understanding media: The extension of man. New York: McGraw Hill, 1964.
4. Giubilini A, Savulescu J. The artificial moral advisor: The "Ideal Observer" meets artificial intelligence. *Philosophy & Technology* 2018;**31**:169–88; Klincewicz M. Artificial intelligence as a means to moral enhancement. *Studies in Logic, Grammar and Rhetoric* 2016;**48**(1):171–87. For related discussion, see also Lara FD, Deckers J. Artificial intelligence as a Socratic assistant for moral enhancement. *Neuroethics* 2020;**13**(3):275–87.
5. Erler A, Müller VC. AI as IA: The use and abuse of artificial intelligence (AI) as human enhancement through intellectual augmentation (IA). In: Jotterand F, Ienca M, eds. *The Routledge Handbook of the Ethics of Human Enhancement*. London: Routledge, 2023: 189–201.

[6] Searle J. Is the brain's mind a computer program? *Scientific American* 1990;**262**(1):26–31.

[7] "Künstliche Intelligenz", *Theo.Logik – Religion Inside*, BR2 2023; available at https://www.br.de/mediathek/podcast/theo-logik/kuenstliche-intelligenz-4/1985661 (last accessed 29 June 2023).

[8] "Du blir dummare", *Idébyrån*, SVT 2023; available at https://www.svtplay.se/video/Kv1YzLZ/idebyran/du-blir-dummare (last accessed 29 June 2023). These kinds of worries about how AI might make us less inteligent can be compared with previous worries about whether search engines and other technologies might make people less intelligent. See, for instance, Carr N. Is Google making us stupid? What the internet is doing to our brains. The Atlantic 2008: https://www.theatlantic.com/magazine/archive/2008/07/is-google-making-us-stupid/306868/ (accessed on 5 aug. 2023)

[9] Bratsberg B, Rogeberg O. Flynn effect and its reversal are both environmentally caused. *PNAS* 2018;**115**(26):6674–8.

[10] Menn J. Technology ethics campaigners offer plan to fight 'human downgrading'. *Reuters* 2019 Apr 24; available at https://www.reuters.com/article/us-tech-ethics-idUSKCN1S002A (last accessed 29 June 2023).

[11] Vallor S. Moral deskilling and upskilling in a new machine age: Reflections on the ambiguous future of character. *Philosophy & Technology* 2015;**28**:107–24.

[12] Danaher J. The rise of the robots and the crisis of moral patiency. *AI & Society* 2019;**34**:129–36.

[13] Bryson J. The artificial intelligence of the ethics of artificial intelligence. In: Dubber M, Pasquale F, Das, S. eds. *The Oxford Handbook of Ethics of AI*. Oxford: Oxford University Press, 2021: 2–25.

[14] Chalmers D. Could a large language model be conscious? *PhilArchive*; available at https://philpapers.org/rec/CHACAL-3 (last accessed 29 June 2023).

[15] Railton P. Ethics and artificial intelligence. *Uehiro Lectures 2022*, University of Oxford. Video recordings; available at https://www.practicalethics.ox.ac.uk/uehiro-lectures-2022 (last accessed 29 June 2023).

[16] See, for example, Floridi L. AI as agency without intelligence: On ChatGPT, large language models, and other generative models. *Philosophy & Technology* 2023;**36**:15; available at https://link.springer.com/article/10.1007/s13347-023-00621-y. (last accessed 8 August 2023).

[17] Minsky M. *Semantic Information Processing*. Cambridge MA: MIT Press; 1968.

[18] Nyholm S. *This is Technology Ethics: An Introduction*. Hoboken: Wiley-Blackwell; 2023.

[19] Turing A. Computing machinery and intelligence. *Mind* 1950;**59**(236):433–60.

[20] Jefferson G. The mind of mechanical man. *British Medical Journal* 1949;**1**(4616):1105–10.

[21] McCarthy J, Minsky M, Rochester N, Shannon CE. A proposal for the Dartmouth summer research Project on artificial intelligence; available at http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf. (last accessed 29 June 2023).

[22] Russell S, Norvig P. *Artificial Intelligence: A Modern Approach*. Hoboken: Prentice Hall; 1995.

[23] See note 16, Floridi.

[24] See note 18, Nyholm, at chap. 1.

[25] Verbeek PP, *Materializing Morality: Understanding and Designing the Morality of Things*. Chicago: Chicago University Press; 2011.

[26] Clark A, Chalmers D. The extended mind. *Analysis* 1998;**58**(1):7–19.

[27] Levy N. Neuroethics and the extended mind. In: Sahakian B, Illes J, eds. *Oxford Handbook of Neuroethics*. Oxford: Oxford University Press; 2011:285–94.

[28] Danaher J. Why internal moral enhancement might be politically better than external moral enhancement. *Neuroethics* 2019;**12**(1):39–54.

[29] This, for example, is what some of those who discuss AI systems as a form of moral enhancement have mind in mind: they are interested in using recommender systems that enable people to act in more ethical ways, by their own lights or by external criteria.

[30] See note 27, Levy 2011, at 291.

[31] See note 28, Danaher 2019, at 43.

[32] Vold K, Hernandez-Orallo J. AI extenders and the ethics of mental health. In: Jotterand F, Ienca M, eds. *Artificial Intelligence in Brain and Mental Health*. Berlin: Springer; 2021:177–202, at sect. 3; See also note 3, Hernandez-Orallo, Vold 2019.

https://doi.org/10.1017/S0963180123000464 Published online by Cambridge University Press

[33] Chalmers D. Do large language models extend the mind? *Human and smart machines as partners in thought? A hybrid workshop on large language models*, UC Riverside; 2023 May 10.

[34] See note 6, Searle 1990.

[35] See note 19, Turing 1950.

[36] See note 16, Floridi 2023.

[37] List C, Pettit P. *Group Agency*. Oxford: Oxford University Press; 2007.

[38] As it happens, Christian List argues in a recent article that when groups of people become organized and form "group agents," then this is also a form of, or similar to, artificial intelligence. List C. Group agency and artificial intelligence. *Philosophy & Technology* 2021;**34**:1213–42.

[39] Maslen H, Savulescu J, Hunt C. Praiseworthiness and motivational enhancement: 'No pain, no praise'? *Australasian Journal of Philosophy* 2019;**98**(2):304–18.

[40] See note 2, Porsdam Mann et al. 2023.

[41] See note 38, List 2021.

[42] I have benefited from input from Tomi Kushner, a very helpful peer reviewer, and the feedback from an audience at the University of Glasgow where I presented this material at the "Human Enhancement and Well-Being" conference organized by Emma Gordon.

Cite this article: Nyholm S (2024). Artificial Intelligence and Human Enhancement: Can AI Technologies Make Us More (Artificially) Intelligent? *Cambridge Quarterly of Healthcare Ethics* 33: 76–88, doi:10.1017/S0963180123000464