Philosophical Pragmatism and Design as a Workaround to Common Fallacies in Executive Function Enhancement through Artificial Intelligence: A Reflection on Neuralink and ChatGPT

Guadalupe "Lupita" Vásquez Fabela[1,2] Maximiliano Barajas Sánchez[3,4]

[1]Parsons School of Design, The New School, New York, USA
[2]Universidad Nacional Autónoma de México (UNAM), Mexico City, Mexico
[3]Universidad Autónoma Metropolitana (UAM), Mexico City, Mexico
[4]Université Gustave Eiffel, Marne-la-Vallée, France

vasql438@newschool.edu maximiliano.barajas@cua.uam.mx

ABSTRACT

The cycles of expectation-disappointment in AI can be seen as an invitation to make a reflection on the philosophical foundations of our approaches, especially when it comes to the development of technologies that are invasive to the human body. Technologies like Neuralink are based on the possible manipulation of chemical phenomena in the brain, but the approach involves an ontological framework where chemistry is subordinated to physics or considered "just applied physics". Since the previous statements are now being challenged by philosophers of chemistry and science in general, it is necessary to reevaluate the technologies that are grounded in reductionist approaches, especially when they are aimed at invading and controlling the body under the assumption that it is possible to "correct" any conditions that actually emerge as a concept in a different level of reality, such as human behavior, that can be influenced by factors other than just the brain activity, and even factors outside of the body. On the other hand, the use of LLM's like ChatGPT have been proven to affect mental health, and the impact can be attuned through a strategic user experience design based on the processes and stages where LLM's like ChatGPT are used for.

ARTIFICIAL INTELLIGENCE TODAY

When it comes to Artificial Intelligence (AI), there have been cycles of optimism and high expectations followed by disappointment, particularly around its general capabilities (Mitchell, page 1, 2019), and its potential of outperforming human beings in all sorts of tasks. It has come to light that the scientific community might be at risk of fallacious reasoning when making predictions about the performance of AI compared to human beings, particularly mereological fallacies, even if the knowledge of these fallacies has been there for decades.

(Mitchell, 2019, p3.) mentions two of the relevant fallacies to this case of study: Easy things are easy and hard things are hard. Intelligence is all in the brain

Easy things are easy and hard things are hard

This relates to the things humans execute with minimal thought or effort. There are things that anyone can do with ease, like "looking out in the world and making sense of what we see, carrying on a conversation, walking down a crowded sidewalk without bumping into anyone" (Mitchell) that paradoxically turn out to be hard for machines to do. Actually, this is called "Moravec's paradox," named after roboticist Hans Moravec, who wrote, "It is comparatively easy to make computers exhibit adult level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility" [39]." (Moravec cited by Mitchell)

This indicates that the level of complexity we project onto certain tasks is influenced by the perception of customary human capabilities, at least at a preliminary level.

Intelligence is all in the brain

The sentence "intelligence is all in the brain" represents a mereological fallacy. As explained by Smit and Hacker, "if someone commits the mereological fallacy, then he ascribes psychological predicates to parts of an animal that apply only to the (behaving) animal as a whole". The mereological fallacy is not strictly speaking a fallacy in the sense that it is not an invalid argument, but it is still an illicit predication that "leads to invalid inferences and arguments, and so can loosely be called a fallacy" (Smit H., Hacker P. M. S., p.1, 2014).

The validity of ascribing psychological properties to parts of an animal or human has been discussed from ancient times and it's still a debatable topic. (Smit and Hacker p. 2) Smit and Hacker then point out that this fallacy has been misunderstood: the fallacy "(...) does not imply that understanding the brain is unimportant for understanding mental phenomena". They then make clear that the participation of a part (e.g. an organ) in an activity doesn't imply that the activity can be ascribed to just that part.

"For example, we cannot walk unless the motor cortex is functioning normally, but that does not mean that we walk with our brain." (Smit H., Hacker P. M. S., p. 6, 2014)

Conceiving the brain as a kind of computer in which cognition takes place while assigning the body a secondary role of just taking part in the input (perception) and output (behavior) has been the predominant approach in AI. This approach is called the "information-processing model of mind" and emerged in psychology in the mid-twentieth century.

