Max Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence*, New York: Random House–Knopf, 2017, 543 pp (eBook).

Max Tegmark is a Swedish-American physicist and an MIT professor who loves thinking about life's big questions, from cosmology to artificial intelligence (AI), and has wrote the book *Life 3.0: Being Human in the Age of Artificial Intelligence* published by *Random House* on August 29, 2017., in which he explores the future of life, technology and AI, and its relationship to human beings.

If you are wondering about the title of the book, let me take a minute to explain it to you. Tegmark divides life into three stages: *Life 1.0, Life 2.0* and *Life 3.0. Life 1.0* represents the simple biological evolution, where both the hardware and software are evolved rather than designed. The hardware and software are determined by its DNA and can't be changed in a single organism's lifetime but can gradually evolve over many generations. *Life 1.0* can only survive and replicate. Examples of *Life 1.0* are bacteria and other single-celled organisms. Humans, on the other hand, are examples of *Life 2.0,* life whose hardware is evolved, but whose software is largely designed. By hardware Tegmark means the body, and by software, our concepts, ideas and extended abilities such as language; all the algorithms and knowledge from our senses, thoughts and emotions that we use to process the information and then decide what to do — "everything from the ability to recognize your friends when you see them to your ability to walk, read, write, calculate, sing and tell jokes" (2017: 42). *Life 2.0* represents the cultural revolution, except it can survive and replicate, it can also redesign much of its software by studying, thinking, writing, joking or inventing new technologies. Finally, *Life 3.0* is the AI which can design and upgrade both its software and hardware. It can replicate itself from scratch and build new bodies relatively quickly from raw materials, plus, it can also learn from its surroundings, gather and store information, learn from the past to avoid mistakes, enabling it to advance enormously and to transform itself more directly and quickly than our creativity enables us to do. In this sense, *Life 3.0* is the master of its own destiny and, when created, will represent the technological evolution of life.

Tegmark says that the boundaries between the three stages of life are not so rigid. For example, if bacteria are *Life 1.0* and humans are *Life 2.0*, then a mouse could be between 1.0 and 2.0, let's classify it as *Life 1.1*. Although we might think that its software is also designed because of its ability of learning, the fact is that it can't develop a proper language to communicate or some other methods that can efficiently help him transfer the gathered knowledge to next generations. What a mouse learns during its life largely gets lost when he dies and the newborn has to learn from scratch by watching the elders. Similarly, we might argue that humans living in the 21st century are between 2.0 and 3.0. In fact, considering the current progress with prosthetics, artificial limbs and medicine in general, we are more like *Life 2.1* already. We can perform some hardware changes like replacing a missing body part, installing a pacemaker or a hearing aid, strengthening our *immune* system and curing many diseases with medications and so on,

but we cannot design our body to be immensely different like having 4 arms, being 10 meters tall or having a thousand times bigger brain (45–46).

In summary he divides the development of life into three stages, distinguished by life's ability to design itself:

- *Life 1.0* (biological stage): evolves its hardware and software
- *Life 2.0* (cultural stage): evolves its hardware, designs much of its software
- *Life 3.0* (technological stage): designs its hardware and software.

Even if *Life 3.0* doesn't yet exist on Earth, it can arrive during this century and perhaps even during our lifetime. What will then happen and what this progress in AI will mean for us humans is the topic of this book, says Tegmark (46).

Now about the book. *Life 3.0: Being Human in the Age of Artificial Intelligence* is a relatively long book (almost 400 pages in printed edition), organized into eight chapters that take the reader on a journey that starts at the beginning of time, and describes the evolution of intelligence and technology in relation with humans. Here you can see the structure of the book given by the author (75–76):

		Short Chapter Title	*Topic*	*Status*
		Prelude: Tale of the Omega Team	Food for thought	Extremely Speculative
The history of intelligence	1	The Conversation	Key ideas, terminology	Not very speculative
	2	Matter Turns Intelligent	Fundamentals of intelligence	
	3	AI, Economics, Weapons & Law	Near future	
	4	Intelligence Explosion?	Superintelligence scenarios	Extremely Speculative
	5	Aftermath	Subsequent 10,000 years	
	6	Our Cosmic Endowment	Subsequent billions of years	
The history of meaning	7	Goals	History of goal-oriented behavior	Not very speculative
	8	Consciousness	Natural & artificial consciousness	Speculative
		Epilog: Tale of the FLI Team	What should we do?	Not very speculative

As the author points out, it's possible to skip some chapters because they are mostly self contained. For example, if you are not new to AI, skipping chapter 2 will get you right to the question "What does it mean to be human in the age of AI?" On the other hand, readers new to the AI field will get a nice introduction and terminology of the field in chapters 1 and 2 (74).

