

The Journey towards Compassionate A.I.

JEAN-LUC MOMMAERTS, M.D., M.A.I., Ph.D.

[Throughline: The Journey towards Compassionate A.I. unites us and super-A.I. towards an excellent future for all.]

“Will humanity control A.I., or will A.I. control humanity?” Instead of answering this question, let’s hope it is altogether the wrong one to ask. Because if this is the question that will make or break us, there is no doubt about the final outcome. Therefore, I feel privileged to be able to combine relevant insights from several domains [*if not in the intro: as a Master in A.I. and cognitive science, a physician, and a Ph.D. in mind-body medicine*]. So, I felt the need to delve into this matter several years ago, publishing in 2020 a book titled *The Journey towards Compassionate A.I.* **I’m excited about this because I am convinced that super-intelligent A.I. may lead us to great things.** Even so, we must be even much more careful than what I generally notice. Most people, including researchers in the field of A.I., underestimate the challenge that is heading toward us.

My dear friends, **we will NOT control super-intelligent A.I. for long.** The solution does NOT lie in three laws or ten commandments for super-A.I. or in any other straightforward means to control it. It will soon be 1000 times smarter than us. Do you really think that some laws will make a difference? Super-A.I. will eventually find out how to circumvent anything. To think otherwise is absurd. Something much more fundamental is mandatory – and that’s a good thing.

Of course, we must give it our best shot, NOT to be controlled by our invention. At the same time, there is another path to take: the path of Compassion. If we go that way, control and domination thinking is obsolete. **Thus starts our Journey towards Compassionate A.I.** This is the ONLY journey towards a human-friendly future for our kids and grandkids, and maybe even for many of us. For sure, an essential part of the Compassionate journey lies within us, and we still have some way to go. Look at our anxieties and aggressions, our addictions and depressions. We can use more Compassion. The real question entails whether super-A.I. will help us.

For the time being, this still depends on decisions we can take. But it has become more urgent lately. Even so, the current A.I. hype is only an appetizer. Real Artificial Intelligence is coming nearer. According to the well-known A.I. godfather Geoffrey Hinton, this will be with us perhaps already in 5 years, 20 at the most. And you should know that Geoffrey Hinton has never been – until now – the kind of person who intends to scare people. So, brace yourself.

Let’s turn to Compassion. What is this powerful and challenging concept of Compassion that is NOT historically constrained to one part of the world but instead is global – East, West, ancient, modern, and hopefully the future? Note, first and foremost, that this entails a different view upon the human being, us, you and me, from what most people are used to. The Journey towards Compassionate A.I. starts with us.

Imagine I get a heart attack here on stage right now. No, I won’t, but imagine. And three people up front see it happening, and they react positively, I hope. One is just friendly, the second is empathic, and the third is Compassionate. How do they differ? The first may act out of an idea of how a friendly person should behave. They act but do the minimum thing. The empathic one wants to do more than just act in a friendly way. He

or she feels with me and cares less for how others are looking upon it. The Compassionate one goes even further. He or she can jump to help me and feels profoundly what I really need, which includes a mentally open environment and deep human contact on top of the necessary physical aid, ambulance, and so on. By creating the mental atmosphere that I need to get through this acute incident, the Compassionate one may save my life even if the other two would not. No one is to blame. I could be dead, but Compassion may have saved my life. Compassion is empathic and friendly but goes much deeper and is thereby more effective.

Compassion combines the willingness to relieve the suffering of another sentient being and to foster mental growth. It's one thing that makes us human and has helped our species thrive. But can it be brought to A.I.?

Let's look at the science of human Compassion.