(Mitchell, p. 6, 2019). Mitchell (p. 7) mentions that AI has historically "aimed at a kind of 'pure intelligence' (...) independent of emotions, irrationality and constraints of the body such as the need to eat and sleep" and then she points out an interesting fact about the relationship between intelligence and irrationality:

"Related to the theory of embodied cognition is the idea that the emotions and the "irrational" biases that go along with our deeply social lives—typically thought of as separate from intelligence, or as getting in the way of rationality—are actually key to what makes intelligence possible."

The series of unfulfilled expectations behind AI might be a sign of a deep-rooted problem that might have something to do with conceptualizations. Setting up the foundations for a more adequate ontological framework for AI requires us to analyze the fallacies, or illicit predications in the case of mereological fallacies. There is an opportunity for reconceptualizing phenomena and optimizing contemporary research in the field of technological interventions in cognitive processes. In a time in history where Artificial General Intelligence is expected to be found outside of human body, it is pertinent to question the philosophical frameworks that make up the reality we want to manipulate.

CONTEMPORARY MENTAL HEALTH INTERVENTIONS INVOLVING AI: NEURALINK'S EXAMPLE

There are advocates that consider certain mental health conditions, like autism and ADHD (Attention Deficit and Hyperactivity Disorder), as simply different ways of perceiving and processing the world, rather than a pathology to be cured or inferior abilities that must be improved. Pathologizing these conditions would have more to do with the societal context and a set of unfulfilled expectations. (Patrick Dwyer [†], p. 1)

Beyond these possible ways of conceptualizing the problem, there is a strong factor that justifies the application of new conceptual frameworks: the possibility of developing a wide range of alternative treatments that improve the quality of life of patients and their ability to interact with their environments.

The latest research efforts usually involve AI, a promising tool to improve mental health issues, including cognitive deficits. A few years ago Neuralink, an emerging enterprise by Elon Musk, was introduced as a company aimed at improving neurological and mental health issues, like "autism and schizophrenia", through implantable brain-machine interfaces. That kind of symbiosis might seem at first as a solution to Moravec's paradox, where humans perform the easy tasks that are hard to automate, while the little devices can provide humans with the abilities that are easy to automate, but hard to perform at a human level. Still, this approach has not exactly escaped from the cycles of optimism and disappointment in AI ([†]). Neuralink, optimistically, mentioned at first that improving those conditions was one of their objectives, but the objectives were later redefined, at least temporarily, directing the advances towards the treatment of various kinds of paralysis ([†]). This is a more feasible advance, but not exactly groundbreaking: "systems with similar capabilities have been published in the peer-reviewed literature as long as 17 years ago (21 in 2025) and other simpler systems are being used clinically to treat a variety of neurological disorders" (Aswin Chari et al., p. 2, 2021).

That approach of Neuralink might still be guilty of fallacious reasoning: it seems like, in some way, the complexity of mental disorders is being reduced not only to the brain, but to electrical interactions. The fact that symptoms emerge in levels of reality that have more to do with day-to-day interactions than just electric interactions seems to be ignored, or compartmentalized.

There is also another important inconvenient with approaches like Neuralink and that is the invasiveness of the method: neurosurgery is considered by experts a risky procedure and many of them consider Neuralink's device implantation a procedure whose benefits are still not proven to outweigh the risks (p. 5, Chari A., neurosurgeon). These kinds of technologies are still being developed and the perceived weak spots might or might not be taken care of in later stages.

EXECUTIVE FUNCTION

There is a wide range of mental health conditions that might potentially be treated with technology, AI particularly. The approach of this article will be limited to executive dysfunction, a cognitive deficit associated with mental health conditions like ADHD, autism, Parkinson's disease, etc. Given the complexity and particularities of each condition, we'll limit the discussion to executive dysfunction in ADHD.

Executive function –also known as executive functions or executive functioning– is defined as the set of intellectual behaviours involving the frontal lobes, "including planning, self-monitoring, problem solving, reasoning, and working memory" (Ardila, p. 159). It's a subject of debate whether executive function is a component of intelligence. For Ardila, "(...) general intelligence is related to metacognitive executive functions but not to emotional/motivational executive functions" (Ardila p. 162).

