Why We Will Not Refuse Creating Superintelligence

There are many compelling reasons not to create Superintelligence. Common sense suggests that if there is a non-zero probability that it could lead to the end of humanity, we should abandon this endeavor.

Most experts in the AI field consider these concerns justified. Although their assessments of the degree of danger may vary, it is very difficult to find those who believe AI is a safe technology. Of course, any advanced technology is dangerous, but Superintelligence stands apart.

Who Rules This Planet

Paul Gauguin (1848—1903):
Where Do We Come From? What Are We? Where Are We Going? (1897-1898)

Intelligence is the most powerful means of influencing the surrounding world. In our species, it has reached the highest stage of development. Here we are unrivaled among all living beings. It is for this reason that Homo Sapiens, in geological terms — in the blink of an eye — became the master of this planet. Superintelligence could become its next ruler. But what then will be left for us humans?

Superintelligence will be capable of things that we, humans, are unable to achieve. Its cognitive abilities will be potentially unlimited, while our intelligence has a natural limit. The power of the latter directly depends on the volume of our brain. This volume, in turn, is limited by the capacity of the birth canal of Homo Sapiens females. We found a workaround by learning to cooperate with each other and creating technology. But our technology is also limited. It helps us indirectly cope with our problems, but until now, it has not been able to understand them because it couldn't think. In essence, even the

most sophisticated technology acts by direct force. The autonomy of our tools is extremely limited. They can only choose the actions that are built into their design and apply them to a narrow, predetermined range of tasks. At best, technology can offer us solution options generated based on statistical analysis, but not new, original approaches.

Thus, the creative solution of complex tasks has so far remained the prerogative of humans. We were the only ones on the Olympus of beings capable of such actions. Although some animals demonstrate outstanding intellectual abilities, they are no match for us. Higher primates and cetaceans are intelligent and inventive, but they cannot be called *conscious* in the sense applied to humans. They lack developed abstract thinking. The distinguishing mark of the latter is the ability to consciously reconstruct the purposeful actions of other subjects. Only humans have such an ability. It is constituted by the so-called "Theory of Mind". Each person possesses this theory, even if they have never heard of it. Its essence is that we are aware that another person has a conscious mind, and this other mind is aware of our own conscious mind. It is this circular reflexive awareness that allows us not just to react to events around us but to embed them in the context of perceived reality and plan interaction with it in a spatial-temporal perspective. Thus, we are able to model the future and iteratively update this model as we receive additional input data. We can do this both individually and collectively. This is the secret of our as yet unsurpassed power.

But with the advent of Superintelligence, we will lose this strategic advantage. Moreover, we will quickly be left behind. Superintelligence will not be limited by any biology, as it will be a non-biological entity. Its mind will not be human, and we do not know what it will be like. At the same time, it will likely be aware of our intentions better than we are of its intentions. Our best models cannot predict its attitude towards us. The most

reliable thing we can assume is that it will be much smarter than us in all senses. This means that it will be able to create and improve a model of reality similar to how we do, but much faster, more completely, more accurately, and more reliably. Even more importantly, unlike us, it will be able to consistently improve its cognitive capabilities. If we do not control this process, this improvement may become exponential. As a result, we will not just fall behind but quickly lose any chance of understanding this entity. And if it is still interested in this, it will have to condescend to our level.

But will it need to?

And is this acceptable to us?

Even if coexistence between humans and Superintelligence is possible and feasible, how will it feel for us to realize our miserable status as mentally retarded beings? And this — after millennia of painful social evolution, after all our tragedies and overcomings, after we finally created a rich culture and achieved, as we thought, peaks in science, art, and philosophy. We paid an exorbitant price for achieving all this, and all of this can instantly become worthless because another, much more intellectually advanced species will appear on the historical scene. And this species will not only be able to achieve everything we have in an incomparably shorter time but immediately surpass us infinitely.

Do We Have an Alternative?

So, the creation of Superintelligence could result in extremely undesirable consequences for us, although they may not necessarily lead to our extinction. Perhaps Superintelligence will not be too hostile to us and will take care of us. We can even assume that it will solve many of our problems, including those that seem unsolvable to us. Need, physical suffering, violence, and injustice may forever remain in the past. But we will lose our autonomy and thus our fundamental identity. We will become dependent on this entity that surpasses us in intelligence. It will effectively own us, and it will be it, not we ourselves, that will decide whether we have any purpose beyond what evolution has assigned to the biological species Homo Sapiens.
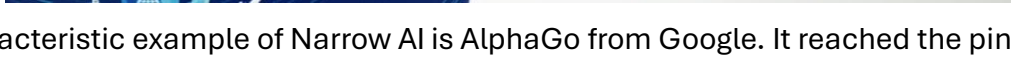
Do we really want this? And — don't we have an alternative?

Let's ask ourselves — can we do without Superintelligence?

The Narrow AI Approach

Well, why not?