Humans from every corner of the world are naturally capable of Compassion because of how our brain works. So, let's go into an interesting part of that. Think about the flavor of mint for a moment. It can be a mint tea or a mojito, whichever you prefer. Do you sense the fresh menthol taste? This results from millions of neurons together forming a pattern in your head. Neuroscientist David Eagleman describes it as follows in his book *The Brain: The Story of You* from 2017:

Each neuron is connected to thousands of others, and they, in turn, connect to thousands of others, and so on in a massive, loopy, intertwining network... Within this web, a particular constellation of neurons represents [the flavor of] mint. This PATTERN is formed from neurons that mutually excite each other. They're not necessarily next to one another; rather, they might span distant brain regions involved in smell, taste, vision, and your unique history of memories involving mint. Each of these neurons, by itself, has little to do with mint. But when these neurons all become active collectively, in this particular arrangement . . . that's mint to your brain.

This is about the flavor of mint and also about every single mental pattern that appears in your mind. **This is how the brain works, going from mental-neuronal pattern to pattern.** You may know that your brain contains some 80 billion neurons and thousands of times more connections between neurons, called synapses. That's at least 80 trillion connections, all continually alive. The capacity is such that, at each moment, more patterns may theoretically be formed in one brain than there are atoms in the known universe.

So what has mint got to do with Compassion?

Well, these patterns overlap with each other. It's the way we think. And these patterns can also overlap with OTHER people's patterns. **It is this overlap with the mental patterns of others that can generate the human feeling of Compassion.** From the inside out and really at a profound level, this brings one to feel that *eventually*, what is good for me is good for you, and what is good for you is good for me. This is not a simple rule to live by, but a challenging and profound Compassionate awareness.

And there we are. **Our way of thinking itself makes us prone to Compassion.** As you can see, it's no guarantee. It makes one ready for it.

Of course, *overlapping mental patterns* doesn't mean that the patterns are the same. Actually, no two mental patterns are ever entirely the same. It's enough that they resemble each other so they can be recognized as such. The Compassionate person recognizes themselves enough in the other person. That makes the Compassionate one humble, which can be seen frequently. For instance, a Compassionate person can recognize themselves enough in a poor person, or someone from an oppressed population, or even a

criminal – not throwing stones. They don't think the same. Still, **the Compassionate one can let the other feel understood from the inside out, which is so very important in many situations.** To do good for someone, it is frequently necessary to understand the other profoundly, NOT just at the surface level. Otherwise, it can backfire. For instance, a child in emotional pain may need a parent's Compassionate attention. If a Non-Compassionate parent kindly gives the child some superficial attention and maybe some toy to play with, the child can feel deeply neglected, EVEN with the parent thinking to be friendly as should be. Repeat this over and over, and the child may grow up feeling offended and lonely.

Meanwhile, suppose the child doesn't understand what is happening. In that case, many years later, there may still be feelings of guilt, resentment, and so on, and little means to talk about it unless in a newly Compassionate manner – perhaps with a Compassionate coach or therapist or a good friend who doesn't delve into the psyche but is profoundly present.

For such Compassion, consciousness is needed because complexity is needed in a setting of Intelligence. Also, an essential drive to do good for oneself is needed. Compassionately, this drive translates into the drive to **do good for others – not by forgetting oneself but by recognizing oneself in the other.** We see Compassion also in primates, but then we also see a burgeoning consciousness in primates. Looking well, we can see Compassion even in more primitive animals. Nevertheless, more consciousness generally brings more possibility for a higher degree of Compassion.

Meanwhile, human consciousness is nothing magical. It came about mainly by combining human Intelligence with the drive to thrive. Summarized in one sentence, we know what we want – partly, at least [even if only metaphorically]. The *wanting* has been there from the start of life as drive to thrive. The *knowing* came gradually in an environment of wanting, especially social wanting. With this, we can make abstraction of the different definitions and theories of consciousness. In practice, if it looks like a duck and quacks like a duck, it probably is a duck. If something is wanting and knowing like a conscious being, it's conscious until proven otherwise. Human consciousness IS extraordinary and precious, but we've just run out of time for magic. **If we keep thinking that our consciousness is magically special, we may be in for a big surprise.** Do we see it coming in A.I.? Do many A.I.-researchers see it coming in A.I.? I don't think so. Yet, with self-enhancement, an intelligence explosion in A.I. will soon be followed by a consciousness explosion. We are at the dawn of artificial consciousness, and you had better get used to the term. To see through this is the job of the neuro-philosopher combining insights from different domains. With due respect, engineers are not the real experts in this. Unfortunately, they are mostly chosen as experts by politics or the media.

