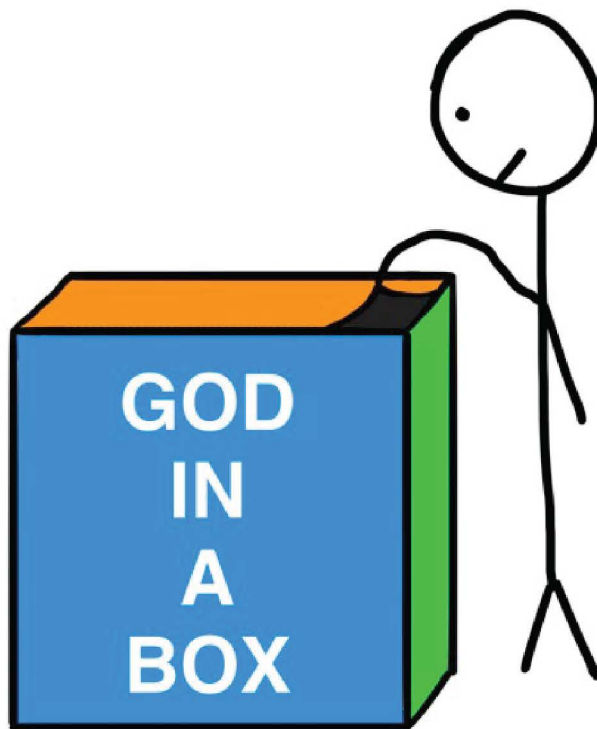




WAIT BUT WHY

The AI Revolution: Our Immortality or Extinction



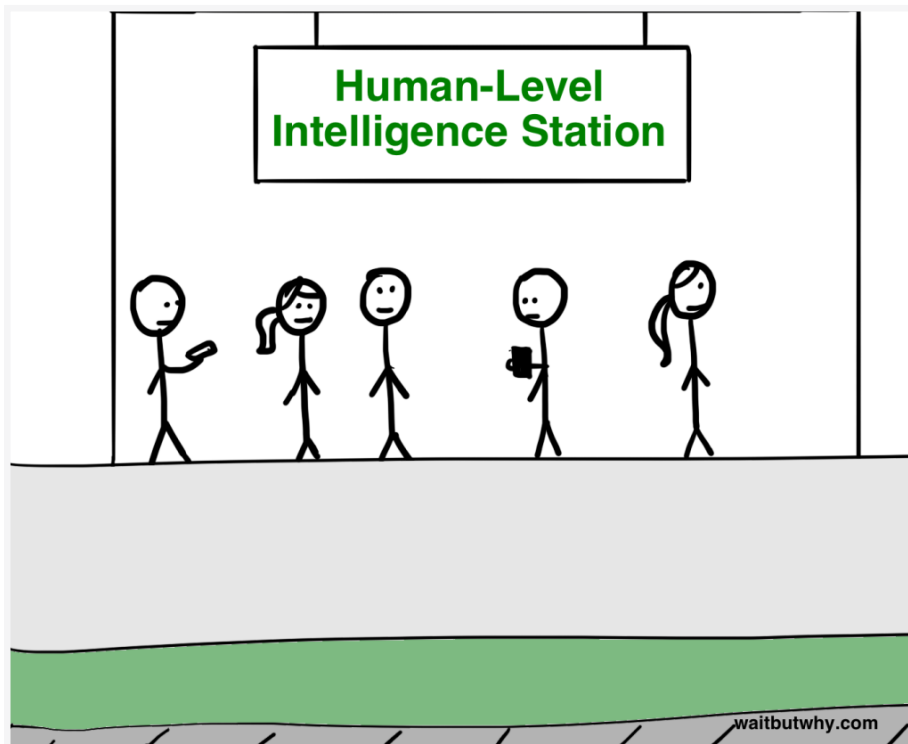
By Tim Urban

