On November 30, 2022, ChatGPT was released by OpenAI. Bernadette Matthew, a Ph.D student at the Indian Institute of Technology Delhi, summed up its surprising competence: “Chatting with ChatGPT is like chatting with a real person. If I had known this earlier, I could have saved myself so much time and work.” It became literally the most rapidly-adopted consumer product, of any type, of all time. Having tracked the AI conversation for over a decade, at that point, I'd call it the single most watershed moment of the global conversation. It definitely became easier to get audiences with political staffers.

ChatGPT(as of the May 2024 version as we write this) is still visibly stupid, if you know how to come at it from the right angle to provoke it into predictable stupidity. When I ask the May 2024 ChatGPT to solve physics problems involving bouncing balls and momentum, it makes simple algebra errors and screws up the math *qua* math.

Those particular examples might or might not still be around at the time you, the reader, crack open this book. When the first versions of AI-drawn images came out, people noticed that set of AIs was bad at drawing hands, and put out viral images like this:

Less than a year later, DALL-E has figured out how to draw hands:

AI incapability is a moving target.

[break]

But if you're still alive to read this book, I'd bet that the current set of consumer-accessible AIs probably have some set of stupidities, that dedicated online probers have managed to turn into neat viral examples. And yeah, maybe if you took some poor ordinary human with the sort of amnesia where they can't form new memories, and had thousands of people repeatedly probe them to find questions to which they'd repeatedly give stupid answers, then you could make that human look stupid too. But (as of May 2024) I think the stupidity is real; ChatGPT 4 is not quite up to human par.

ChatGPT 4 wasn't the sort of device where, if anyone builds it, everyone dies. I don't just mean this in the trivially observable sense that I survived to write this, but in the sense that it wasn't what we were originally worried about two decades earlier.

So what then is this "intelligence" stuff, that we are worried may be made "artificially", if it's not just the property of "talking like a person"?

[break]

I will not start by defining intelligence in words. If a stone-age human was trying to discuss fire with their fellow tribe-members, they'd do better by pointing at the hot flickery red-orange stuff and saying "let's talk about that stuff over there", rather than saying "I define 'fire' to be the physical manifestation of the fire-god's anger".

The phenomenon precedes any attempt at a definition. We'll give a definition afterwards, but don't make the mistake of the caveman who chortles "You can't even define fire precisely, nor does your theory account for the fact that some fires burn blue. Therefore, your claim that I should flee before the forest fire consumes me is bunk."

The forest fire is allowed to burn you to ashes, and yes, it is allowed to burn you even before anybody has presented you with a satisfactory definition of fire.

[break]

So I begin by pointing at "intelligence" as it exists within the world.

I gesture at humans, and at mice. Mice build nests in the walls of our houses. We fell forests to erect those houses, and we also mine metals and erect skyscrapers. This is a difference between us and mice. (If this sounds obvious, that's fine, it's meant to be; there's nothing wrong with beginning from the obvious, when somebody asks you to point at things.) If you look at the visible processes leading up to the difference, it's that humans have discovered truths about metals and metallurgy, which, even if somebody tried to explain to a mouse, the mouse could not learn. This is widely believed to have something to do with a difference between mouse brains and human brains. We agree with this widespread belief. We point to the different powers of mouse brains and human brains, the different things they can learn and do in the world; this is not a definition of intelligence, but it is a beginning thing to point to in the world, when we want to say what it is that we are talking about.

Humans have no rivals among particular animal species in the world, when it comes to building skyscrapers. You could, however, say that humanity already has a rival at figuring out interesting tricks, an ancient rival, far older than humanity itself: namely, the process of natural selection. DNA encodes blueprints for making animals, sometimes errors happen in those blueprints, sometimes the resulting animal does a better job of reproducing itself and replicating its DNA; hundreds of millions of years after this process got started, it coughed out mice. Humanity built skyscrapers, but even we have never built a whole entire mouse. Our biotechnology industry is barely getting started on making tweaks to cells.

On the other hand, humanity does seem to be learning and gaining power much *faster* than natural selection. We have not built a mouse, but we have also not spent anything like the hundreds of millions of years that natural selection takes to build mice from scratch. DNA's

structure was discovered less than a century ago (1953). Our current airplanes can't reproduce in the wild, but two hundred years ago they weren't flying at all.

[break]

Our airplanes can also carry 20 tons while crossing the Pacific ocean in 10 hours, and that is something an eagle or an albatross is legitimately not close to doing. It's genuinely not clear biological evolution ever would or could get that kind of speed and carrying capacity, even if some environmental challenge selected for faster, bigger birds over a hundred million years; there are basic obstacles to doing that sort of thing with proteins (the basic building blocks of biology) rather than, say, metals. And natural selection is not a sort of thing that can just start over with a new non-protein system; natural selection works by blindly tweaking previous systems. Proteins are literally as old as life on Earth, and there is no known life that does not run on proteins; not because proteins are optimal, but because natural selection is not a sort of thing that can start over and envision a completely different system.

Pigeons can reproduce, and heal their injuries, and the airplanes of humanity cannot. But humanity somehow got to airplanes through a *faster* route than natural selection building the pigeon, and we point at those things which let us work faster. We used sketches and designs and computer simulations and small-scale models, before producing an Airbus A320 from raw materials. We built Airbus A320s in our extended imagination (counting written blueprints as a form of augmented imagination) before building them in reality; without fully building and testing and discarding literal millions of intermediate forms, the way natural selection would have needed to solve that problem if it could solve that problem at all.

