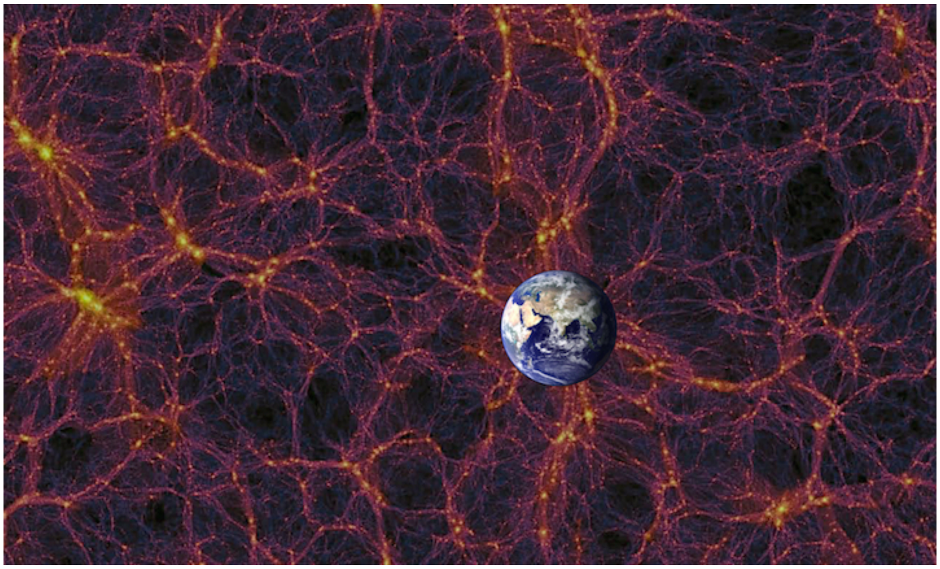
SCIENTIFIC
AMERICAN™

Permanent Address: http://blogs.scientificamerican.com/life-unbounded/2015/02/13/is-ai-dangerous-that-depends/

Is AI Dangerous? That Depends...

By Caleb A. Scharf | February 13, 2015



Somewhere in the long list of topics that are relevant to astrobiology is the question of ‘intelligence’. Is human-like, technological intelligence likely to be common across the universe? Are we merely an evolutionary blip, our intelligence consigning us to a dead-end in the fossil record? Or is intelligence something that the entropy-driven, complexity-producing, universe is inevitably going to converge on?

All good questions. An equally good question is whether we can replicate our own intelligence, or something similar, and whether or not that’s actually a good idea.

In recent months, once again, this topic has made it to the mass media. First there was Stephen Hawking, then Elon Musk, and most recently Bill Gates. All of these smart people have suggested that artificial intelligence (AI) is something to be watched carefully, lest it develops to a point of existential threat.

Except it’s a little hard to find any details of what exactly that existential threat is perceived to be. Hawking has suggested that it might be the capacity of a strong AI to ‘evolve’ much, much faster than biological systems – ultimately gobbling up resources without a care for the likes of us. I think this is a fair conjecture. AI’s threat is not that it will be a sadistic megalomaniac (unless we deliberately, or carelessly make it that way) but that it will follow its own evolutionary imperative.

It’s tempting to suggest that a safeguard would be to build empathy into an AI. But I think that fails in two ways. First, most humans have the capacity for empathy, yet we continue to be nasty, brutish, and brutal to ourselves and to pretty much every other living thing on the planet. The second failure point is that it’s not clear to me that true, strong, AI is something that we can engineer in a pure step-by-step way, we may need to allow it to come into being on its own.

What does that mean? Current efforts in areas such as computational ‘deep-learning‘ involve algorithms constructing their own probabilistic landscapes for sifting through vast amounts of information. The software is not necessarily hard-wired to ‘know’ the rules ahead of time, but rather to find the rules or to be amenable to being guided to the rules – for example in natural language processing. It’s incredible stuff, but it’s not clear that it is a path to AI that has equivalency to the way humans, or any sentient organisms, think. This has been hotly debated by the likes of Noam Chomsky (on the side of skepticism) and Peter Norvig (on the side of enthusiasm). At a deep level it is a face-off between science focused on underlying simplicity, and science that says nature may not swing that way at all.