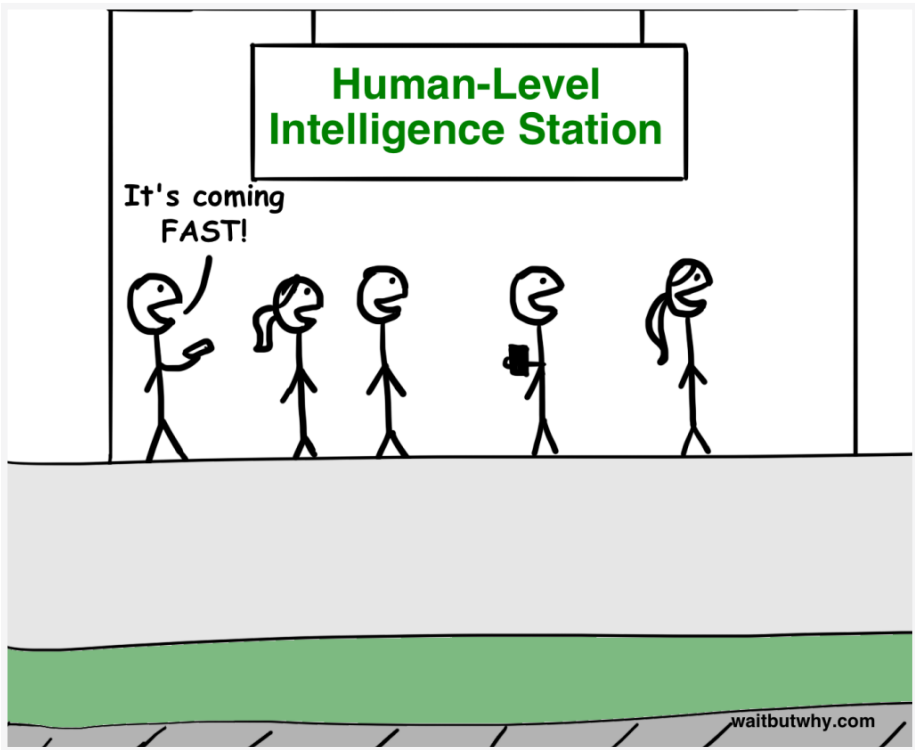
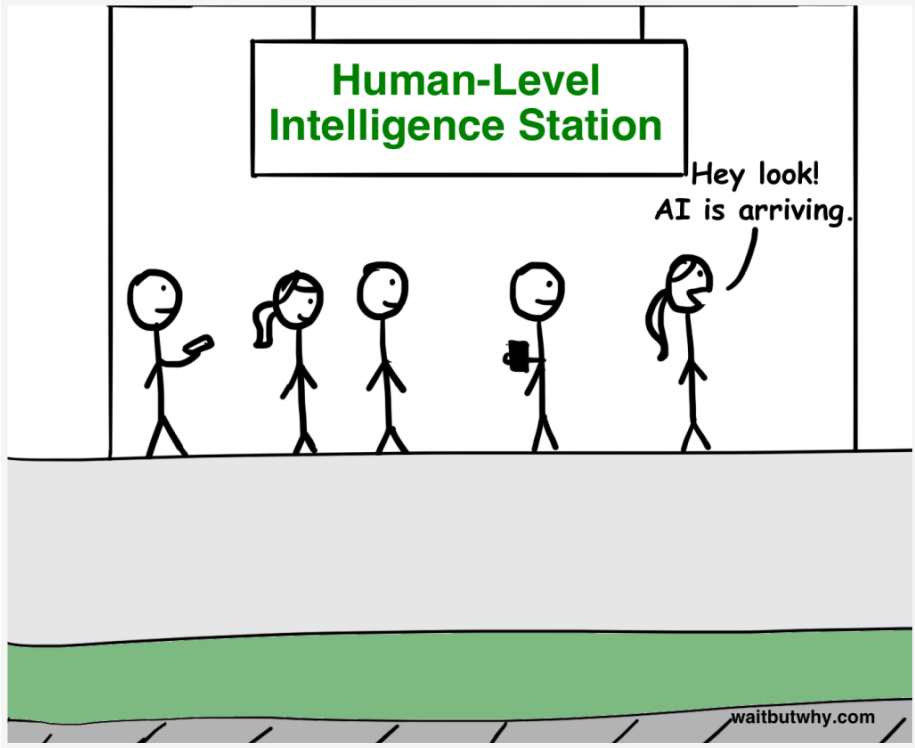
Note: This is Part 2 of a two-part series on AI. Part 1 is [here](#).