We gesture at the Airbus A320 that flies faster than a pigeon, and was *designed* much *much* faster than pigeons evolved. We say that this has to do with humanity making things in a different way than evolution does: quicker to learn; more able to leap through large gaps in the design space, by making many simultaneous changes that depend on each other; able to start over from scratch with new materials; and at least sometimes able to correct errors by visualizing designs in our imaginations, without needing to watch an obviously flawed plan fail. This is not meant as a definition of "intelligence", but it has something to do with the subject matter we want to talk about.

[break]

This is a somewhat different notion of "intelligence" than you might pick up from Hollywood movies. In Hollywood, an intelligent character is one who speaks in a British accent, and is shown to win in chess games, and occasionally presents Mr. Bond with some incredible technological marvel (if they're a good guy) or makes planning mistakes that a 5-year-old could spot (if they're bad).

And in common parlance, "intelligence" means a trait that is possessed by mathematicians and physicists (and especially Einstein), and dispossessed by the village idiot, with salespeople and musicians somewhere in the middle.

"Mathematicians are intelligent", one might say, "but intelligence isn't all there is to success--many of the most successful humans are charismatic businessmen, or successful pop stars." But charisma is not synthesized in the kidneys! Charisma is a mental process; it is a way that your brain moves around your muscles that leads to desirable outcomes; in our attempts to gesture at "intelligence", we gesture also at charisma. Charisma may not feel like "thinking hard", it may not even feel somatically like it's happening inside your head, but charisma still isn't synthesized inside your kidneys.

English does not have a standard word which uniquely means only this concept we're trying to point at, for the thing which humans have more of than mice and have differently from natural selection, for the thing that includes both engineering and charisma as ways that a brain directs a body to solve a complicated challenge from reality.

By rights, we should perhaps coin a new word, for our notion, something like "mindpower", so that when we refer to machines that are adept at “mindpower”, we are free from any connotation that they're adept at chess, but clueless when it comes to romance.

[break]

Alas, the word "intelligence" is the best that English has. Worse, it appears in the very widespread term "artificial intelligence" that's usually used to refer to a related subject matter. We ask you to bear with us, as we try to gesture at some thing that seems to exist in the world and doesn't quite have a *standard* word in English which means only and exactly that thing. And we beg you to understand, that not everything that everyone uses the word "intelligence" to mean, is the thing we're trying to point to.

Imagine if some culture used the word "fire" in a way where it referred to both the orangey-bright hot stuff and also a burning sense of passion, and maybe also, had a different word for fires that burn naturally versus those that people start or tend. An aspiring physicist needs to be able to say, "I want to start by talking about only the orangey-bright hot stuff -- all the orangey-bright hot stuff, whether it's built by people or found in nature" -- and not be shouted down by people saying "But you said 'fire', and when *I* use that word, I mean..." In technical textbooks there exists a convention where the author can say, "I am temporarily, locally using this word to mean only the following thing..." and this is allowed so long as they use the word *consistently* and without trying to fluctuate between nonstandard and standard meanings. We know this isn't a textbook, but we're doing that to 'intelligence' now. We don't know how the conversation can move forward otherwise. English doesn't, actually, have a word that everyone already uses to mean only and exactly the thing we want to point to.

[break]

Having tried to point at the thing-in-the-world, we'll tentatively venture one piece of a definition of intelligence in words.

In this definition, we'll talk about the work that intelligence does, rather than the mechanisms that perform the work.

If you ask a wise physicist to define an "engine", they will tell you about what sort of *work an engine does*, and they may tell you some principles that govern that work. They might say: "an engine converts non-mechanical energy into mechanical energy", and tell you about conservation of energy and how knowing that helps.

They won't try to come up with some verbal definition that simultaneously describes the inner machinery of both a combustion engine and a hamster wheel. The innards of a rocket engine differ from the innards of an electric motor differ from the innards of a nuclear reactor, and there's not much that can usefully be said of all those innards at once, except that they perform the job of converting energy into mechanical energy, subject to various physical laws governing the conversion.

[break]

On our viewpoint, we'd say there's two kinds of work involved in intelligence; importantly different even though they are interrelated, intertwined, and somewhat interconvertible. There's even mathematical support for this perspective, and current machine learning systems partially reflect these unspoken mathematical reasons. We mention this, not because we're going to throw the math at you right away, but because if you're wondering why we picked these two kinds of work and not twenty other different plausible-sounding kinds of work, the answer is that this is what the math supports. More on this later.

The first kind of work we'd say is done by an engine of intelligence is "guessing what will next be seen, before you see it" -- not just, say, predicting the outcome of a sports game that you want to bet on; but also the wordless anticipation when you look out the window that your eyes will see a blue sky or a gray sky, rather than flashing orange and green.

The second kind of work is "moving or choosing in a way that brings you to a particular destination or outcome" -- not just choosing the path you take to drive across a city to get to a particular grocery, say, but also how your brain sends motor signals to your muscles to keep you upright.

We'll call them "prediction" and "steering".