Why can't AI Cognize the External World? Criticism Based on Extended Cognition

Zhi OuYang [1], QiXiang Gao [1*]

[1] School of Philosophy, Inner Mongolia University, 010070, Hohhot, China
[1*] Transportation Institute, Inner Mongolia University, 010070, Hohhot, China

* These authors contributed equally.

ouyangzhi@mail.imu.edu.cn, *gaoqixiang@mail.imu.edu.cn

Abstract. Under the concept of the metaverse, people rethink the problem of artificial intelligence. If AI technology is greatly advanced, can AI in the metaverse constitute a cognitive system with the world we humans live in? This paper argues that AI cannot extend their cognition to the external world, because the "life-world gap" prevents them from communicating with the external world. And the "life-experience gap" in different life worlds makes AI actually cavemen in the metaverse era, they can only cognize shadows that are incredibly close to the external world, and for the external world, they may only understand in abstract logic, but in reality, they do not understand everything about the world we live in.

Keywords: AI, Metaverse, Extended cognition

1 Introduction

In the last few decades, a large number of theories about cognition have been proposed in the philosophical community. Among these theories, when we focus our attention on the exploration of the boundaries of cognition, we can see that they form two different tendencies. A traditional tendency of cognitive science is that human cognition is only the result of the calculation of abstract symbols, so traditional cognitive science usually claims that cognition is computational.[1]A point of interest is that they believe that the boundary of cognition can only stay inside the body because they believe that cognition begins with the input of the brain and ends with the output of the brain, and they do not consider the external world.[2]29-30Another tendency is that the boundary of cognition is not only the body but also the environment can become a part of the cognitive process. This view was first mentioned literally in a book called *The Embodied Mind*[3], Andy Clark and David Chalmers (Abbreviated as C&C) developed the theory in *The Extended Mind*, introducing the idea of "extended cognition", where they argue that the external environment is part of human cognition, or in other words, that human cognition extends to the external environment! So, if technology has developed by leaps and

S. Yacob et al. (eds.), *Proceedings of the 2023 7th International Seminar on Education, Management and Social Sciences (ISEMSS 2023)*, Advances in Social Science, Education and Humanities Research 779,
https://doi.org/10.2991/978-2-38476-126-5_131

bounds, metaverse and Strong AI have become possible, can AI transcend the boundaries of the metaverse to construct a cognitive system with the external world?

In this paper, I will argue that AI in the metaverse is actually "metaverse-cavemen" who cannot cognize the external world. I will first try to explain the core arguments of the Extended cognition theory. Second, I will try to put myself in the shoes of the rebuttals and think about the questions they might ask about this topic. In the third part of the paper, I will discuss this topic from two perspectives, namely the "life-world gap" and the "life-experience gap", and finally conclude that AI cannot extend their cognition to the external world, they may be able to use the same discourse system as we do, but they may not be able to recognize the world we live in. Therefore, they cannot extend their cognition to it.

2 Extended Cognition and AI

2.1 Core Arguments of Extended Cognition

In Andy Clark and David Chalmers' 1998 article *The Extended Mind* they argue that human cognition is not limited to the brain, but extends beyond the body and that the body and the environment co-construct the cognitive process. "Where does the mind stop and the rest of the world begin?"[5]7 The question posed at the beginning of this essay actually shows this view of theirs. To illustrate Extended cognition, C&C designed a thought experiment: there is a computer screen that displays various two-dimensional shapes with "sockets" corresponding to these images, and a person sitting in front of the computer screen is asked to match these shapes with the "sockets". The first person was asked to mentally rotate the shapes to align them with the sockets. The second person uses a device (such as a keyboard, or a mouse) to do it. The third person has been implanted with some neural implant, so he can think as efficiently as the second person within the process.[5]

There is no doubt that we would all usually assume that both the first and the third person are actually the same in this cognitive process because both are calculated within the body. And apparently the same cognitive structure in the second person operating with the external device as in the third person implanted with the device. Because the difference between them is simply that the second person's cognitive processes run on a computer, while the third person runs them in his own mind via a neural implant. Thus, since we would consider the third person to be like the first person in that they are both cognitive, it is natural that the second person using the external device can also be cognitive, except that he does not have the external device implanted in his mind. So, "we cannot simply point to the skin/skull boundary as justification, since the legitimacy of that boundary is precisely what is at issue."[5]

