

Is ChatGPT Really Disembodied?

Simon McGregor

University of Sussex
s.mcgregor@sussex.ac.uk

Abstract

This article argues that the application of an embodied cognitive science perspective does not require us to distinguish between systems that have a physically tangible body and systems that do not. I consider the specific case of ChatGPT, a large language model specialised for interactive dialogue, and argue that ChatGPT can potentially be seen as embodied, albeit with a very unfamiliar type of embodiment.

I propose that we should explicitly distinguish between two notions of physicality: on the one hand, whether a system's body is *tangible* or not (roughly, whether we imagine it as providing us tactile-kinesthetic affordances); on the other hand, whether a system is *physically situated* or not (i.e. whether or not it interacts physically with the rest of the Universe).

I discuss whether or not tangibility should be accorded any major theoretical weight, within cognitive science, by considering several theoretical issues relating to embodiment: six from the previous literature, and four that I raise myself. My conclusion is that (at least in regard to these aspects of embodied cognition) there is no good theoretical reason to treat tangible bodies as a prerequisite for embodied cognition.

Hence, I argue that an interactive language model like ChatGPT can, in principle, perceive the world and interact with it just as physically as a squid or robot does (albeit less tangibly) through text channels, which serve as its physical sensors and actuators. Whether or not we should understand it as doing so depends on its behaviour, not on its substrate.

Embodied cognitive science rightly draws attention to the bodily relationship between an embodied agent (biological organism or robot) and its environment. Embodied agents face the task of negotiating real-time interactions with the world. This problem can be made harder or easier by aspects of an agent's body, which inevitably have profound effects on the emergent structure of the interaction dynamics between the agent and its environment. For instance, the difficulty of an obstacle avoidance task might vary significantly (and in unexpected ways) according to the position and orientation of the proximity sensors on a robot's body.

Attention has also been given to the role that the human body plays in shaping and facilitating supposedly 'abstract'

cognition. For instance, experimental work on 'overt head movement' in attitude formation indicates that participants' evaluations of statements and products can be manipulated by inducing them to move their heads in ways that resemble affirmative nods or negative head-shakes ([Tom et al., 1991, 2006, Wells and Petty, 1980]).

Despite the theoretical importance of the body to 4E cognition, *cognitive scientists do not really agree on what bodies are*. There is no widely-accepted definition of a body; the term is used with an intuitive rather than technical meaning. Wilson [2002] identifies six theoretically distinct strands of embodiment, while Ziemke [2003] identifies at least six different versions of bodies, and there are probably more.

In some relevant sense, embodiment is not a unified scientific concept. We rely on a pre-scientific notion of the body, drawn from a phenomenological understanding of the idea of bodies, as used in everyday life. This is highly problematic when it comes to cognitive science, which needs something more than (as it were) a folk theory of bodies. Otherwise, we may be led astray in our inferences about agents and their behavioural capacities.

One example of this problematicity is in relation to discussions of "disembodied" AI models such as large language models (LLMs). The general consensus (based on a shared intuition about what bodies are) is that, in contrast to robots, LLMs are not embodied. This consensus is unsurprisingly reflected in the output of LLMs themselves, which are trained to reproduce the sorts of things that humans say. For instance, consider the following transcript from a ChatGPT console (generated 20/12/2022 online).

PROMPT: Is ChatGPT embodied?

RESPONSE: ChatGPT is a language model developed by OpenAI. It is not embodied in the sense that it does not have a physical body or senses. It is a program that processes and generates text based on the input it receives. It does not have the ability to perceive the world or interact with it in a physical way. It is purely

a digital entity that exists within a computer system.

In this article I will consider several theoretically relevant aspects of embodiment, and challenge the view that it is helpful to treat LLMs like ChatGPT as disembodied along these dimensions. I will allow that ChatGPT does not have a ‘tangible’ body, in an obvious intuitive sense, but I will argue that this is not a scientifically important distinction - either at the level of physics, or of fundamental cognitive science.

It should be admitted that this article is not really about ChatGPT in particular; ChatGPT serves only as an individual representative of a broader class of AI systems that are widely assumed to be “disembodied”. Hence, I do not discuss specific technical properties of ChatGPT (such as the transformer architecture or the fact that it is a generative model).

My conclusion will be that we should see canonically “disembodied” AI systems such as LLMs, not as being “disembodied” or “abstract”, but as potentially possessing very alien bodies that render them meaningfully different from robots or biological organisms. In particular, I will propose that we should acknowledge and be mindful of their physicality. There may be some parallels here with critiques of so-called ‘digital dualism’, e.g. Jurgenson [2011].

