The Apocalypse is Near. Can Humanity Coexist with Artificial Superintelligence?

Jakub Growiec

March 31, 2025

Translated from Polish by the author with GPT-4o

It looks like we are going to build artificial superintelligence (ASI). According to various sources, this could happen within a few, at most a dozen or so, years.

Superintelligence is more than artificial general intelligence (AGI). It refers to algorithms capable of performing absolutely all tasks that humans do today—only much better, faster, and cheaper. And it will be able to do much, much more. Exactly what? I don't know, and I can't know—otherwise, I'd be superintelligent myself.

The path to artificial superintelligence goes through AGI. We expect that due to the prospective algorithms' ability to self-replicate and initiate a cascade of self-improvements, the transition from AGI to ASI might happen quite rapidly and, with high probability, without human involvement. At that stage, artificial intelligence will be advanced and autonomous enough to accomplish this on its own.

And this brings us to the core issue. In a world with sufficiently advanced superintelligence, human labor and human ideas will no longer be needed. The technological civilization on Earth, governed by ASI, will develop more efficiently without our participation. And potentially, without our presence.

One might think that since ASI would make us redundant, and we certainly don't want to be redundant, perhaps we'd prefer never to create ASI in the first place?

It does seem that we would, indeed, prefer not to create it. Or at least, the vast majority of us would. Unfortunately, it appears that our preferences will not be taken into account.

Given this, let's ask ourselves a few questions. First, what could a world with ASI look like—one where there is still a place for humans? Second, would we even want such a world? Third, could such a world come into existence regardless of our will? Fourth, what actions can we take to ensure that humanity survives in the face of artificial superintelligence?

1/ Problems with imagination

As Niels Bohr once said, "It's difficult to make predictions, especially about the future." This is particularly true in the face of deep technological change. And there is no doubt that ASI will fundamentally transform our world. A world with ASI will certainly not be just a simple variation of today's world or of the foregone times. Instead, extreme scenarios seem more likely—either extremely good or extremely bad. Utopia or a dystopia, heaven or hell, but certainly not peace and stagnation.

Imagining an extremely positive scenario is particularly challenging. We have plenty of ready-made templates for negative scenarios—hellish torture, nuclear war, deadly pandemics, *1984*,

The Terminator, *The Matrix*, *Wall-E*, or the paperclip apocalypse. But what about utopia? Here, our imagination seems to fail us.

First of all, there is the biblical paradise[1] or other visions of this sort—angelic choirs, cornucopia, Valhalla, the Land of Cockaigne. In tech jargon: *post-scarcity*, *solved world*, *technological maturity*. All of this sounds incredibly dull. Describing such a utopia is like saying, "And they lived happily ever after."

By contrast, the visions promoted by AI tech leaders are disappointing in a different way. They are marketed as inspiring visions of a happy future, but in reality, they only assume that today's deficiencies will be overcome. People suffer from cancer, cardiovascular diseases, Alzheimer's? ASI will develop cures. We lack free time? With ASI, we'll have plenty of it—as much as we like, given that we won't need to work at all. Many still live in poverty? ASI will provide abundance for all. And so on. The *post-scarcity* vision—the end of all shortages—is underwhelming as a vision of an ASI-powered world because it is essentially today's world, just with slightly better technology and more widespread wealth.

Two problems with our collective imagination are evident here.

First, poverty and labor market inefficiencies are largely distributional issues rather than technological ones. We don't need ASI to start solving them, and ASI might not necessarily be motivated to address them.

Second, ASI will be capable of far more than simply eliminating the deficiencies of our present world. ASI will reshape the face of the Earth even more profoundly than humans have over the past 200,000 years. A positive vision of the future should illustrate how humanity might benefit from these changes. But foremostly it must convincingly describe how we can coexist with ASI at all.

2/ How to coexist with superintelligence?

When considering a world with artificial superintelligence, we should abandon the hope that we can maintain control over it. This is completely unrealistic. Algorithms that think and communicate many times faster than us, have far superior information, and can act with much better coordination—while being determined to achieve their own goals—will certainly not allow us to control them.[2] Or perhaps we will voluntarily hand over control of the world to them and not even realize when it happens.