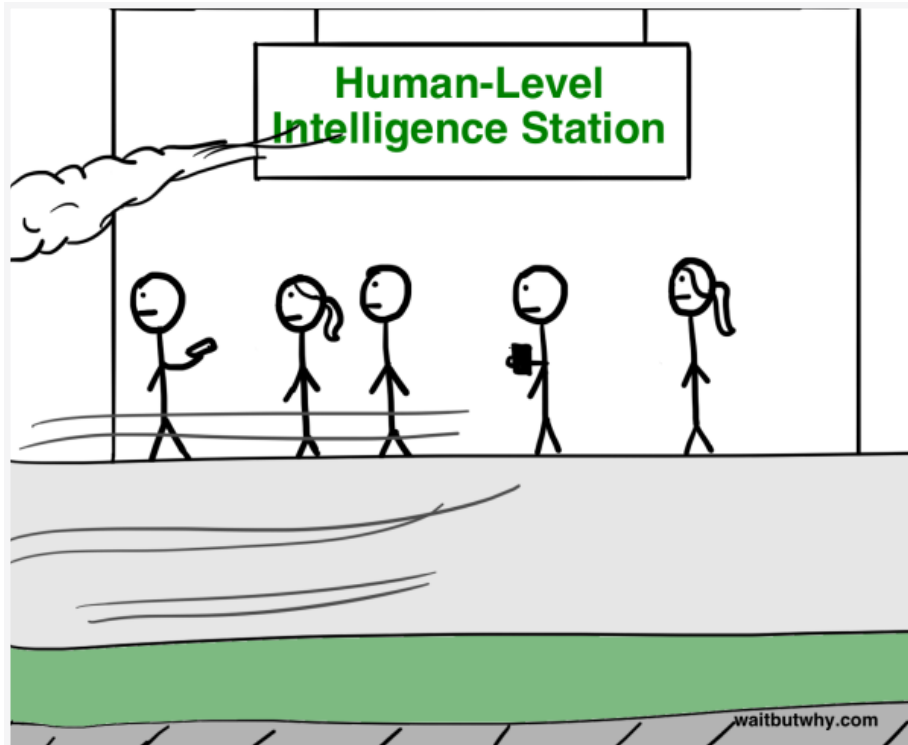
We have what may be an extremely difficult problem with an unknown time to solve it, on which quite possibly the entire future of humanity depends. — Nick Bostrom

Welcome to Part 2 of the “Wait how is this possibly what I’m reading I don’t get why everyone isn’t talking about this” series.

Part 1 started innocently enough, as we discussed Artificial Narrow Intelligence, or ANI (AI that specializes in one narrow task like coming up with driving routes or playing chess), and how it’s all around us in the world today. We then examined why it was such a huge challenge to get from ANI to Artificial General Intelligence, or AGI (AI that’s at least as intellectually capable as a human, across the board), and we discussed why the exponential rate of technological advancement we’ve seen in the past suggests that AGI might not be as far away as it seems. Part 1 ended with me assaulting you with the fact that once our machines reach human-level intelligence, they might immediately do this:







This left us staring at the screen, confronting the intense concept of potentially-in-our-lifetime Artificial Superintelligence, or ASI (AI that's way smarter than any human, across the board), and trying to figure out which emotion we were supposed to have on as we thought about that. ¹

Before we dive into things, let's remind ourselves what it would mean for a machine to be superintelligent.

A key distinction is the difference between *speed superintelligence* and *quality superintelligence*. Often, someone's first thought when they imagine a super-smart computer is one that's as intelligent as a human but can think much, much *faster* ²—they might picture a machine that thinks like a human, except a million times quicker, which means it could figure out in five minutes what would take a human a decade.

That sounds impressive, and ASI *would* think much faster than any human could—but the true separator would be its advantage in intelligence *quality*, which is something completely different. What makes humans so much more intellectually capable than chimps isn't a difference in thinking speed—it's that human brains contain a number of sophisticated cognitive modules that enable things like complex linguistic representations or longterm planning or abstract reasoning, that chimps' brains do not. Speeding up a chimp's

¹ Tiny orange footnotes are boring and when you read one, you'll end up bored. These are for sources and citations only.

1

If you don't know the deal with the notes, there are two different types. The blue circles are the fun/interesting ones you should read. They're for extra info or thoughts that I didn't want to put in the main text because either it's just tangential thoughts on something or because I want to say something a notch too weird to just be there in the normal text.