Tegmark begins his book with a fictional "what-if" premise, a very plausible but imaginative tale of the Omega Team, a corporate team of brilliant researchers who, with a commitment to helping humanity, secretly build an AI called *Prometheus*. With strong security measures, this superintelligent machine not only makes billions for its creators, but takes over the world and transforms it positively. Using this AI technology, the Omega Team accomplishes the most dramatic transition in history; eliminating all previous national power structures they create a world alliance and consolidate a single global power which runs the planet, ending state conflict; improving the quality of life, education and health; increasing the entire planet's standard of living and enabling life to flourish into the far future throughout the cosmos. This prelude can be read as a SF thriller but in fact it's much better than what Hollywood has come up with so far on this subject, and I believe it works very well as an introduction to the book because it's presented

more like a non-fictional description of a business and political development rather than a sci-fi scenario. As the book progresses, the author occasionally includes fictional scenarios that fit the description of the Omega Team and Prometheus. This is genuinely thought-provoking and brings the technology and its human implications vividly to life. A great way to start the book and attract the attention of the reader (10–34).

After introducing the "Most Important Conversation of Our Time", giving the terminology and clearing some common misconceptions, in the second chapter of the book Tegmark gives a detailed description of intelligence, memory, computation and learning. He then discusses these qualities in the context of whether they are limited to humans (*Life 2.0*) or applicable to machines (*Life 3.0*) as well. The author gives an overview of intelligence from its origin, billions of years ago and through its development to the present days. He then explores the current state of research into machine learning and some breakthroughs in the field of AI.

In the third chapter, Tegmark goes on and discusses some of the main issues regarding AI and its impact on humanity in the near future. He considers the short-term effects of the development of AI such as space exploration, laws, AI weapons, jobs and wages, and the quest for human-level intelligence or AGI (Artificial General Intelligence). He often cites examples like *IBM Deep Blue*, *IBM Watson*, *Google DeepMind* (computer programs that can beat humans in chess, *Jeopardy* and *Go*) as well as self-driving cars, financial software and computer games which, in my opinion, brings the topic closer to the reader. I found this chapter very interesting and practical because the reader can begin to understand why securing AI is not only critical for the near future, but also how an inadequate security of AI could lead to catastrophic consequences in a much farther period of time in terms of public safety, financial stability, transportation, energy, healthcare, space exploration, and so on.

After discussing the current progress and possible issues in AI, Tegmark takes the reader on a journey through the "intelligence explosion" that will happen if one day we succeed in building human-level AGI, referring to the scenario of Prometheus and the Omega Team overtaking the world.

The fifth chapter includes a broad range of very interesting possible scenarios and consequences that could occur between intelligent machines and humans in a more distant future (*The Next 10,000 Years*), all depending on how we design AI's path, and whether the superintelligence will stay on those paths or decide to take a path for itself. Tegmark describes both positive and negative relations and potential outcomes, from a peaceful coexistence of humans and machines to the enslavement of machines and even to the complete overtaking of machines and extinction of humanity. This chapter can be very interesting for philosophers because it deals with concepts of political philosophy such as libertarianism, totalitarianism, egalitarianism, Orwellianism, freedom, social structures, political power, property rights and so on, all in a relationship of integration in a society between humans and intelligent machines.

The sixth chapter is a speculation about life's future potential aided by technology and how could it flourish in the next billion years and beyond,

not only in our Solar System but in all the possible cosmos. Tegmark describes the various ways the superintelligence could develop, whether it would become a rogue, how humans would interact with it, and would it prevent the predicted end of our universe. It takes a physicist to imagine how far life could life progress if limited only by the laws of physics. This part of the book is pretty astonishing even if most of it can hardly be achievable due to various limitations and possible cosmic wipeout. There is a lot of futurology in Tegmark's book which can be a little bit frustrating, especially about the things that we are highly unlikely to be able to predict, though at least the author recognizes this and points it out.

