Max Tegmark

# Life 3.0: Being Human in the Age of Artificial Intelligence

Made by Blinkist

These key insights in blinks were written by a team of experts at Blinkist. We screen the world of nonfiction to choose the very best books. Then, we read them deeply and transform them into this concise format that brings you the most inspiring ideas from the text.

Maybe these blinks will inspire you to dig deeper, or maybe they're enough to start you thinking and then on to something new. However you read blinks, we hope they help you become an even brighter you.

## What's in it for me? Learn what the future will bring.

For thousands of years, life on earth has been progressing and evolving. No species exemplifies this more than humans.

Max Tegmark imagines us now moving toward the final evolutionary stage: *Life 3.0*. In this era of humanity, technology will live independently, designing both its own hardware and software, and the repercussions for the very existence of humankind are immense.

Such artificial life does not yet exist on earth. However, we are faced with the emergence of non-biological intelligence, commonly known as artificial intelligence (AI).

In these blinks, you'll be taken on a journey charting possible versions of the future. You'll also learn what exactly is

involved in the creation of AI and how AI differs from human intelligence. Along the way, you'll grapple with some of the biggest philosophical questions concerning what it means to be human.

In these blinks, you'll learn:

- about the holy grail of AI research;
- what kind of chaos exists in your coffee cup; and
- how AI might put jobs at risk.

## AI could represent the future of life, but it's a controversial subject.

The story of how life emerged on earth is well known. Some 13.8 billion years ago, the Big Bang brought our universe into being. Then, about four billion years ago, atoms on earth arranged themselves in such a way that they could maintain and replicate themselves. Life had arisen.

As the author posits, life can be classified into three categories according to levels of sophistication.

The first stage of life, *Life 1.0*, is simply *biological*.

Consider a bacterium. Every aspect of its behavior is coded into its DNA. It's impossible for it to learn or change its behavior over its lifetime. The closest it comes to learning or improvement is

evolution, but that takes many generations.

The second stage is *cultural*, *Life 2.0*.

Humans are included here. Just like the bacterium, our "hardware" or bodies have evolved. But unlike simpler life-forms, we can acquire new knowledge during our lifetimes. Take learning a language. We can adapt and redesign ideas that we might call our "software." And we make decisions using this knowledge.

The final stage is the theoretical *Life 3.0*, a form of technological life capable of designing both its hardware and software. Although such life doesn't yet exist on earth, the emergence of non-biological intelligence in the form of AI technologies may soon change this.

Those who hold opinions about AI can be classified by how they feel about the emerging field's effect on humanity.

First up are the *digital utopians*. They believe that artificial life is a natural and desirable next step in evolution.

Second, there are the *techno-skeptics*. As the name suggests, they don't believe that artificial life will have an impact anytime soon.

Finally, there's the *beneficial AI movement*. These people aren't sold on the idea that AI will necessarily bring benefits to humans. They therefore advocate that AI research be specifically directed toward possible universally positive outcomes.

*“One of the most spectacular developments during the 13.8 billion years since our Big Bang is that dumb and lifeless matter has turned intelligent.”*

**Capabilities for memory, computation, learning and intelligence aren't distinctly human attributes.**

What makes us human? Our ability to think and learn? One might think so.

Researchers in AI, however, are generally opposed to such a notion. They claim that the capability for memory, computation, learning and intelligence has nothing to do with human flesh and blood, let alone carbon atoms.

Let's begin with intelligence. Though there's no universally accepted single definition, the author likes to think of intelligence as the "*ability to accomplish complex goals.*"

Machines might be increasingly able to outperform us in defined tasks such as playing chess, but human intelligence is uniquely broad. It can encompass skills

like language learning and driving vehicles.

However, even though *artificial general intelligence* (AGI) doesn't yet exist, it's clear that intelligence isn't just a biological faculty. Machines can complete complex tasks too.

Intelligence – just like capacities for memory, computation and learning – is what's known as *substrate independent.* That is, an independent layer that does not reflect or depend upon an underlying material substrate.

So, for example, human brains can store information, but so can floppy drives, CDs, hard drives, SSDs and flash memory cards, even though they're not made of the same material.

But before we get to what this means for computing, we need to understand what computing is.

Computing involves the transformation of information. So, the word “hello” might be transformed into a sequence of zeros and ones.

But the rule or pattern which determines this transformation is independent of the hardware that performs it. What’s important is the rule or pattern itself.

This means that it’s not only humans who can learn – the same rules and patterns could exist outside of the human brain too. AI researchers have made huge strides developing *machine learning*: machines that can improve their own software.

So, if memory, learning, computation and intelligence aren’t distinctly human, then what exactly makes us human? As research in AI continues apace, this question is only going to prove harder to answer.