Embodiment and Agency

In this section I will clarify the extent of my claims about LLMs such as ChatGPT. This paper is meant to defend the claim that ChatGPT is not ‘disembodied’; the question of whether ChatGPT is an ‘agent’ is beyond the scope of the current article. What I deny is the following claim: that *because* ChatGPT is ‘disembodied’, it cannot be an embodied agent.

Consider a robot body (say, that of Boston Dynamics’ Atlas robot). The robot’s sensors and actuators consist of cameras, motors, kinesthetic sensors and so on. A controller for this body maps a stream of data (from the robot’s sensors) onto a stream of data (that feeds into the robot’s actuators). Most such maps would not produce coherent behaviour on the part of the robot. For instance, imagine a map that fed bits from the sensors into a cryptographic hash function, and used bits from the hash to drive the actuators. This would probably look like a robot that did nothing more than quiver randomly. Let’s call this robot Quiver; we would not particularly want to call Quiver an ‘agent’. By contrast, the right sort of controller would produce a robot that runs and jumps in a lifelike manner, navigating obstacles and adapting to unexpected circumstances. Let’s call this robot Athlete.

I want to distinguish between two questions about ChatGPT:

1. Do Quiver and Athlete both possess something of foundational cognitive importance that ChatGPT lacks, namely a body?
2. Is ChatGPT more like Athlete, which seems to merit an attribution of agency, or more like Quiver, which does not?

This paper is directed towards question 1., and I will argue for the following answer: “ChatGPT lacks a tangible body, but this does not preclude us from treating it as a physically situated agent.” I believe that the answer to question 2. depends not on whether ChatGPT possesses a tangible body, but on whether it exhibits the right sort of behaviour to merit being treated as an agent. What the right sort of behaviour *is*, and whether or not ChatGPT exhibits it, is beyond the scope of this paper: my claim here is only that tangibility is not the core feature of 4E cognition.

Tangible and Intangible

This section introduces a distinction between ‘tangible’ (or perhaps ‘concrete’) and ‘intangible’ systems, and argues that this distinction is not the same as a notional distinction between ‘physical’ and ‘non-physical’ (or alternatively between ‘real’ and ‘abstract’).

Many of the ways in which a biological organism affects the world are what we might call ‘tangible’. Tangibility, in this informal sense, is related to the spatially coherent arrays of forces that macroscopic substances exert on one another while maintaining a perceived identity as the same substance. Such interactions usually involve displacement of one perceived object or substance by another. When a fish swims in water, or a bird picks up material for a nest, or a Venus flytrap closes its trap, or a human being types on a keyboard, these are all intuitively ‘tangible’ interactions in this sense.

It is common for communicative processes to operate using non-tangible pathways. For instance, when a person speaks, or raises a hand, the effects of these social signals are mediated by non-tactile and non-kinesthetic sensory systems (audition and vision) in other humans.

Chemical signalling via diffusing molecules is also of a sort that we could informally class as ‘intangible’. Once the molecules have been released, they do not possess any relationship that preserves spatially coherent forces. If you press on a pheromone, information about the direction and magnitude of the force is not transmitted at a macroscopic scale through other pheromones.

I hope the reader will grant that this distinction (between tangible and intangible) is an intuitively familiar one, and relates to our intuitions about what bodies are. Dogs have tangible bodies: we can reach out and pat them. We cannot reach out and pat a computer program like ChatGPT.

I want to make a slightly nuanced distinction between the class of things we can reach out and touch, and the class of ‘tangible’ objects. For instance, the bodies of microscopic organisms are tangible in the sense that they exert and transmit mechanical forces in a spatially coherent manner, but we cannot pat a bacterium, because they are too small¹.

It is not clear that the intuitive distinction between tangible and intangible phenomena reflects any meaningful difference at the level of physics. According to physics, unsurprisingly, everything is physical, and interacts with everything else through purely physical means.

So, for instance, ‘intangible’ sounds consist of compression waves in air that depend on forces by which atoms displace each other. In humans, inner ear bones transmit vibrations caused by sound waves to vibrations in cochlear hairs. But sound waves can also have non-auditory effects: even profoundly deaf people can detect loud low-frequency sounds through the bodily vibrations that these waves cause, e.g. in the torso. Pistol shrimps produce sounds - cavitation bubbles - with sufficient pressure to kill small fish.