The remaining chapters explore concepts in physics, goals, ethics, the subject of consciousness and meaning, and then investigate what society can do to help create a desirable future for humanity. Tegmark believes that, in the future, when we create intelligent machines we could consider them, in some sense, as our descendants; we would be very proud of what they can do, they would have our values, and would do all the great things that we couldn't do. Even if they choose to eliminate us, they will live on and continue the story of life in our part of the observable universe. But what if those machines are zombies without any consciousness? Then if we humans eventually go extinct there will be nobody experiencing anything. It's like our whole universe had died for all intents and purposes. Tegmark believes that it's not our universe giving meaning to us, but we as conscious beings are giving meaning to our universe. The meaning comes from our experience. If there's nobody experiencing anything, our whole cosmos just goes back to being a giant waste of space. For these various reasons is very important to understand what it is about information processing that gives rise to what we call consciousness. Tegmark discusses what consciousness could be, saying that "consciousness is the way information feels when being processed in certain ways" and speculates that it must be substrate-independent, similarly to remembering, computing and learning (474–475). Tegmark argues that the risks of AI come not from malevolence or conscious behavior intrinsically, but rather from the misalignment of the goals of AI with those of humans. In Tegmark's words, "the real risk with artificial general intelligence isn't malice but competence. A superintelligent AI will be extremely good at accomplishing its goals, and if those goals aren't aligned with ours, we're in trouble" (407). Still, there are a lot of questions that humans should try to answer before any superintelligence is created, says Tegmark. He finishes his book optimistically, describing his work at the Future of Life Institute he has founded, which aims to ensure that we develop not only technology, but also the wisdom required to use it beneficially.

There is no doubt that the progress of AI can become an issue that needs thinking, writing and discussing about, and I believe Tegmark did a great job with *Life 3.0*. The book probably needs to be read alongside Nick Bostrom's *Superintelligence: Paths, Dangers, Strategies* and other recent books in the field to get a full picture. My opinion is that it would be even more effective if it was a bit shorter. Some chapters are intensely exciting and informative, others, including *Our Cosmic Endowment: The Next Billion Years and Beyond,* are a little bit too long and pretentious. Some chapters feel like fillers, put there just to make the book thicker, they add little to no useful

information on AI. On the other hand, the long awaited chapters on the ethical questions and consciousness, which would have made the book more interesting for me, especially from the perspective of a philosopher, are just a scratch on the surface and do not delve into the depth of these issues.

Nevertheless, *Life 3.0: Being Human in the Age of Artificial Intelligence* by Max Tegmark is a great book that I'd recommend to anyone interested in the topic of AI, the long-term effects of future technology and its ramifications on all aspects of mankind. People working in the AI should definitely read this book so that they understand the broader concerns surrounding this area. I intend to read it again as the discussion on AI will get more and more interesting and gain importance over time.

Personally, this topic is very close to me. During the last couple of years I took a deep interest in AI research, consciousness and the possibility of creating human-level AGI and *Superintelligence*. Last year, while I was finishing my master thesis on the philosophical problems of AI, Max Tegmark's book was published and instantly hit the bestseller lists in September and October 2017. Too bad it wasn't published earlier because it would be of great help for me, it's the kind of book I was looking for at that time, simple, well-rounded and up to date with the recent events in AI.

I believe that *Life 3.0: Being Human in the Age of Artificial Intelligence* is a very accessible and highly readable book even for readers with no background knowledge in the field of AI. Due to Tegmark's simple style of writing and avoiding fancy words, he successfully gave clarity to the many faces of AI, starting from the history of the field to the implications of recent accomplishments in AI and the more detailed analysis of how we might get from where we are today to human level AGI or even to *Superintelligence,* a general intelligence far beyond the human level. The author talks about every possible argument and every point of view regarding AI that it's hard to find the main conclusion, but he presents multiple viewpoints which gives the reader a well-rounded perspective to come to his own conclusions. Max Tegmark is an interesting and provocative thinker; he uses stories that seem like SF novels to show the possible ways that AI could develop. He did an amazing job explaining the most likely outcomes in a simple manner that even readers lacking technology knowledge could understand it. With the description of an AI evolution closer than we imagine it, he enables the reader to look its possibility, pros and cons, as well as its impact on humanity (jobs, laws, weapons) with a perspective of its future potential.

This book could be seen as a challenge for humans interested in the future of life, intelligence and consciousness, a challenge on how to create a benevolent future civilization of humans merged with a possibly even greater intelligence than our own. I truly believe that this will be the most important conversation of our time and we should ask ourselves what we can do to improve our future coexistence with AI and avoid the risks that might get us in trouble.

IVAN SAFTIĆ
University of Rijeka, Rijeka, Croatia