## AI is advancing rapidly and will impact human life in the near future.

Machines aren't anything new to humans. We've been using them for manual tasks for millennia. If you define your self-worth by your cognitive skills, such as intelligence, language and creativity, these machines pose no threat. However, recent breakthroughs in AI might begin to worry you.

The author had his own "holy-shit" moment in 2014 when he witnessed an AI system playing an old computer game named Breakout. That's the game where you hit a ball against a wall by maneuvering a paddle.

At first, the AI system did poorly. But it soon learned and eventually developed an intelligent score-maximizing strategy that even the developers hadn't thought of when they played themselves.

It happened again in March 2016, when the AI system AlphaGo beat Lee Sedol, the world's best Go player. Go is a strategy game that requires intuition and creativity because there are many more possible positions in the game then there are atoms in the universe, so mere brute force analysis is not practical. But the AI system still sailed to victory, seeming to display exactly the kind of creativity required.

AI systems are also advancing quickly in the field of natural languages. Just consider how much the quality of translations provided by Google Translate has improved lately.

It's clear that AI will impact all areas of human life in the near future. Algorithmic trading will affect finance; autonomous driving will make transportation safer, smart grids will optimize energy distribution and AI doctors will change healthcare.

The big issue to consider is the effect AI will have on the job market. After all, as AI systems can outperform humans in more and more fields, we humans may even become unemployable.

Let's turn now to other potential impacts of AI development.

## Creating human-level AI could result in a superintelligent machine taking over the world.

Until now, AI has been applied fairly narrowly in limited fields like language translation or strategy games.

In contrast, the holy grail of AI research is the production of AGI that would operate at a human level of intelligence.

But what would happen if this holy grail were found?

For starters, the creation of AGI might result in what's known to AI researchers as an *intelligence explosion*.

An intelligence explosion is a process by which an intelligent machine gains *superintelligence*, a level of intelligence far above human capability.

It would achieve this through rapid learning and recursive self-improvement because an AGI could potentially design an even more intelligent machine, which could design an even more intelligent machine and so on. This could trigger an *intelligence explosion* that would allow machines to surpass human intelligence.

What's more, superintelligent machines could take over the world and cause us harm, no matter how good our intentions.

Let's say, for example, that humans program a superintelligence that is concerned with the welfare of humankind. From the superintelligence's perspective, this would probably be akin to a bunch of kindergartners far beneath your intelligence holding you in bondage for their own benefit.

Quite probably you would find this a depressing and inefficient situation and take matters into your own hands. And

what do you do with incompetent, annoying human obstacles? Control them, or better yet, destroy them.

But maybe we're getting ahead of ourselves; let's look at some other, less terrifying, scenarios that might occur.

**Various AI aftermath scenarios are possible, ranging from the comforting to the terrifying.**

Whether we like it or not, the race toward AGI is underway.

But what would we like the aftermath of attaining it to look like?

For instance, should AIs be conscious? Should humans or machines be in control?

We have to answer basic questions, as we don't want to end up in an AI future for which we're unprepared, especially one which could do us harm.

There are various *aftermath* scenarios. These vary from peaceful human–AI coexistence to AIs taking over, leading to human extinction or imprisonment.

The first possible scenario is the *benevolent dictator*. A single benevolent superintelligence would rule the world, maximizing human happiness. Poverty, disease and other low-tech nuisances would be eradicated, and humans would be free to lead a life of luxury and leisure.

In the same vein, there's a scenario involving a *protector god*, where humans would still be in charge of their own fate, but there would be an AI protecting us and caring for us, rather like a nanny.

Another scenario is the *libertarian utopia*. Humans and machines would peacefully coexist. This would be achieved through clearly defined territorial separation. Earth would be divided into three zones. One would be devoid of biological life but full of AI. Another would be human only. There would be a final mixed zone, where humans could become cyborgs by upgrading their bodies with machines.

This scenario is a little fantastical, however, as there's nothing to stop AI machines disregarding humans' wishes.

Then there's the c*onquerors' scenario*, which we looked at in the last blink. This would see AIs destroy humankind, as we'd be seen as a threat, a nuisance or simply a waste of resources.

Finally, there's the *zookeeper scenario*. Here a few humans would be left in zoos for the AIs' own entertainment, much like we keep endangered panda bears in zoos.

Now that we've examined possible AI-related futures, let's look at the two largest obstacles to current AI research, namely goal-orientedness and consciousness.

*“Humans will become as irrelevant as cockroaches.”* –
Marshall Brain

**Nature, humans included, has goals, and researchers are striving to simulate this behavior for AI.**

There's no doubt that we humans are goal-oriented. Think about it: even something as small as successfully pouring coffee into a cup involves completing a goal.

But actually, nature operates the same way. Specifically, it has one ultimate purpose: destruction. Technically, this is known as *maximizing entropy*, which in a layperson's terms means increasing messiness and disorder. When entropy is high, nature is "satisfied."