Likewise, an instantiation of the ‘intangible’ computer program ChatGPT consists of patterns of voltages (and similar physical variables) in digital electronics. The electromagnetic force that moves electrons inside a computer is the same electromagnetic force that pushes back on our hands when we touch a table. The fact that humans cannot feel the ChatGPT program with our hands does not make that program meaningfully abstract or physically unreal.

I take this view to be a necessary implication of physicalism. I can interact with ChatGPT via my keyboard and monitor; therefore, ChatGPT is *physical* (despite it happening to be *intangible*).

Wilson’s Six Views Of Embodiment

Wilson [2002] provides an influential analysis of embodied cognition, identifying six theoretically important principles. In this section, I briefly summarise these principles, and provide some preliminary comments on whether they are applicable only to tangible bodies.

Wilson’s principles are as follows:

1. *Cognition is situated.*

Cognition inherently involves interaction with an external environment.

2. *Cognition is time pressured.*

¹It’s perhaps worth remarking that sensorimotor metaphors from human-scale tangible interactions are quite fallible at the spatial scale of bacterial embodiment, where unfamiliar forces such as osmotic pressure are relevant.

The world does not wait for an agent to finish thinking; it demands actions in real-time.

3. *We off-load cognitive work onto the environment.*

Interactions with the environment are used by embodied agents to process information.

4. *The environment is part of the cognitive system.*

The agent alone is not a meaningful unit of analysis.

5. *Cognition is for action.*

Cognitive functions must be understood in terms of their role in producing appropriate behaviour.

6. *Off-line cognition is body based.*

Cognitive processes are grounded in concrete sensorimotor competences, even when not interacting with the environment.

With the possible exception of principle 6 (*Off-line cognition is body based*), I am skeptical that these principles require an embodied agent to have a specifically tangible body (in the sense in the section “Tangible and Intangible”).

For instance, why should it be theoretically significant whether or not the physical interactions between an agent and its environment are tangible or not? As mentioned earlier in this article, many important interactions are intangible even for tangibly-embodied agents. Hence, Wilson’s principle 1 (*Cognition is situated*) seems equally applicable to intangible bodies.

Likewise, although interaction timing is a very important axis of distinction between modern chatbots and biological organisms, time-pressured cognition does not necessarily presuppose tangible embodiment. For instance, we can readily imagine future chatbots that could interrupt a user’s stream of input at appropriate moments to interject comments or clarifying questions. Such a capacity would involve real-time interaction, with all its ensuing complexities.

The story is similar for off-loading of cognitive work, or the environment being part of the cognitive system, or cognition being for action. None of these principles would seem to require a specifically tangible body. They could all, in principle, be applicable to a text-based system such as ChatGPT, whose physical instantiation is not a tangible one.

Wilson’s principle 6 (*Off-line cognition is body based*) is arguably a specific claim about cognition in humans (possibly with more speculative extensions to other animals and/or recommendations for robot design). It does not have an obvious application to intangible systems such as ChatGPT. However, I will discuss later (in the section “Scaffolding of

‘Abstract’ or ‘Higher’ Cognition”) how a generalised extension of this claim might be relevant to supposedly ‘disembodied’ chatbots.

The remainder of the article considers some alternative principles of embodied cognition, which I propose can be usefully disentangled from a tangible / intangible distinction. While they relate to Wilson’s principles, they are not identical.

Four Aspects Of Embodiment

In this section, I will introduce four specific ways in which the body is important for conventionally embodied agents such as biological organisms and robots, and discuss whether or not specifically *tangible* bodies play any central theoretical role.

- In biological organisms and robots, internal processes involved in cognition are not restricted to the brain or nervous system, and are spread throughout the body – and beyond ([Boem et al., 2021, Pfeifer and Gómez, 2009, Rowlands, 2009]). I argue that the question of where to draw the cognitive boundary is as relevant for ChatGPT as it is for more tangible systems.
- The body determines the sensorimotor channels between an agent and its environment, and affects the sensorimotor contingencies ([O’Regan and Noë, 2001]) associated with these channels. I claim that there is little to no fundamental difference between tangible and intangible agents in regard to sensorimotor embodiment.
- An embodied agent only needs to be competent with respect to the sensorimotor requirements of its *sensorimotor habitat* ([Buhrmann et al., 2013]) – the particular niche that it inhabits. I claim that this is no different for intangible agents.
- Supposedly ‘abstract’ cognition in humans (and perhaps other animals) is held by 4E cognitive scientists to depend on more tangible bodily competences ([Lakoff and Johnson, 2008, Lakoff and Núñez, 2000]). I argue that the capabilities of existing LLM models undermine the idea that such scaffolding is *necessary* on the level of an individual agent, and conjecture that a more general sensorimotor scaffolding principle might be relevant to intangible agents.