He then proposes a characterization of these two executive functions:

"(a) "Metacognitive –or intellectual– executive functions" which includes temporality of behavior, problem solving, abstracting, planning, anticipating the consequences of behavior, strategy development and implementation, and working memory (the usual understanding of executive functions, generally measured in neuropsychology executive functions tests) (Ardila p. 160)

(b) "Emotional/motivational executive functions", which are responsible for coordinating cognition and emotion. In other words, they have the ability to fulfill basic impulses following socially acceptable strategies. (Ardila, p. 161)

A treatment to increase executive function by only taking into account the metacognitive or intellectual aspect of it might at least miss out on the opportunity to take advantage of emotional or motivational aspects; at worst, it might be deemed conceptually inadequate to work. Trying to influence emotions or motivation from the brain structure, conceiving it as a machine, might turn out at least more complex and perhaps way harder than influencing the intellectual component. It comes to our attention that executive function might not only be influenced by the brain's structure; it is also the context that plays an important factor in the processes that cause executive function to emerge.

The influence of context in conceptualizations

It's perhaps obvious that motivation and emotions are, in varying degrees, dependent on factors outside of the human body. Marketing can be a perfect example of that: certain offers or perceived benefits can influence our behaviors. Intelligence also emerges as a function of the context, although not exactly in the same way. Wechsler (1944) defined intelligence as "the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment" (Ardila p. 161). While it's not clear whether intelligence is only "caused" by internal (fisonomical) factors or external ones (the upbringing or the specific tasks, for example), it certainly depends on the context to be expressed or to emerge.

Since executive functions depend on metacognition/intelligence and emotions/motivation, and both are context-dependent, it follows that executive function is also a function of the context. It is interesting to note that executive function disorders emerge when it is time to complete tasks or goals. The diagnosis of this situation isn't made by x-ray, MRI or some other laboratory tests that would detect something abnormal in the body; it is the patients facing a strong impact in their daily functioning that takes them to the doctor. While it is true that treatments aimed at modifying something "internal", like medication and therapy, help patients improve their symptoms, it is also true that a person, be it a teacher, parent or caretaker, that reminds the patient what to do or what directions to take (external agents) might also improve the condition, as Hosenbocus proposes:

"Children with EF (executive function) disorders can achieve a sense of success and avoid getting into difficulties as long as they have support from another person, a parent, teacher, mentor or friend to act as a "surrogate frontal lobe" (the region of the brain associated to EF) to guide them and keep them on track." Hosenbocus S, Chahal R. A review of executive function deficits and pharmacological management in children and adolescents.

However, as patients grow older, it becomes more complicated and even looked down upon to obtain this kind of assistance: adults are expected to be independent. Can human-computer interactions be one solution for that? One of the aims of this article is to show the possibility of applying technology for improving cognitive function in a non-invasive way. This non-invasive way can take the form of a symbiosis AI-human not invading the body, like in the form of chips inserted in the brain, but as an external factor to help patients reach their goals and/or increase their executive function while maintaining independence. Even if some people might not consider it valid to call EF to such a setting, it would still reduce the negative impact of EF deficits.

PHILOSOPHICAL FOUNDATIONS OF AN EXTERNAL SYMBIOSIS BETWEEN HUMAN BEINGS AND AI

Advances in contemporary philosophy of chemistry are intimately related to technologies like Neuralink, based on the possible intervention to the signals between neurons. This communication between neurons is modeled based on the chemistry of neurotransmitters and their electrical signals. (†) Philosophy of chemistry plays an important role in this matter: it becomes a necessary tool to even begin questioning the validity or authority of technologies that from a reductionist standpoint aim to influence and control phenomena that might only emerge at a different level of reality.

What are the alternatives? An opportunity for philosophical pragmatism

According to Peirce, pragmatists deny that any conceivable description of reality could be sufficiently complete and exact for all purposes. We cannot really acquire a conception of reality as it is, which would require what some philosophers call the "God's-Eye View", referring to a capacity to conceive reality that goes way beyond human capabilities. Such a conception of reality is impossible for humans to experiment, therefore we are forced to settle for a limited conception of reality based on what we can access through our senses.

"Given that the privileged God's-Eye View does not exist, there is no 'true' ontology: all ontologies have the same metaphysical status because all of them are constituted by equally objective descriptions" (O. Lombardi, The ontological autonomy of the chemical world , p. 11).