Let's return to the cup of coffee. Pour a little milk in, then wait a short while. What do you see? Thanks to nature, you now have a lukewarm, light brown, uniform mixture. Compared to the initial situation, where two liquids of different temperatures were clearly separate, this

new arrangement of particles is indicative of less organization and increased entropy.

On a bigger scale, the universe is no different. Particle arrangements tend to move toward increased levels of entropy, resulting in stars collapsing and the expansion of the universe.

This goes to show how crucial goals are, and currently, AI scientists are grappling with the problem of which goals AI should be set to pursue.

After all, today's machines have goals too. Or rather, they can exhibit goal-oriented behavior. For instance, if a heat-seeking missile is hot on your tail, it's displaying goal-oriented behavior.

But should intelligent machines have goals at all? And if so, who should define those goals? For instance, Marx and Hayek each had a distinct vision when it came to the future of the economy and

society, so they would undoubtedly set very different goals for AI.

Of course, we could begin with something simple, like the Golden Rule that tells us to treat others as we would ourselves.

But even if humanity could agree on a few moral principles to guide an intelligent machine's goals, implementing human-friendly goals would be trickier yet.

First of all, we'd have to make an AI *learn* our goals. This is easier said than done because the AI could easily misunderstand us. For instance, if you told a self-driving car to get you to the airport as fast as possible, you might well arrive covered in vomit while being chased by the police. Technically, the AI adhered to your stated wish, but it didn't really understand your underlying motivation.

The next challenge would be for the AI to *adopt* our goals, meaning that it would agree to pursue them. Just think of some politicians you know: even though their goals may be clear, they still fail to convince large swaths of the population to adopt the same goals.

And finally, the AI would have to *retain* our goals, meaning that its goals wouldn't change as it undergoes self-improvement.

Huge amounts of scientific research are currently being devoted to just these ideas.

Even though nature's goal is maximum entropy, particles nonetheless rearrange themselves into complex organisms. Why? Because a living organism dissipates energy faster and thereby increases entropy.

## AI researchers are deliberating the meaning of consciousness and the subjectiveness of AI experience.

The question of what consciousness is and how it relates to life is hardly new. AI researchers are faced with the same age-old issue. More specifically, they wonder how lifeless matter could become conscious.

Let's come at it from a human perspective first. As a physicist like the author would put it, conscious human beings are just "food rearranged," meaning that the atoms we ingest are simply rearranged to form our bodies.

Consequently, what interests AI researchers, then, is the rearrangement that intelligent machines would have to undergo to become conscious.

It shouldn't be a surprise that no one has an answer right now. But to get closer,

we have to grasp what's involved in consciousness.

It's tricky. We might like to imagine consciousness is something to do with awareness and human brain processes. But then we're not actively aware of every brain process. For example, you're typically not consciously aware of everything in your field of vision. It's not clear why there's a hierarchy of awareness and why one type of information is more important than another.

Consequently, multiple definitions of consciousness exist. But the author favors a broad definition known as *subjective experience*, which allows a potential AI consciousness to be included in the mix.

Using this definition, researchers can investigate the notion of consciousness through several sub-questions. For instance, "How does the brain process

information?" or "What physical properties distinguish conscious systems from unconscious ones?"

AI researchers have also deliberated how artificial consciousness or the subjective AI experience might "feel."

It's posited that the subjective AI experience could be richer than human experience. Intelligent machines could be purposed with a broader spectrum of sensors, making their sensory experience far fuller than our own.

Additionally, AI systems could experience more per second because an AI "brain" would run on electromagnetic signals traveling at the speed of light, whereas neural signals in the human brain travel at much slower speeds.

It might seem like a lot to wrap your head around, but one thing is clear: the potential impact of AI research is vast. It points to the future, but it also entails

facing some of humankind's oldest philosophical questions.

## Final summary

The key message in this book:

**The race for human-level AI is in full swing. It's not a question of *if* AGI will arrive, but when. We don't know what exactly will happen when it does, but several scenarios are possible: humans might upgrade their "hardware" and merge with machines, or a superintelligence may take over the world. One thing is certain – humanity will have to ask itself some deep philosophical questions about what it means to be human.**

## Got feedback?

We'd sure love to hear what you think about our content! Just drop an email to remember@blinkist.com with the title of this book as the subject line and share your thoughts!

## Suggested further reading: *The Age of Spiritual Machines* by Ray Kurzweil

*The Age of Spiritual Machines* (1999) is your guide to the future. These blinks explain the new age of machines and what robotic intelligence will mean for life as we know it.

**Nice work! You're all done with this one.**

We publish new books every week at blinkist.com.
Come and see – there's so much more to learn.

Inspired to read the full book?

Get it here.