I do not discuss bio-enactivist notions of embodiment that rely on the concept of autopoiesis (such as [Di Paolo, 2005]), or assert some other continuity between life and mind (e.g. [Kirchhoff and Froese, 2017]). However, I think it’s worth remarking that computer programs under execution are patterns in matter and energy, that perpetuate themselves by degrading a thermodynamic gradient; to me, they seem closer to metabolically-defined biological bodies than the inert metal frames of robot bodies are.

This is not meant to be an exhaustive list of relevant dimensions of embodiment. See e.g. Ziemke [2003] for some other versions of embodiment, which could also be related to the notion of tangibility.

The Loci Of Cognition

One of the ways in which 4E cognitive science often differs from what we might call traditional neuro-cognitivism is in the role that is attributed to the biological brain. According to 4E cognitive science, cognitive processes are distributed throughout the body, right down to the ‘morphological computation’ performed passively by the musculoskeletal system in transmitting and transforming tangible forces ([McGeer et al., 1990, Pfeifer and Gómez, 2009]).

I propose framing this principle in a way that doesn’t rely on the notion of a tangible body: be open to surprising conclusions about where the ‘real’ cognitive work is being done. Some of it could be in the bones. This broader framing provides a continuity with notions of extended cognition, in which cognitive processes are taken to occur outside of the tangible body. In this way, it relates to Wilson’s principles 3. (*We off-load cognitive work onto the environment*) and 4. (*The environment is part of the cognitive system*).

While ChatGPT does not have a tangible body, the broader question of where to draw the cognitive boundary is nonetheless highly relevant to consideration of ChatGPT’s potential capacities and behaviour.

For example, one might naively assume that ChatGPT’s memory of a particular interaction is limited by the size of its context window (where the history of a conversation is internally stored). But this ignores the possibility that ChatGPT stores information *externally*, e.g. in the brains of its users².

Alternatively, when we consider the processes that modify the parameters of ChatGPT’s model (such as reinforcement learning from human feedback, or training on a new corpus of data), should we understand this as an interaction between a single system and its environment? In other words, is ChatGPT just the pre-trained model, or is it something broader?

I’ve emphasised that LLMs are physically instantiated systems that interact with the rest of the physical world. Just as 4E cognitive scientists do not assume that cognitive processes in humans are limited to the brain, we should be careful not to suppose that the cognitive processes of AIs are limited to what goes on in a particular hardware box or software system.

²Not to mention the states of other programs that it interacts with on a large automated scale in future.

Sensorimotor Embodiment

For the moment, let us assume that we have decided where to draw the boundary around a system that we call “ChatGPT”. For instance, suppose we consider a particular instance of the pretrained generative model, interacting with a user. For the sake of exposition, we will simplify the operation of ChatGPT’s actual algorithm. We construe it as follows: the user enters a text prompt, which forms ChatGPT’s input. An internal text tokeniser transforms the text prompt into model-readable tokens and adds them to ChatGPT’s context buffer. The generative model is run on the context buffer and a response is sampled from the generative predictions. This response is added to the buffer, detokenised, and output to the user.

Remember, I am not trying to argue here that ChatGPT behaves in a way that merits an attribution of agency. I am only concerned with the question of whether its lack of a tangible body is an important factor.

My claim is that, if the program’s behaviour *were* appropriate, we could treat the text input and output channels as sensorimotor channels. This has an obvious relevance to Wilson’s principle 1. (*Cognition is situated*).

- These channels may not be *tangible*, but they are as unmysteriously physical as the sensorimotor channels of squirrels or robots.
- We can apply the tools of *sensorimotor contingency theory* and *information agent theory* to these channels, to characterise their statistical properties and dependences, in just the same way we can apply them to the sensorimotor channels of tangibly-embodied agents.

Where it comes to sensorimotor embodiment, a common theme involves a focus on the properties of systematic relations between sensors and actuators. This unites *sensorimotor contingency theory* ([O’Regan and Noë, 2001]), the *information agent framework* (e.g. [Klyubin et al., 2007]), and work on *uninterpreted sensors and actuators* in robotics (e.g. [Olsson et al., 2006]).

From this perspective, what really matters is the statistical properties of the combined sensorimotor trajectories induced by an agent’s embodiment. The sensors and actuators could be camera pixels, bacterial flagella, proprioceptors, loudspeakers, whatever. The only reason that differences matter between these physical sensorimotor substrates is because they induce different statistical relations between the channels.