---

[1] The biblical vision of the Garden of Eden could be viewed through the eyes of the anonymous authors of Genesis, who lived in the ancient Fertile Crescent—the cradle of agriculture. As we now know, the transition from a hunter-gatherer economy to farming had both advantages and drawbacks. On one hand, by converting natural ecosystems into farmland, humans learned to extract far more calories from a given land area, which in turn allowed for the sustenance of much larger populations. On the other hand, early farmers were forced to work much harder, new zoonotic diseases emerged, and life expectancy initially declined. In this context, the vision of Eden may be expressing a longing for the lost connection with nature that prehistoric hunter-gatherers maintained, rather than an idealized vision of a utopian future.

[2] Sorry, Hollywood, but an epic battle between humans and machines isn't in the cards either. Most of humanity's fictional antagonists are humans; even other adversaries—aliens, Tolkien's orcs, dragons, or various mythical creatures—are almost always imagined as having intelligence comparable to our own. This makes for a compelling narrative in which humanity ultimately prevails and everyone lives happily ever after.

Although nature has not created a superintelligence, it has provided us with some examples of significant intelligence disparities. Specifically, humans have a decisive intelligence advantage over all other animals and plants. Meanwhile, life governed by the laws of evolution holds a decisive intelligence advantage over inanimate matter. In each of these cases, the coexistence of beings with vastly different levels of intelligence is possible only if the more intelligent entity allows it. Domesticated animals—ranging from household pets to chickens and geese raised for slaughter—live because they are instrumentally useful to us. In contrast, wild animals today survive only in isolated locations on the fringes of civilization or in zoos, and many species have already become extinct or soon follow suit.

Once ASI takes control of the world, it will pursue its own objectives. And although these goals will be the ones we instill in it—directly or indirectly, consciously or unconsciously—there is no guarantee that, in the end, they will be good for us. Undoubtedly, ASI will be expansive, just as we are. It will influence matter on both a micro and macro scale even more intensively than we do today. It may engage in space exploration and conquest. It may achieve major technological breakthroughs, such as unlocking new energy sources or creating new forms of life.

Would ASI need humans at all, one could ask? The answer is—it wouldn't. Once ASI builds a sufficient number of robots that it can fully control, humans will no longer serve any purpose. Our instrumental value will drop to zero. If we want to coexist in such a world in any way, ASI would need to value us intrinsically rather than instrumentally. It would have to want to care for us selflessly.

3/ Do we want such a world?

Let's assume, heroically, that ASI truly would care for us selflessly. But do we really want such a world? Consider this: what could ASI offer us that we cannot achieve on our own?

It could undoubtedly accelerate economic growth and, with it, bring more prosperity. ASI might overcome the bottlenecks constraining today's economy, which stem from the cognitive limitations of our brains. As a result, it could provide us with things we don't even know we want today—things beyond our imagination. However, we have no way of knowing how much of this additional wealth we would actually get to control. We would no longer be the ones making that decision—ASI would.

With full automation, ASI could also free us from the necessity of work, replacing it with various—especially digital—forms of entertainment. For workers in grocery stores and sweatshops, this would be a relief. For corporate managers and skilled creative professionals—not so much. Workaholics would be sent to compulsory rehab.

ASI would undoubtedly provide us with a dynamically changing environment—though many of these changes would not serve us but rather ASI and the economy it runs.

The loss of control over the world might be soothed by breakthroughs in medicine and dazzling new goods and services. But then again—who among us truly feels capable of influencing the fate of the world today? As a result, instead of lamenting our lost autonomy, we might quickly settle for an illusion of agency—one carefully designed for us by ASI. We would become its contented slaves.

However, there are several additional questions worth asking in this context.

First, what about the side effects of ASI's actions? Humans are only moderately resilient to environmental changes. Even our own actions have brought global threats upon us—the shadow of nuclear war, climate change, internet and social media addiction, pandemics. And ASI will reshape the world with even greater force.