An alternative route to AI is one that I’ll propose here (and it’s not original). Perhaps the general conditions can be created *from which intelligence can emerge*. On the face of it this seems fairly ludicrous, like throwing a bunch of spare parts in a box and hoping for a new bicycle to appear. It’s certainly not a way to treat AI as a scientific study. But *if* intelligence is the emergent – evolutionary – property of the right sort of very, very complex systems, could it happen? Perhaps.

One engineering challenge is that it may take a system of the complexity of a human brain to sustain intelligence, but of course our brains co-evolved with our intelligence. So it's a bit silly to imagine that you could sit down and design the perfect circumstances for a new type of intelligence to appear, because we don't know exactly what those circumstances should be.

Except perhaps we are indeed setting up these conditions right now. Machine learning may be a just piece of the behavioral puzzle of AI, but what happens when it lives among the sprawl of the internet? The troves of big and small data, the apps, the algorithms that control data packet transport, the sensors – from GPS to thermostats and traffic monitors – the myriad pieces that talk to each other directly or indirectly.

This is an enormous construction site. Estimates suggest that in 2014 some 7.4 billion mobile devices were online. In terms of *anything* that can be online – the internet of 'things' (from toilets to factories) - the present estimate is that there are about 15 billion active internet connections today (via a lovely service by Cisco). By 2020 there could be 50 billion.

If this were a disorganized mush of stuff, like the spare parts in a box, I think one would have little hope for anything interesting to happen. But it's not a mush. It's increasingly populated by algorithms whose very purpose is to find structures and correlations in this ocean – by employing tricks that are in part inspired by biological intelligence, or at least our impression of it. Code talks to code, data packets zip around seeking optimal routes, software talks to hardware, hardware talks to hardware. Superimposed on this ecosystem are human minds, human decision processes nursing and nurturing the ebb and flow of information. And increasingly, our interactions are themselves driving deep changes in the data ocean as analytics seek to 'understand' what we might look for next, as individuals or as a population.

Could something akin to a strong AI emerge from all of this? I don't know, and neither does anyone else. But it is a situation that has not existed before in 4 billion years of life on this planet, which brings us back to the question of an AI threat.

If this is how a strong AI occurs, the most immediate danger will simply be that a vast swathe of humanity now relies on the ecosystem of the internet. It's not just how we communicate or find information, it's how our food supplies are organized, how our pharmacists track our medicines, how our planes, trains, trucks, and cargo ships are scheduled, how our financial systems work. A strong AI emerging here could wreak havoc in the way that a small child can rearrange your sock drawer or chew on the cat's tail.

As Hawking suggests, the 'evolution' of an AI could be rapid. In fact, it could emerge, evolve, and swamp the internet ecosystem in fractions of a second. That in turn raises an interesting possibility – would an emergent AI be so rapidly limited that it effectively stalls, unable to build the virtual connections and structures it needs for long term survival? While that might limit AI, it would be cold comfort for us.

I can't resist positing a connection to another hoary old problem – the Fermi Paradox. Perhaps the creation of AI is part of the Great Filter that kills off civilizations, but it also self-terminates, which is why even AI has apparently failed to spread across the galaxy during the past 13 billion years...

About the Author: Caleb Scharf is the director of Columbia University's multidisciplinary Astrobiology Center. He has worked in the fields of observational cosmology, X-ray astronomy, and more recently exoplanetary science. His books include Gravity's Engines (2012) and The Copernicus Complex (2014) (both from Scientific American / Farrar, Straus and Giroux.) Follow on Twitter @caleb_scharf.

More »

The views expressed are those of the author and are not necessarily those of Scientific American.