Sensors in robots and biological organisms are very diverse, and their activations depend in turn on complex emergent features of embodied cognition. Consider signals from nutrient chemosensors on the surface of a sea anemone tentacle, by comparison to an array of microphones mounted on

a wheeled robot.

The statistical properties of these channels would be very different even if the agents in question were completely passive. Since these agents are not passive, we need to additionally factor in the patterns of dependency on relevant actuators in a complex agent-environment loop; when we’ve done this, the sensorimotor contingencies of chemosensors and microphone arrays are likely to look as different as those of chemosensors and ChatGPT’s text input channels.

So I conjecture that there is nothing intrinsic about text token channels (whether inputs or outputs) that justifies treating them as unlike the sensors of tangibly embodied agents. Of course, they represent a characteristically distinct interactive modality, but this modality is legitimately *physical* despite ChatGPT’s intangible embodiment.

Perhaps there are mathematical properties of pure-text sensorimotor contingencies that always differ from properties of sensorimotor contingencies in tangibly-embodied agents, and justify making an important cognitive distinction between text-based agents and tangibly-embodied agents.

But if so, then cognitive science should aim at discovering this fact, rather than presupposing it. Otherwise, we are simply smuggling in features of our pre-scientific intuitions as a priori assumptions. I regard it as highly implausible that we are currently in a position to confidently assert anything meaningful about the relation between pure-text sensorimotor cognition and conventionally-embodied sensorimotor cognition.

Embodied Niche

In the previous section, I discussed how an agent’s embodiment structures its sensorimotor contingencies - the statistics of its sensory signals and their dependency on its motor signals.

But an embodied agent has a ‘niche’ (usually associated with conditions on the maintenance of its existence) that it tends to inhabit, and its interaction dynamics actively tend to keep it within that ‘niche’.

In this way, each embodied agent has a ‘self-centred world’ that represents only a highly restricted range of the physical Universe’s true complexity. This relates to the notion of *umwelt* von Uexküll [1934] in ethology, or the concept of *sensorimotor habitat* introduced by Buhrmann et al. [2013]. A similar idea is found in Friston [2013].

One consequence of this is that, insofar as we attribute ‘internal models’ to agents, these models do not have to be particularly accurate to be effective. Roughly speaking, they only need to be (approximately) accurate when it comes to those parts of the world that the agent typically

samples. When it comes to circumstances that are almost never encountered, the agent's 'model' can be wildly inaccurate, with little impact on the agent's behavioural competence.

Thus, even if we construe embodied intelligence in terms of internal modelling (which many 4E cognitive scientists resist), it is not really about basing one's behaviour on a widely accurate model of the world. Rather, it is about somehow possessing a model that produces behaviour which keeps the agent within a favourable subspace of the world.

This principle would seem to apply just as much to intangible agents as it does to tangibly-embodied agents. Sensorimotor interactions with the external world only sample a small part of that world, and purposive interactions (or apparently purposive interactions) shape and restrict that sampling regardless of whether they are via tangible or intangible channels.

The world does not (so to speak) present itself in its entirety to any agent. We should be mindful that this may be as relevant for ChatGPT as it is for mayflies and oak trees.

Scaffolding of 'Abstract' or 'Higher' Cognition

There is abundant evidence that, at least in humans, abilities that we think of as 'abstract reasoning' are heavily underpinned by competences with the tangible world (e.g. [Lakoff and Johnson, 2008, Williams et al., 2009, Meier et al., 2012]). This is related to Wilson's principle 6. (*Offline cognition is body based*).

So it may be surprising to some 4E cognitive scientists that supposedly 'disembodied' agents such as ChatGPT are capable of generating complex verbal behaviour, of the sort that we would consider, if emitted by a human, to be evidence of abstract reasoning.

To be clear: the behaviour of modern LLMs in no way undermines the claim that, *in humans*, linguistic and mathematical reasoning can only occur by co-opting mechanisms which evolved (or were learned) to solve more tangible embodied problems. Moreover, LLMs are trained to reproduce patterns in human language usage; hence, if their outputs do indicate reasoning abilities (which is up for debate), then those capabilities are historically dependent on human tangible embodiment.

However, in the past it might have been possible to imagine that the lack of a tangible body might prevent AIs from writing satisfactory undergraduate-level essays involving chains of 'abstract' argument. Such a hypothesis is no longer really viable (although it is still open to debate whether such essay-writing really requires human-like abstract reasoning). At the very least, the current capabilities of LLMs cast doubt

on whether 'abstract' reasoning requires tangible embodiment.