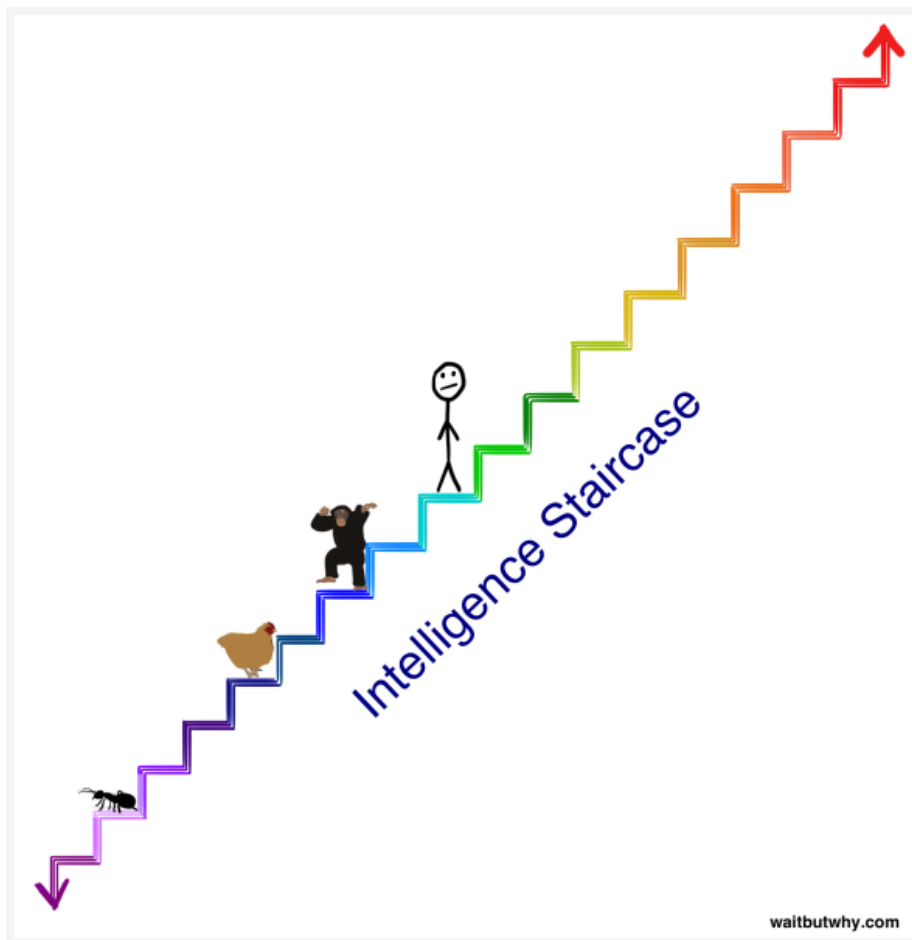
2

The movie *Her* made speed the most prominent superiority of the AI character over humans.

brain by thousands of times wouldn't bring him to our level—even with a decade's time, he wouldn't be able to figure out how to use a set of custom tools to assemble an intricate model, something a human could knock out in a few hours. There are worlds of human cognitive function a chimp will simply never be capable of, no matter how much time he spends trying.

But it's not just that a chimp can't do what we do, it's that his brain is unable to grasp that those worlds even *exist*—a chimp can become familiar with what a human is and what a skyscraper is, but he'll never be able to understand that the skyscraper was *built by* humans. In his world, anything that huge is part of nature, period, and not only is it beyond him to build a skyscraper, it's beyond him to *realize that anyone can build a skyscraper*. That's the result of a small difference in intelligence quality.

And in the scheme of the intelligence range we're talking about today, or even the much smaller range among biological creatures, the chimp-to-human quality intelligence gap is *tiny*. In an earlier [post](#), I depicted the range of biological cognitive capacity using a staircase: ³



3

A) The location of those animals on the staircase isn't based on any numerical scientific data, just a general ballpark to get the concept across. B) I'm pretty proud of those animal drawings.

To absorb how big a deal a superintelligent machine would be, imagine one on the dark green step two steps above humans on that staircase. This machine would be only *slightly* superintelligent, but its increased cognitive ability over us would be as vast as the chimp-human gap we just described. And like the chimp's incapacity to ever absorb that skyscrapers can be built, we will never be able to even comprehend the things a machine on the dark green step can do, even if the machine tried to explain it to us—let alone do it ourselves. And that's only two steps above us. A machine on the second-to-highest step on that staircase would be to us as we are to ants—it could try for years to teach us the simplest inkling of what it knows and the endeavor would be hopeless.

But the kind of superintelligence we're talking about today is something far beyond anything on this staircase. In an intelligence explosion—where the smarter a machine gets, the quicker it's able to increase its own intelligence, until it begins to *soar* upwards—a machine might take years to rise from the chimp step to the one above it, but perhaps only hours to jump up a step once it's on the dark green step two above us, and by the time it's ten steps above us, it might be jumping up in four-step leaps every second that goes by. Which is why we need to realize that it's distinctly possible that very shortly after the big news story about the first machine reaching human-level AGI, we might be facing the reality of coexisting on the Earth with something that's here on the staircase (or maybe a million times higher):

