

Why the Vocabulary of AI Cannot Be Fixed by Argument: A Baconian Alternative

2026-09-04

Many put the right diagnosis on why debates on the consciousness of LLMs, on their intelligence or the individuality of their agents, are leading nowhere: we do not yet know what these properties really mean *for a LLM*. We do have plenty of definitions, but most seem to be contextual and not able to capture what happens with this new artifact called an LLM. My point is that this is a bigger problem of how do we use natural and scientific language, and its limits in the face of a new phenomenon, and not strictly one of definitions. Two decades of effort put into this attempt to better talk about AI, failed for a more similar reason than one may suspect at first, one that Bacon analyzed at the dawn of early modern science in *Novum Organum*: we are forced to use words from natural language that describe phenomena in a subjective way. Bacon saw that a new language is needed, one in which the words are anchored in objective experiments, and not in subjective perception. He failed to apply his project to “heat”, mainly because you cannot build a new language word by word. But his lesson stood, and was incorporated by Newton and other into what will become “scientific language”: you cannot fix scientific terms by argument. It happened again when QM emerged and natural language spilled back in because nothing else was available; the atom was “a planetary system”, or, later on, when Einstein and Schrödinger found it intolerable that “a particle has no position”. The situation today in AI is eerily similar.

Let us start with a clear example. Is a cat conscious? No doubt, no matter what definition or set of criteria we employ. But why is it then that most current thinking on machine consciousness assumes that the right kind of language processing would itself amount to consciousness? Many conscious beings may not ever talk, and yet, we consider coherent talk as a mark of consciousness. Is it? Why?

The problem traces back to Block’s split: phenomenal and access consciousness. “What is like to be something?” vs. “a state available for reasoning, for speech, and for rational control of action”. This is already scientific, and not natural language.

It is *access consciousness* that is invoked in LLMs’ case, because it appears that

there is *something* that is accessed to produce coherent speech. This has its own problems, because as many observed, this is our intentional stance at work: we do not really understand why they are so coherent and effective, and we basically project on them a reason so that they would make sense to us.

But we'll not get into this. We'll just observe a strange fact: almost nobody says "LLMs have access consciousness" instead of "LLMs are conscious". There are reasons for that, the term is not made for the phenomena we are dealing with, so that it is stretched, and bended, up to the point of becoming a new problem rather than a solution.

Block's definition of access consciousness seems flawed, because it allows the extension of "rational control of action" to the basic minds. For many, a cat has "rational control of action" when hunting mice, a wholesale import of content and representationalism into basic minds. The consequence is that reasoning and speech sit at the same level as the cat's "rational control of action", making *access consciousness* available to both humans and basic minds. Why then, not to talking machines?

This view can be attacked in many ways, from the basic minds' side, but one can do it from the human side also: Does Verstappen when taking corners at 200 km/h, avoiding late cars and some times accidents in front of him, uses *access consciousness* by "rational control of action?" Does a great pianist, when playing a difficult score, invokes "rational control of action?" I am quite sure that if they would do it, they would perform rather badly.

If we take out this rather strange faculty of access consciousness called "rational control of action", we are left with reasoning and language, that is, with language. And we can better redefine *access consciousness* as whatever manipulates language to describe or to reason about our perceptions. (I will not get here into why you cannot reason without language). In a way, the access consciousness overlaps with that little voice in your head that is incessantly reasoning about things in the world and what do you need to do. Verstappen, when overpassing, is not relying on his little inner voice, text is a too narrow a channel to manage the huge amount of impressions, feelings and intuitions that he needs to take in and much too slow for that. (By the way there are studies, the best F1 drivers are not those with faster reflexes but those that can intuit events before events happen. They do not *react*, they *predict*) And, of course, cats hunting mice do not have little inner voices, reasoning about trajectories; they are more like Verstappen.

If we agree that access consciousness is the one incessantly reasoning about the world using words, the next question will be: does an artefact that is reasoning about the world using words has `access_consciousness`?