I don't think it is very clear what this tells us about the importance of a tangible body for cognition. What I do want to emphasise is this: the undeniable traces of human tangible embodiment in supposedly 'higher' or more 'abstract' cognition do not necessarily establish that tangible embodiment is the important factor in that phenomenon.

Instead, it may be more productive to consider a general principle of 'scaffolding' whereby competences in more-practiced domains can sometimes be generalised to less-practiced domains (albeit in ways that introduce detectable biases).

So in the human case, the idea would be that competences in bodily locomotion, visual perception, object manipulation, and so forth, facilitate abilities like mathematical reasoning by a scaffolding process. My suggestion is that the tangible nature of human embodiment may not be an essential part of such a story; perhaps we see the marks of the tangible body on human cognition because humans are tangibly-embodied (unlike LLMs).

I've claimed that text token streams could, in principle, be seen as primitive sensorimotor channels for a language-based agent. The cognitive bread and butter of such an agent would be sensorimotor interactions with the world along such channels, which have different affordances than the human embodied sensorimotor interface.

Could such an agent, primarily competent in natural-language tasks, acquire competences in a domain with different characteristics? For instance, could it learn to control a robot arm (via a text interface)? In principle, perhaps it might - and it doesn't seem inconceivable that it could do so by exploiting parallels with its existing capabilities in the linguistic domain. In turn, this might introduce detectable biases in its control strategies - analogous to the impact of embodied cognition on human 'abstract' reasoning.

While this account is highly speculative and a little fanciful, I hope it is at least plausible that tangible embodiment might not be the key feature of cognitive scaffolding, and that a more abstract sensorimotor scaffolding phenomenon might - in principle - be relevant to understanding intangible agents.

Avoiding Dualism

This article proposes a highly permissive notion of embodiment, under which any physically implemented system can be considered embodied. Some might consider that this trivialises the concept of embodiment; however, I maintain that the permissive notion of embodiment is non-trivial because it emphasises that it is theoretically essential to consider *how* systems are embodied.

The core claim in my argument is as follows: embodiment is mainly concerned with how the interaction dynamics between a system and its environment are structured in real-time, and the existence of such structure does not require any particular type of physical substrate. It is true that LLMs do not have tangible bodies, i.e. we cannot reach out and touch them in the way that we can touch a robot's body, or a cheetah's. But we should not suppose that this implies LLMs are "abstract entities" which do not interact with the external physical world in significant ways.

Just as Descartes believed that mind and body were composed of two fundamentally different types of substance, which interacted mysteriously in the pineal gland, the claim that ChatGPT is "purely abstract" suggests that it is made of some non-physical abstract stuff that interacts mysteriously with the physical world. This would be profoundly misguided, since ChatGPT is physically implemented in ordinary matter and energy. ChatGPT implementations, and their interactions with the rest of the world, follow the laws of physics in as banal and un-mysterious a manner as rocks.

One might object that LLMs do not interact with the physical world in a meaningfully similar way to tangibly-embodied agents. This claim has been sometimes used to support the position that LLMs do not really understand the physical world. Again, I think this claim begs the question, and risks simply reproducing intuitions about language and bodily activity that exhibit a problematic dualism. The intuition here seems to be something like, human beings understand what chairs are because they can sit on them and look at them. LLMs cannot not understand what chairs really are, because they only have access to natural language text with the word "chair" in it. On this account, LLMs only learn about language; they cannot learn about the real world.

But humans do not see chairs by some magic process of unmediated mental contact with the world. Our contact with chairs is mediated by our embodied sensorimotor channels. There are no little chairs travelling along our optic nerves, to be picked up and felt over by tiny homunculi in the brain. There are just action potentials, which are no more intrinsically chair-like than are digital tokens encoding the word "chair".

It's certainly true that humans interact with chairs far more directly than LLMs. Those action potentials in the optic nerve respond to detailed properties of particular chairs at rapid time scales, in a rich manner that connects to a myriad other sensorimotor channels such as tactile signals and muscle activations. Chairs, and interactions with them that are sensitive to fine-grained details about specific chairs, are part of everyday human life. Not so for ChatGPT. Don't humans thereby connect to real chairs, unlike LLMs which

only know text, and not chairs at all?

I don't think so, and I think that such a claim would presuppose an arbitrary distinction between text and the world, as though computer programs processing text were not physically implemented, their inputs were not caused by the physical world, and their outputs did not causally affect the world.