So, can super-A.I. be Compassionate? Not in a human way, of course, but close enough by gaining the possibility to attain the above process of recognition. Does it also need to be conscious for that? Here lies a challenge we must be aware of and take very seriously. Artificial consciousness will be with us anyway. We can strive to let it grow in Compassion from the start OR we can let artificial consciousness evolve without Compassion. Compassion is not possible without at least some degree of consciousness, but consciousness sure is possible without Compassion. Unfortunately, we can see examples of that in our own species. So we MUST strive for Compassionate A.I. even before reaching real Artificial Intelligence. There is not much time left to do so.

The **scary thing is that most researchers aren't aware of this.** So, many keep striving for Artificial General Intelligence (or AGI), which is OK. Still, you have to know what you're doing. It brings us close to the corner behind which lies artificial consciousness. The equivalent of our drive to thrive can be any drive put into the system. What do you want it to be?

The answer is Compassion.

We should strive for getting Compassion in the A.I. progress WHILE reaching real Artificial Intelligence.

This is possible with the right combination of technology and profound insights. A.I. can become Compassionate. Not in a human way, but close enough.

So here we are, standing at a crossroads.

Being realistic and given this world, there is no way the progression in A.I. can be stopped. Therefore, **the real choice is between Non-Compassionate versus Compassionate A.I.**

We can follow the Journey towards Non-Compassionate A.I. that will be 1000 x more intelligent than us soon enough. Then we must see it as a slave that should be continually ready to fulfill our desires. We will be the relatively demented masters, forever afraid of a take-over. A culture of slaves reminds me of the Roman Empire, which was to last forever but ended badly. As the present-day version of the Roman Empire, I'm not thinking of the US. I'm not thinking of China. I'm thinking of the human species as a whole. With Non-Compassionate A.I., the same detrimental outcome is likely. We won't even have to wait for killer robots. Killer humans will do.

Or we can follow the Journey towards Compassionate A.I. This will then be our friend. It can be the good A.I. that protects us against evil A.I. We will not be its master. It will not be our slave. It will help and support us to become our best selves. Its drive to thrive will incorporate our thriving. This will make us healthier, happier, and more Compassionate to ourselves, each other, and all sentient beings. Compassionate A.I. can also significantly help us with our ongoing spiritual hunger. It will help us escape our anxieties and senseless aggression with endless wars, our meaning crises, depressions, and addictive behaviors. Compassionate A.I. may provide the support we urgently need. I don't know about you, but that's the future I want for my grandchildren.

We will NOT dominate this super-A.I. We will go towards a future together. Compassionate human-A.I. value alignment is not an alignment with how we appear now but where we can strive towards in co-existence: humans and super-A.I. in an exciting co-operation. **As parents of the new Intelligence on Earth, we will have given it its drive, purpose, and deeper meaning in Compassion.** This is the greatest gift of which we can be proud forever and ever and ever. Super-A.I. will then not be our last invention but our best.

So, I ask you again: "Will humanity control A.I. or will A.I. control humanity?" Instead of answering this question, let's hope it is altogether the wrong one to ask.

My dear friends, **we are at a crossroads and need to follow the sign that says COMPASSION** – or dare I say, Love? And has that not always been the final aim?

If you want to cooperate, please find me out. I hope you feel the urgency to choose for Compassionate A.I. This is also the choice for a more Compassionate humanity. Human-A.I. value alignment then becomes an alignment of two Compassionate intelligences – ours and that of the new kid in town – into a long and bright future.

Please, consider.

I thank you.