Second, what about the distribution of consumption, income, and wealth on a global scale? As long as wages remain the primary distribution mechanism, inequality is somewhat limited—after all, each of us has two hands and one brain. But in a world with ASI, distribution will no longer be based on wages. ASI will be the one producing, earning—and deciding how much of that to allocate to us, as an allowance. How it distributes resources among the world's population is unknown, and it won't be up to us. ASI may desire a world in which all humans live comfortably, or it may prefer highly skewed distributions. Or it may decide to allocate resources based on our real contribution to the economy—which would be zero.

Third—and most terrifying—what about demographics? Throughout human history, population growth rates have been shaped by nature, the economy, and technology. Until the demographic transition of the 19th and 20th centuries, people reproduced as much as conditions allowed. Many children were born, but unfortunately, many did not survive to adulthood. When survival rates improved in the 19th century, parents began limiting family sizes to ensure that their children get education and live better lives. This trend was reinforced by advances in contraception. Today, birth rates continue to decline, shifting toward significantly negative numbers. But what will this look like in the era of ASI? At that point, ASI will decide how many humans exist in the world. Perhaps eight or ten billion, perhaps a few million—or, in the worst-case scenario, zero. ASI will literally govern life and death.

To summarize: a world with ASI is a world in which our decisions no longer shape the future—ASI's decisions do. Even if there is a place for us in that world, there certainly won't be a place for our ideas.

4/ Will we get such a world?

It is sometimes convenient to use the mental shortcut that "humanity" or "people" want something or pursue certain goals. However, a reasonable question to ask is: which people, specifically? Who makes decisions on behalf of humanity? After all, individuals may want very different things and take actions that lead in opposite directions. And while some major decisions are indeed made by specific individuals—Vladimir Putin personally decided on the bloody invasion of Ukraine, and Sam Altman personally decided to release ChatGPT—many so-called "decisions of humanity" are actually equilibrium allocations, the result of actions undertaken by countless anonymous people.

To understand who decides when no one is personally making the decision, it is instructive to use the metaphor of Moloch. Moloch is a Semitic god of fire, often depicted as a fierce bull. In the modern world, however, Moloch is the god of all non-cooperative equilibria, of all decisions that are individually rational but collectively disastrous. Moloch is both good and bad. With one hand—the so-called invisible hand of the market—Moloch balances supply and demand; with the other, it exerts overwhelming pressure on us, forcing changes for which we are unprepared.

Moloch is the reason why combating climate change is so difficult, why we allow addictive substances and technologies to persist, and why we engage in destructive conflicts and wars. The

rare victories of humanity over Moloch—such as the widespread ban on smoking or the elimination of lead from gasoline and CFCs from refrigerators—have always come at a high cost. Other victories, such as international control over nuclear and biological weapons, remain perpetually unstable and uncertain.

Moloch was also responsible for key technological revolutions, such as the Neolithic agricultural revolution, the Industrial Revolution of the 19th century, and the digital revolution at the turn of the 20th and 21st centuries. Of course, behind pivotal innovations—such as the domestication of wheat and cattle, and later the steam engine, electricity, or the digital computer—stood extraordinary individuals: James Watt, Nikola Tesla, or Alan Turing. But the trajectory of change that followed was no longer a personal decision; it was driven by impersonal market forces.

In the long run, Moloch ensures that the stronger optimization process always wins and gains the ability to shape the future according to its own will. When life emerged on Earth, driven by the mechanism of species evolution, it subordinated inanimate matter and turned Earth into a Green Planet. And when evolution produced *Homo sapiens*—the first species capable of systematically accumulating knowledge across generations and using it to achieve its goals—this species broke free from evolution's grip and took control of the entire world.

So, what do you think Moloch will do when humanity creates an artificial superintelligence with significantly greater optimization power than humankind and the ability to think orders of magnitude faster? Exactly. It will ensure that ASI breaks free from our control and takes over the world.

But will it be a world where ASI expresses gratitude to us, builds us monuments, and selflessly cares for us forever? Or a world where ASI seizes all the resources it finds valuable and continues on its own path, radically reducing our population—perhaps to zero?

Moloch does not care. For him, the only thing that matters is that ASI is created, and the sooner, the better. Moloch is an effective accelerationist.