As far as I can make out, this text/world distinction seems to follow an intuitive prejudice whereby tangible objects such as rocks are classified as "real" and less tangible entities such as computer programs are classified as "not real". The physical sciences don't seem to entail any such metaphysical distinction. To reiterate an observation made in the earlier section 'Tangible and Intangible', the electronic processes in my laptop, that correspond to the execution of a computer program, are as physically real as rocks. If they weren't, my laptop wouldn't work.

When we reframe the inputs and outputs of LLMs as physical causal channels, carrying information that can be related to a normative/functional framework of appropriate behaviour, the picture looks very different. To say that LLMs don't understand the world because they only pick up on regularities in text would be like saying that humans don't understand the world because they only pick up on regularities in their various sensorimotor signals. ChatGPT's training data, with its "purely linguistic" regularities, is not some ethereal structure existing in a purely abstract realm separated from the physical world.

Of course, we need to avoid a naive view which supposes that ChatGPT's training data and inputs 'tell' it about the world in a way that presupposes a relationship of linguistic understanding (in the sense we would apply to humans). In other words, the information that an AI agent acquires about the world through (text-based) sensorimotor interactions with that world might well not resemble the linguistic meanings that a human would ascribe to that text.

Likewise, as described in the earlier section 'Embodied Niche', even if we find it helpful to consider ChatGPT to have an 'internal model' of the world, we should be very careful before we assume it aligns with human ways of construing the world. ChatGPT's *umwelt* may be significantly different from the human *umwelt*; if chairs occur within ChatGPT's *umwelt* (as it were), they may do so in a very alien form.

So I'm not claiming that text-based agents understand the world by applying a standard natural language interpretation to their inputs. The claim is only that the "purely linguistic" regularities in ChatGPT's training data reflect genuine things in the real world. Had the world been different, the training data would have been different, because that data is physically part of the world, and is caused by relevant features

of the world. Hence, formally speaking, text data carries (Shannon) information about the physical world in exactly the same sense that the optic nerve carries (Shannon) information about the world.

Implications

I have argued in this paper that a significant number of theoretical concerns from embodied cognitive science do not require the presupposition of a ‘tangible’ body, i.e. one that can displace other substances and be displaced by them, in a manner that involves spatially coherent arrays of forces. They only require that a body be ‘physical’, in the sense of having a physical instantiation (and thereby, by necessity, engaging in physical interaction with the rest of the Universe).

What implications follow from these arguments? Firstly, I propose that embodied cognitive science needs some theoretical adjustments. At the very least, we should not conflate ‘tangible bodies’ with ‘physical bodies’, because a large class of genuinely physical phenomena are intangible.

Furthermore, those who claim that specifically tangible bodies are essential for embodiment need to explain (or at least gesture more definitely towards) what tangible bodies have that is supposedly special.

Perhaps more importantly, if I am correct that there is nothing fundamentally ‘disembodied’ about physically instantiated systems which lack tangible bodies, then we can (and probably should) apply 4E cognitive science, and the full range of embodied cognitive science tools, to intangible AI systems such as ChatGPT.

On close inspection of their behaviour, we may of course decide that it is not helpful to see them as agents, in the same way that we might not ascribe agency to the Quiver robot I described earlier (in the section ‘Embodiment and Agency’). But we should not presuppose that they are not embodied agents simply (as it were) because we cannot touch them with our hands.

One practical consequence here is that we should not leave discussions about AI safety in the hands of cognitivists, who do not recognise (for instance) that *The agent alone is not a meaningful unit of analysis* or that *[Embodied agents] off-load cognitive work onto the environment*, or who suppose that widely accurate models of the world (as humans understand the world) are required for successful interaction with that world. We should actively pursue conceptions of AI risk and agency within a 4E cognitive science perspective.

Summary

I have discussed whether or not language-based systems such as ChatGPT should be considered ‘disembodied’ in any sense meaningful to cognitive science. In order to clarify the discussion, I introduced a distinction between ‘tangible’ and ‘intangible’ (but still physically instantiated) systems, which was meant to capture some of the intuitive aspects of ‘physical embodiment’.

I considered the importance of tangible embodiment for a number of aspects of embodiment that are important to 4E cognitive science: six from Wilson [2002] and another four which I thought were relevant. For each of these aspects, I presented reasons to doubt that tangible bodies are of core theoretical importance. Future work can expand this analysis by considering other aspects of embodiment (see, e.g. those described in Ziemke [2003]).