It is expected to experience some degree of resistance after reading the previous paragraph. Putnam invites us to reflect on the false dichotomy that represents classifying reality as either objective or subjective. He liberates us from this false dichotomy through internalist realism, a posture that does not aim to conceive one true reality; instead, what we could conceive as our "objective" reality actually emerges when we create a conceptual framework in which we define objects. In other words, we can give up the duty of describing reality "as it is" and still be objective based on what we can access epistemologically through our senses. (Putnam, 1981, p. 49).

Once we become aware of this false dichotomy, we are free to explore different ontologies.

INTERNAL AND EXTERNAL RELATIONS ONTOLOGIES

Philosophers of science, like Bernal, have pointed out a fundamental difference between the ontological models derived from classical physics and ontological models that outperformed the latter at modelling phenomena where high selectivity was involved, like chemical reactions. Bernal points out that "corpuscularism belong to a class of ontologies that conceive relations as being external". That means that the identity of things is independent of their relations. "This allows for the qualities of things –e.g. shape and size – to determine their relations." (Bernal, †, p. 86)

Bernal argues that "when relations are selective, this turns into a hard and unrewarding enterprise" (Bernal, †, p. 86).

Internal relations ontologies fit this criteria:

“In an ontology of internal relations there is no object previous to its relations, as they constitute the thing. Qualities of the object cannot play any role in determining its relations, as the latter precede the former. Instead, relations of the object are used to construct its qualities. A theory about internal relations exploits complex relation patterns to construct rich characterizations of the entities.” (Bernal)

Trying to conceive certain phenomena, including chemistry, from external relations ontologies has been complex and unrewarding (Bernal). We could then say that chemistry is not reducible to physics, described by external relations ontologies. When it comes to cognitive processes, can we reduce them to physical or chemical processes? We presented these two ontologies to bring light to the restrictions and problems it represents the application of external relations ontologies, as well as the opportunities that internal relations ontologies can bring to the table. It is important to stress the fact that we are not trying to justify the reduction of cognitive processes now to chemical phenomena with a new ontological framework; we are justifying the adoption of new ontological frameworks to start exploring what remains unexplored, this time in human cognition. Ontological frameworks that require a reductionistic view of reality might have been successful at some level describing simpler phenomena outside of human bodies, but the validity of such frameworks remains unknown when it comes to describing processes taking place in human bodies.

We can certainly say that symptoms precede the diagnosis of EF impairments, which means we observe symptoms at a higher level of reality: the person and the context (the tasks to be performed). The truth is that most of the concepts used to describe the cognitive phenomena that the scientific community aims to explore nowadays emerge at a different level of reality. This level of reality involves both the person and the context. Technological advances can perhaps benefit from more holistic approaches.

To pinpoint what is wrong at a cell communication level can turn out complicated and perhaps unrewarding. Trying to selectively influence motivational, emotional, and metacognitive aspects based on electric signals and chemical reactions might add to the complexity. Neuralink or any technology that bases its approach in chemical phenomena can benefit from exploring different ontological frameworks that can represent an opportunity to overcome complexity.

A POSSIBLE MODEL TO START EXPLORING AND INFLUENCING EXECUTIVE FUNCTION

We previously defined executive function as the set of cognitive functions that involve planning, self-monitoring, problem solving, reasoning, and working memory. It is worth noting that all of those activities allow humans to reach goals. In the picture below we present a very simplified sketch of

a) a person whose executive function is considered “normal” in the sense that this person can reach his or her daily goals without much problem b) a person with executive function deficits impeding him or her to reach the normal goals c) a person with executive function deficits that either takes medication or gets an implant (like Neuralink’s implants). In this hypothetical case, the “machinery” of a person is modified so that executive function can reach “normal” or functional levels. d) A person with executive function disorders reaching the “normal” goals by getting assistance from another person or by an interface with AI.

The model is oversimplified given the immense variety of goals requiring different cognitive functions, and the human variability in terms of executive function to each one of the goals.