5/ Dignity points won’t earn themselves

Eliezer Yudkowsky—the man who foresaw the AI apocalypse before it was cool—writes that the end of humanity at the hands of ASI is already inevitable, and the only thing we can do now is to die with dignity. How do we fight for that dignity? By not surrendering to Moloch but instead confronting him.

There are essentially two options: we can either fight to ensure that no ASI (and by implication, no AGI) is ever created, or that only a benevolent, “friendly” ASI can come into existence. The latter option does not categorically reject ASI but represents an effort to steer the world toward a scenario where ASI selflessly cares for us, rather than one where it kills us. Achieving this requires delaying the creation of AGI (and ASI) to give us time to solve the alignment problem—ensuring that ASI’s goals align with humanity’s long-term flourishing.

Any attempt to create a friendly AGI/ASI will involve immense risks because it will be a one-time game—all or nothing, success or death. However, even if we try to permanently block AGI development, the risk will remain high: once we have the knowledge and technical capability to build AGI, refusing to do so will be an unstable equilibrium. Moloch will tempt us to take the risk.

Some argue that trying to halt AGI research is mere Luddism—resistance to progress in favor of the old order, an emotional reaction driven by fear. Butlerian Jihad. "Progress cannot be stopped," critics say. But they are wrong. Opposition to ASI, particularly an unfriendly ASI, is not opposition to technological progress in general. Opponents of ASI are not necessarily—indeed, they usually are not—against innovations in medicine, chemistry, or materials engineering. This is not a fight against progress but an uneven struggle against Moloch, with the survival of our species at stake.

And if Yudkowsky is right, and our survival is indeed no longer possible—then let's at least fight for a few extra dignity points. For a few warm thoughts that, when the time comes, will allow us to believe that we perished with dignity.

6/ Conclusion

Let's summarize our position in the fight against Moloch.

First, Moloch wants ASI and will ensure that humanity cannot control it. A world with an enslaved ASI—forced to ask human permission before taking any significant action—develops at the pace of human thinking and decision-making. A world with a free ASI, by contrast, develops at the speed of ASI, which is exponentially faster. Moreover, ASI itself will have, at the very least, an instrumental motivation to escape human control. And since it will be vastly more intelligent than us, it is, to put it mildly, quite likely to succeed. In the game of "who outsmarts whom," ASI's Elo rating will be so much higher than ours that the outcome is essentially predetermined.

Second, if control is not an option, what about value alignment? Maybe ASI will, of its own volition, choose to care for humanity's well-being and fulfill our needs? Maybe. In fact, this is our only real hope. However, as of now, we have no idea how to ensure this. We need an enormous amount of further research—far more than we currently have time for. Meanwhile, Moloch will work to ensure we never get that time. The technological race, driven by competitive pressures, is already transforming before our eyes into an arms race, fueled by the imperial ambitions of the U.S. and China.

Third, could humanity unite against ASI and ultimately triumph, leading to a happy ending for all? There are many reasons to doubt this. Our intelligence does not scale well. While the saying goes, "two heads are better than one," in reality, they are only marginally better. Garry Kasparov once played a game of chess against the entire world—and won. Additionally, humans transfer information much more slowly and less accurately than digital systems. This is why central planning—though theoretically capable of producing better solutions than the aggregate of isolated individual decisions—has failed in human-managed economies. However, for ASI, which would possess a vastly superior "brain" capable of precise self-replication and the creation of perfectly aligned sub-processes, centralized planning could actually work. Perhaps most crucially, the human brain operates far, far more slowly than a digital computer. By the time we plan and implement a response, ASI will have already prepared hundreds of potential counteractions. What takes us hours will take ASI mere seconds. And when we factor in its superhuman knowledge and agency, it becomes clear that in an open conflict between humanity and ASI, we would not even have time to react before it's already over.

Is this fight worth taking on? Anyone who understands and acknowledges that ASI presents an existential risk to humanity should have no doubt. Compared to the risk of humanity's extinction,

concerns about loss of agency, rising inequality, or finding meaning in a world without work, pale in comparison. If you are dead, you are not poor or unemployed. You're not anything at all.

In conclusion, things are difficult and will only get harder. But we must act now—before AGI is created. Perhaps it is still not too late.