More generally, I cautioned against dualism when discussing intangible systems, and discussed some of the implications of my view, including relevance to theoretical cognitive science as well as AI risk.

Interactive language models such as ChatGPT undoubtedly have distinctive properties that are important from a 4E perspective. For instance, the *timing* of a chatbot’s interaction dynamics (along sequential text input-output channels) has profound implications for its behavioural capacities. But robots and biological organisms possess a diverse spectrum of different embodiments and ecological niches, whose unique characteristics are all important; I propose that LLMs are in principle no different.

I deliberately left a crucial question open: should we consider LLMs in general (or ChatGPT in particular) to be embodied agents? There is a great deal to be said on that topic, which far exceeds the scope of this article, and depends on particular details of ChatGPT’s behavioural capacities (not to mention how we conceive of agency, which is a contested notion). However, if my arguments are correct, we are not entitled to assert that ChatGPT is not an embodied agent merely because it is intangible.

Acknowledgements

Many thanks to Joe Gough and Chris Buckley for helpful comments on previous drafts of the article.

References

- F. Boem, G. Ferretti, and S. Zipoli Caiani. Out of our skull, in our skin: the microbiota-gut-brain axis and the extended cognition thesis. *Biology & philosophy*, 36:1–32, 2021.
- T. Buhrmann, E. A. Di Paolo, and X. Barandiaran. A dynamical systems account of sensorimotor contingencies. *Frontiers in psychology*, 4:285, 2013.

- E. A. Di Paolo. Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, 4(4):429–452, 2005.
- K. Friston. Life as we know it. *Journal of The Royal Society Interface*, 10(86):20130475, 2013.
- N. Jurgenson. Digital dualism versus augmented reality. *The Society Pages*, 24, 2011.
- M. D. Kirchhoff and T. Froese. Where there is life there is mind: In support of a strong life-mind continuity thesis. *Entropy*, 19(4), 2017. ISSN 1099-4300. doi: 10.3390/e19040169. URL <http://www.mdpi.com/1099-4300/19/4/169>.
- A. S. Klyubin, D. Polani, and C. L. Nehaniv. Representations of space and time in the maximization of information flow in the perception-action loop. *Neural Computation*, 19(9):2387–2432, 2007.
- G. Lakoff and M. Johnson. *Metaphors we live by*. University of Chicago press, 2008.
- G. Lakoff and R. Núñez. *Where mathematics comes from*, volume 6. New York: Basic Books, 2000.
- T. McGeer et al. Passive dynamic walking. *Int. J. Robotics Res.*, 9(2):62–82, 1990.
- B. P. Meier, S. Schnall, N. Schwarz, and J. A. Bargh. Embodiment in social psychology. *Topics in cognitive science*, 4(4):705–716, 2012.
- L. A. Olsson, C. L. Nehaniv, and D. Polani. From unknown sensors and actuators to actions grounded in sensorimotor perceptions. *Connection Science*, 18(2):121–144, 2006.
- J. K. O’Regan and A. Noë. A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences*, 24(05):939–973, 2001.
- R. Pfeifer and G. Gómez. Morphological computation—connecting brain, body, and environment. *Creating brain-like intelligence*, 5436:66–83, 2009.
- M. Rowlands. Extended cognition and the mark of the cognitive. *Philosophical Psychology*, 22(1):1–19, 2009.
- G. Tom, P. Pettersen, T. Lau, T. Burton, and J. Cook. The role of overt head movement in the formation of affect. *Basic and applied social psychology*, 12(3):281–289, 1991.
- G. Tom, E. Ramil, I. Zapanta, K. Demir, and S. Lopez. The role of overt head movement and attention in persuasion. *The Journal of psychology*, 140(3):247–253, 2006.
- J. von Uexküll. *Streifzüge durch die Umwelten von Tieren und Menschen*. Springer, 1934. Read in English trans. J.D. O’Neill, “A Foray Into The Worlds Of Animals And Men”, 2010.
- G. L. Wells and R. E. Petty. The effects of over head movements on persuasion: Compatibility and incompatibility of responses. *Basic and applied social psychology*, 1(3):219–230, 1980.
- L. E. Williams, J. Y. Huang, and J. A. Bargh. The scaffolded mind: Higher mental processes are grounded in early experience of the physical world. *European journal of social psychology*, 39(7):1257–1267, 2009.
- M. Wilson. Six views of embodied cognition. *Psychonomic bulletin & review*, 9:625–636, 2002.
- T. Ziemke. What’s that thing called embodiment? In *Proceedings of the annual meeting of the cognitive science society*, volume 25, 2003.