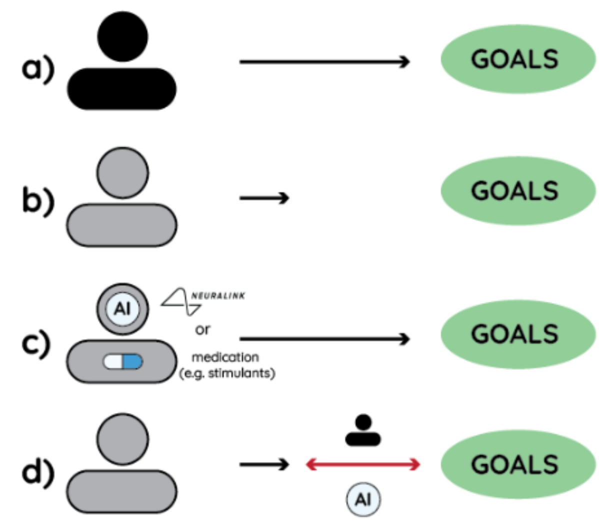

Figure 1: An overly simplified model for executive function.

THE ERA OF CHATGPT: RECENT FINDINGS

Recent research has shown that brain activity and cognitive ability are impacted by the use of LLM’s like ChatGPT. (Yao, Z. and Smith, J., 2025) In a study, the goal was writing an essay to explore how the use of LLM’s affect cognitive indicators, like remembering what was written after the task was done. The research showed that the cognitive impact of using LLM’s as assistants for the goal depends not only on the tool, but also on the stage of the creative process when it is used. Cognitive ability can be reduced, but also increased or stimulated.

This suggests that cognitive enhancement as an emergent property of a hybrid human-interface has a temporal element to be considered, temporal in the sense that it depends on the moment or stage of the creative process where it is utilized. This has already been suggested in the start of this decade by philosophers of science, (Vásquez Fabela, G. S., 2022), and it represents an undeniable opportunity to do research on design strategies that could improve productivity for users while minimizing the negative effects it could have on cognition.

CONCLUSION

While it would be complex to determine what approach is better, especially if we consider the fact that our approach is relatively recent and unexplored, we can present some clear advantages and disadvantages for each one of the approaches.

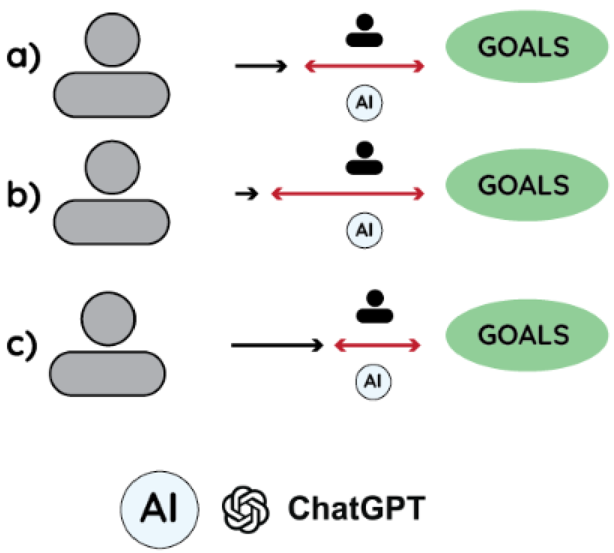

Figure 2: An overly simplified model for executive function influenced by LLM's (ChatGPT).

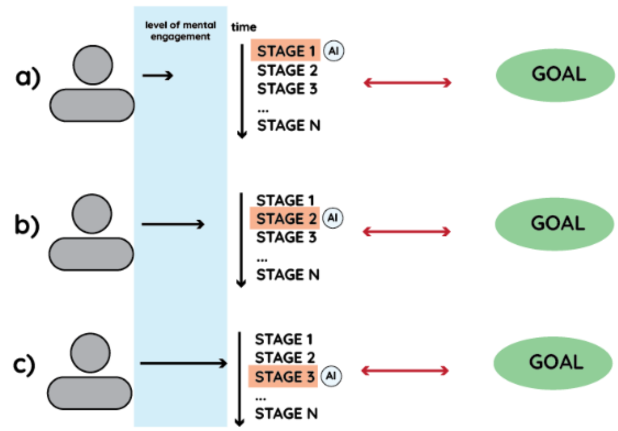

Figure 3: An overly simplified model for executive function influenced by generative AI, showing different levels of cognitive engagement depending on the stage where it is used in the pursuit of a goal.

