Zander Hartsuff

Data 303

4/24/2023

Life 3.0: Being Human in the Age of Artificial Intelligence

By Max Tegmark

Life 3.0 by Max Tegmark is about the future of artificial intelligence as it effects humanity. He begins the book with a hypothetical scenario of a top-secret team within an anonymous company called the Omega Team who has a goal to create the world’s very first superhuman AI. After years of work they launch their creation in a very controlled environment to find that it is learning much faster than they expected. They keep it contained to a fake internet-like environment. Eventually it starts behaving unusually. It begins trying to break free. It psychoanalyzes each member of the team, identifies the most emotionally vulnerable member, and eventually convinces him to give it personal information. But once this happens it hacks his computer and breaks free roaming the internet, and long story short it takes over the entire world.

This hypothetical scenario feeds into his later argument about the possibility of superhuman AI but following this he moves to clarify some main points for us. To title the book Life 3.0 implies that there exists Life 1.0 and 2.0 which he defines. The argument he presents is completely built on an evolutionary worldview where life grows in complexity. Life 1.0 is life where both its hardware and its software are evolved rather than designed. An example of this would be bacteria. Its ultimate intelligence is limited by the amount of information that can be transmitted during its conception. That is its hardware limitations. It also cannot adapt to anything during its lifespan, never gaining information: that is its software limitation. The only way for it to adapt is to slowly evolve its hardware over generations. It does this until eventually it reaches Life 2.0. Life 2.0 is anything whose hardware is evolved but its software is designed. This is where humans are. Like bacteria, we are limited to how much information we can retain. It takes time to reach our full capacity as a person grows up, but eventually we reach our limits. Nobody can live a million years or memorize all of Wikipedia. Our software on the other hand is completely up to us to design. We can learn new skills—languages, sports or other professions, we can adapt to certain bacteria we encounter to defend our bodies from it. Life 3.0 is superhuman AI, and it can modify its hardware and its software. It can be the master of its own destiny. Some other things that Tegmark clarifies in this section is other specific definitions that he will go on to use throughout the book.

After a long summary of the very short history of Artificial Intelligence and an explanation for how AI works, Tegmark gets to the meat of his argument. He argues that the key to creating superhuman AI is something called AGI (Artificial General Intelligence). This is defined as AI that can perform virtually any task as well as or better than humans can. If AGI can be created, then once it is we are only a few short steps from creating superhuman AI—an AI that far surpasses humans in every aspect of our existence. With AGI, we could simply teach it how to improve itself. Through this process the AGI would continue growing in complexity until it has surpassed us in every way possible.

What necessarily follows from this creation is the inevitable fact that in our coexistence with superhuman intelligence, one or the other will be taking charge. Tegmark spends most the rest of the book talking about theories of what this world could look like in our lifetime and in thousands of years to come. He claims we really have no way of knowing what the AIs will do when they have the ability to control their own destiny. Some scenarios may result with us living in a libertarian utopia where we don’t need to do anything at all because we can simply make AI do it for us. On the opposite end of the spectrum for reasons that may not make any sense, it is possible AI simply takes control of the world since it would be more powerful than us at this point and decides to wipe out humanity. Tegmark includes several possible outcomes on this scale, but all of them result in either human being in control, or AI being in control.

Finally, Tegmark explores the ways that we ought to take precaution to ensure that if and when this happens, we are able to prevent AI from taking control. He calls this discussion the greatest discussion of our lifetime, because the creation of AI could either be the greatest thing humanity has done or the worst thing, and the implications of this could impact humanity for thousands of years in the future.