Through this thought experiment, C&C argues that in fact, human cognition is heavily dependent on the external environment. They cite experiments conducted by David Kirsh and Paul Maglio in Tetris, where they calculated that it takes 1,000 milliseconds to simply mentally calculate the rotation of a physical shape, but that with the aid of an external device, "the physical rotation of a shape through 90 degrees takes about 100

milliseconds, plus about 200 milliseconds to select the button."[6] This is an example of how external objects can reduce our cognitive load for us.

According to C&C, "If, as we confront some task, a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is (so we claim) part of the cognitive process. Cognitive processes ain't (all) in the head!"[5]8. Although similar ideas had been proposed by others before C&C, such as Putnam, he propose a thought experiment called "Twin Earth" to explain that "meanings just ain't in the head! "[7]144 Like Putnam, there are other philosophers who believe that human cognition is greatly influenced by the external environment, but C&C continued to argue that these theories were not sufficient to express the true state of human cognition. In C&C's view, although these theories emphasize the importance of the external environment, they actually treat the cognitive process as passive reception, so C&C calls the previous view "passive externalism", and based on the critique of passive externalism, they propose and advocate "active externalism".[5] That is, actually human cognition is not passive, but the human organism is connected with external objects around it, forming a coupling system of two-way interaction, and this coupling system is a cognitive system. In the cognitive process of the coupling system, factors from the outside are active, and all parts of this coupling system play an active role and have a direct influence on the organism.

2.2 Discussion on the AI having extended cognitive abilities

Charmerls says "The iPhone is part of my mind already."[4] In fact, Charmerls speaks to the reality of human beings in modern society, the reality that we are inseparable from our surroundings, and we are always using the environment to help us speed up our cognitive efficiency and help us run our cognitive processes better. For example, when I need to go to the nearest subway station, I will use Google Maps and let it lead me there, but when I don't have Google Maps, the cognitive process I undertake will be much more complicated than now, and the processing efficiency will be much slower.

If we enter the metaverse, in fact, in the view of extended cognition, the metaverse is constructed as a coupling system with the human organism for extended cognition, and our cognition will still depend on the environment created by the metaverse.

For human beings, there is no doubt about the extended cognitive ability of human beings. However, after entering the metaverse era, is human being the only subject in the metaverse? Or furthermore, are humans the only ones in the metaverse capable of extended cognition? In the paper *"Extended cognition in plants: is it possible?*", Parise, Gagliano, and Souza argue that the available evidence suggests that plants extend their cognition beyond their bodies.[8] Thus, the ability to extend cognition is not actually unique to humans, or even to animals.

In a conceivable metaverse era, perhaps humans can share a metaverse world with the "natives" of the metaverse. The "natives" of the metaverse are those life forms that are already in the metaverse, and they are composed of computer algorithms and programs, they do not have a specific entity in the external world, but this does not affect their appearance as a life form. After all, who is to say that life forms must be carbon-

based forms like humans? Artificial intelligence has been advancing rapidly in recent years, and recently ChatGPT, a new type of AI language model, has been widely discussed. Further, if they have cognitive ability, can they recognize the world we humans are living in? According to C&C at *The Extended Mind,* "If, as we confront some task, a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is (so we claim) part of the cognitive process."[5]8 Then, perhaps the "natives" in the metaverse can also have this kind of capacity that is equal to that of human beings so that this kind of cognitive process can produce corresponding functions in the brain of the "natives" in the metaverse, then whether the AI is carbon-based or not, as long as he can have this capability, he can achieve extended cognition. It seems natural, therefore, to assume that AI has the ability to cognize the reality in which we live.