The use of interfaces like Neuralink have proven effective for some tasks, like helping people with paralysis. However, there were also claims about "curing" schizophrenia or autism have been proven too ambitious over the years, in such a way that the company seems to have abandoned them. There is a reluctance to promote the use of these devices on a mass scale, since there needs to be an evaluation of the benefits vs. the possible complications of inserting a chip in the brain, resulting in the use of these interfaces by people with paralysis, mainly.

The use of LLMs, like ChatGPT and DeepSeek, has become popular in the past few years, influencing and shaping productivity at school and work. There is the undeniable benefit of an increase in productivity, but also concerns about the cognitive effect of the use of these technologies in the long term. This becomes increasingly important considering these technologies are increasingly used by kids, teenagers, neurodivergent individuals, and other vulnerable groups, not only to warn against its dangers, but also to explore opportunities that could represent a tangible benefit.

REFERENCES

Ardila, A. (2018). Executive functions: Concepts and clinical applications. Publisher Name.

Bernal, M. (2005). Internal versus external relations ontology in chemistry. Hyle: International Journal for Philosophy of Chemistry , 11(2), 85–104.

Bendix, A. (2019, November 5). Elon Musk says Neuralink could solve autism and schizophrenia. Business Insider . [†]

Chari, A., Henderson, N., & McMahon, G. (2021). Clinical translation of brain–computer interfaces: Hype versus reality. Journal of Neural Engineering , 18(4), 045001. Dwyer, P. (2022). Neurodiversity, autism, and the limits of pathology models. Philosophy, Ethics, and Humanities in Medicine , 17(4), 1–12.

Hosenbocus, S., & Chahal, R. (2012). A review of executive-function deficits and pharmacological management in children and adolescents. J Can Acad Child Adolesc Psychiatry , 21(3), 223–229.

Köse, U. (2018). Yapay zeka: Geleceğin biliminde paradokslar. Popüler Bilim Dergisi , 25(261), 12–21.

Lampos, V., Moura, S., Yom-Tov, E., Cox, I. J., McKendry, R., & Edelstein, M. (2020). Tracking COVID-19 using online search. arXiv Preprint , arXiv:2003.08086.

Lombardi, O. (2014). The ontological autonomy of the chemical world. Foundations of Chemistry , 16(2), 111–136. Mitchell, M. (2019). Artificial intelligence: A guide for thinking humans . Farrar, Straus and Giroux.

Moravec, H. (1988). Mind children: The future of robot and human intelligence . Harvard University Press.

Putnam, H. (1981). Reason, truth and history . Cambridge University Press.

Smit, H., & Hacker, P. M. S. (2014). The mereological fallacy. In P. Biggs (Ed.), Philosophy of mind today (pp. 25–49). Routledge.

Vásquez Fabela, G. S. (2022). Un enfoque pragmatista hacia nuevas ontologías dinámicas en química. Tesis de Licenciatura, Universidad Nacional Autónoma de México, Facultad de Química. https://hdl.handle.net/20.500.14330/TES01000828885.

Yao, Z., & Smith, J. (2025). Large language models in essay writing: Stage-specific effects [Preprint]. arXiv . [†]

Zhu, X., & Goldberg, A. B. (2009). Introduction to semi-supervised learning . Morgan & Claypool.

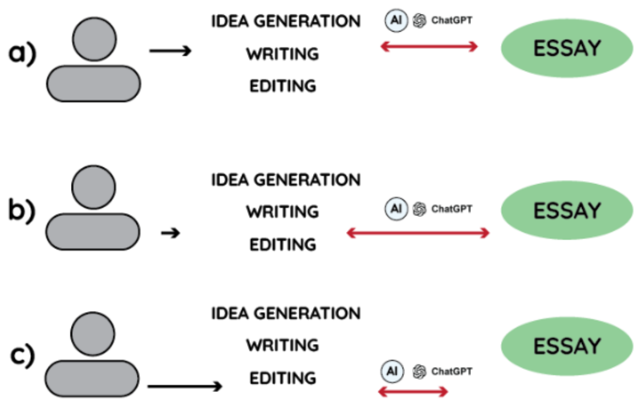

Figure 4: Figure 4. An overly simplified model for executive function influenced by LLM's (ChatGPT), showing different effects in cognition depending on the stage where it is used in the writing of an essay.