3 The two gaps: the reasons AI can't extend cognition to the external world

First of all, it must be admitted that there is no doubt that humans can use technology to speed up their processing efficiency, the same is true in the metaverse. For example, in the real world, most of us have difficulty reading the architect's construction blueprints, but in the metaverse, the architect's construction blueprints can be shown directly to us in a very tangible form, which greatly relieves the burden that human beings need to carry in the cognitive process. But for AI, it is difficult for them to perform this activity of extended cognition, or more radically, they cannot extend cognition.

3.1 Life-world gap: the disconnect between different world

Let us first examine the idea that AI in the metaverse is in a different world than humans. Of course, some people may disagree with this idea, such as, in response to virtual reality, Luo and Chu say: "However, we believe that since the virtual world is a human creation, it necessarily has a human structure, social relations, and historicity. It can thus be conceptualized as a continuation of the life world or a part of it."[12]34 Those who hold the same view would argue that AI in the metaverse is actually in the same world as we are, but merely we are in a different space of the same world. For example, I am in a classroom at a university in China and he is at a library at a university in England, we are in the same world, but we are in different spaces. And, in fact, when humans construct the metaverse world, in order to improve the immersion, we will make the metaverse world more realistic and experience stronger, perhaps in the metaverse, The experiences we experience are likely to be indistinguishable from reality, in this case, it seems that we say that the metaverse and reality are two different worlds is somewhat strong words.

I do not deny that human beings can experience exactly the same feelings in the metaverse as in reality, and I even agree that with the development of science and technology, human beings can achieve the same experience in the metaverse as in reality, and even that experience may be more perfect than in the real world, for example, the

metaverse can help people with disabilities to achieve the same life experience as the able-bodied people. But, in fact, throughout this whole process, we are looking at this from a human perspective and we ignore the "natives" of the metaverse. Why do we think we can feel the same as the experience of reality in the metaverse? Because it is seen from the human perspective, and even the metaverse is deliberately set up according to the human experience model, it naturally caters to the human experience itself.

But what about AI? Do they experience the same things that we do? Is their "life-world" the same as ours? Lakoff and Johnson argue that humans possess the concepts of up, down, left, right, etc. because of the structure of the human body, "Because human beings stand erect and human movement typically involves changing or maintaining this up-down orientation, humans develop or innately possess the concepts up and down."[9]462 But for the AI in the metaverse, their life-world is not the same as ours, and they experience life completely differently than we do, can they understand such orientation concepts as what is up and what is down? Of course, the designer of the metaverse will try to construct a metaverse that is close to the world we live in, even to the extent that its reality seems to us impeccable. But this is only our experience. We cannot actually understand what the metaverse of algorithms and programs looks like to AI, perhaps a string of code so long that we cannot understand, or maybe it is a new experience of life that we have not yet been able to understand. We have to admit that the world we live in is different from each other.

3.2 Life-experience gap: a "metaverse-caveman" hypothesis

Some doubts may arise here, and some may wonder, isn't AI just like us humans? AI is certainly capable of possessing the concepts of up, down, left, and right, without talking too much about the future, in the current video games, those game characters driven by programs in the games can also distinguish where is the road and where is the cliff. For the more advanced AI in the future metaverse, they can certainly do it, and can certainly experience the same experiences as we do.

But let's go back to "Plato's cave". In Plato's cave, there is a group of people who are bound to the cave from childhood, they face the wall all their lives. At that moment, there was a burning flame behind them, and someone was carrying different objects behind them, and because of the flame, the shadows of these objects were projected on the walls that these bound people were facing. So they always thought that these shadows themselves were real. Now, let's imagine a similar scenario where there is a group of people who have been trapped in a machine since they were children, and we are able to feed these people all the information about our world through this machine, assuming that the capacity of this machine is infinite and the speed of inputting information is as fast as the speed of human perception in reality, but the only difference is that the information input by this machine can only be algorithms and programs. This is actually a new-age Plato’s cave for what I might call "metaverse-cavemen". At this time, we are on the outside, through another special machine and these metaverse-caveman communications, and may even be trapped in this machine than normal people know more, and logic more clearly. In such a case, can we say that their life experience is exactly the same as ours? In fact, it is not possible. "Even if the cognitive level of the

intelligence agent improves beyond the ordinary human level and even possesses autonomy, the intelligence agent and humans are still in two different worlds."[10]97 Of course, they receive the same information as we do, but our life-experience is not the same, that is, normal people and people trapped in machines can use the concept of "walking" at the same time, but is the understanding of "walking" by people trapped in machines who have never really walked really the same as the understanding of "walking" by us? "Even if the linguistic intelligence of these robots is so developed as to allow them to grasp indirectly the schema of the human body through internal analogical reasoning, then the human social norms they thus understand have only an abstract (rather than practical) meaning for them, because such norms are really too far from their own 'life-experience' are too far away."[11]10

In short, this "life-experience gap" prevents them from knowing what the world looks like to us. Although they may be able to represent everything in our world in a very clear logical and precise concept, all they know is actually a "shadow" in Plato's cave, the only difference between the metaverse-caveman and Plato's caveman is that the metaverse-caveman is able to see the "shadow" in very high "resolution" due to the high-tech, but it's just a shadow after all, not the thing itself.

4 Conclusion

The metaverse as a hot topic has received much attention in the past few years. At the same time, the metaverse has led people to rethink the question of artificial intelligence. When we review Andy Clark and David Chalmers' theory of extended cognition, they argue that human cognition can extend to the world beyond the body. But perhaps humans are not the only ones who can do it, so let us consider if the metaverse does become a reality and there is AI in the metaverse, can they cognize the external world?

The conclusion maybe is not like that, because of the difference in our life worlds, it is likely that they understand the world completely differently than we do, even if we can understand each other's concepts of words. And because of the difference in our life worlds, we have not experienced the same experiences, so maybe logically we do understand each other, but do we really understand the same concept? Perhaps what they see as "walking" is a concept consisting of a string of codes, while for us humans, we understand "walking" as something other than a string of codes. Let's imagine that a person who doesn't know much Italian goes to Rome and asks someone where the Colosseum is by using very simple Italian words, and the grammar in his statement is even very messy, but it doesn't affect the communication. As an Italian, he can understand very well that the person is going to the Colosseum, but for the AI, they cannot know what the semantics are, they can only logically understand the world we live in. So, the communication process we have with AI is only logical, but as humans, we are not just animals based on logical cognition, we don't just rely on syntax, humans also rely on semantics to communicate with each other.[13]422

Therefore, the AI in the metaverse can't really understand what the external world is, and it can't construct a cognitive system with the external world, unlike us humans, our cognition has been extended into technology, and technology has become part of

our cognitive system, our brain, body, environment, and technology are a coupling system, and this whole system helps us to better cognize the world.

References

1. Shapiro, L. (2007). The embodied cognition research programme. *Philosophy compass, 2(2)*, 338-346.
2. Shapiro, L. (2019). *Embodied cognition*. Routledge.
3. Varela, F. J., Thompson, E., Rosch, E., & Kabat-Zinn, J. (1991). *The embodied mind*. (No Title).
4. Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. OUP USA.
5. Clark, A., & Chalmers, D. (1998). The extended mind. *analysis, 58(1)*, 7-19.
6. Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive science, 18(4)*, 513-549.
7. Putnam, H. (1975). *The meaning of" meaning"*.
8. Parise, A. G., Gagliano, M., & Souza, G. M. (2020). Extended cognition in plants: is it possible?. *Plant signaling & behavior, 15(2)*, 1710661.
9. Lakoff, G., & Johnson, M. (2008). Metaphors we live by. University of Chicago press.
10. Song C. & Li L.(2019).The Autonomy and Responsibility Bearing of the Artificial Intelligent Agent. *Journal of Dialectics of Nature (11)*,95-100.
11. Xu Y.(2017).Embodiment, Cognitive Linguistics and Ethics of Artificial Intelligence. *Journal of Shanghai Normal University (Philosophy & Social Sciences Edition) (06)*,5-11+57.
12. Luo Z. & Chu Z.(2022).Virtual Reality and Embodiment: A Survey Based on Embodied Technology. *Studies in Dialectics of Nature (06)*,29-35.
13. Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and brain sciences, 3(3)*, 417-424.