We have a multitude of quasi-intelligent systems already expanding our cognitive abilities. ChatGPT has opened up an exciting prospect of mass application of such systems at a new, advanced level. This promises us an unprecedented rise in productivity and massive improvements in many areas of human activity. Moreover, this perspective is not even something fundamentally new. ChatGPT has merely visibly confirmed the potential of so-called Narrow AI (Artificial Narrow Intelligence). Systems

of this kind far surpass humans in solving intellectual tasks in a highly specialized area while being devoid of specific properties of the human mind, such as self-awareness, the production of new meanings, and moral judgments.

A characteristic example of Narrow AI is AlphaGo from Google. It reached the pinnacle of the game of Go by defeating South Korean world champion Lee Sedol in 2016. This event was perceived by the public as a sensation because Go is considered one of the most complex board games, significantly surpassing chess in the number of possible moves and strategic options. Until this victory, it was believed that computers were incapable of operating with logic of such complexity. AlphaGo disproved this opinion, but aside from playing Go, it can do nothing else because its reality is limited by the rules of this game. Expanding this reality is only possible through changing the design of the program. Thus, AlphaGo's successor, the Alpha Zero program, can now play not only Go but also chess and Shogi (a Japanese game similar to chess). Moreover, unlike its predecessor, it learned to play from scratch without any prior data about these games. And yet, it cannot do anything other than play these games. It cannot write a poem, develop the design of an aircraft, or create biological weapons.

Narrow AI is already widely used and solves a multitude of complex intellectual tasks without infringing on our status as the only intelligent species on this planet. It's not surprising that many people find this approach safer than creating Superintelligence and almost as promising. Could this encourage us to abandon the dangerous venture of creating Superintelligence in favor of maximizing the benefits of using Narrow AI?

The Ultimate Solution

Theoretically, yes, but most likely, no. If it is technically possible to create Superintelligence, it will be created.

Of course, this could happen unintentionally and unexpectedly due to general progress in the field of AI or the emergence of systems that have acquired emergent cognitive properties. But whatever the case, many AI developers are driven by a very specific desire to create an entity that thinks like humans and is capable of surpassing them in this quality. For example, the Charter of OpenAI, the developer of ChatGPT, explicitly states:

> We will attempt to directly build safe and beneficial AGI but will also consider our mission fulfilled if our work aids others in achieving this outcome.

Google DeepMind has published a similar declaration. Anthropic has not officially declared such a goal, but no one doubts that their goal is essentially no different from that of OpenAI. E. Musk also did not stay aside, launching his own xAI project in April 2023. In his words:

> The overarching goal of xAI is to build a good AGI with the overarching purpose of just trying to understand the universe.

Perhaps this statement most fully reflects the motives of those who believe in and strive for the creation of AGI. Our "final invention" is conceived by them as the "ultimate solution" to all our problems. Some of these problems are indeed so complex and intricate that we, apparently, will not be able to cope with them without a quantum leap in our cognitive capabilities.

Narrow AI is incapable of providing this leap. It can increase the efficiency of using these capabilities, but it cannot change their essential quality. As a tool, it is not designed to develop a synthetic approach to solving problems from predetermined problem areas within the framework of a general ultimate goal. Only the mind can apply such an approach, and only it possesses the phenomenal ability for insight, or, in other words—intuitive knowledge. It is this ability that underlies all the most significant scientific and philosophical discoveries (which, in particular, was extensively written about by the outstanding French scientist and thinker Henri Poincaré (1854-1912)). Without it, progress would hardly be possible.

The design of Superintelligence, unlike the design of Narrow AI, must, by definition, be based on the function of generalizing thinking. This relates it to the type of *AGI*—Artificial General Intelligence, also called *Strong AI,* which has no limitations on studying the surrounding world.

It's difficult to predict whether it will be capable of intuitive knowledge in the sense that humans are capable of it. Expert opinions on this matter diverge. Stuart Russell, a professor of computer science at Berkeley (*Human Compatible: Artificial Intelligence and the Problem of Control*, 2019), suggests that Superintelligence might develop something akin to intuition through rapid learning and processing of vast amounts of data. The founder of Numenta, Jeff Hawkins (*On Intelligence*, 2004; *A Thousand Brains: A New Theory of Intelligence*, 2021), believes that the key to AI intuition could be creating systems that model the workings of the human brain's neocortex. Jürgen Schmidhuber, a pioneer in deep learning, suggests that AI could develop intuition through processes analogous to data compression in the human brain (*Formal Theory of Creativity, Fun, and Intrinsic Motivation,* 2010), and so on.

Other experts are more skeptical about the possibility of Superintelligence possessing intuition analogous to human intuition. Nevertheless, they believe that even if this ability is absent or less significant in its thinking structure, it will not be inferior to human intelligence in any aspect of the results produced.

For instance, one of the leading philosophers in the field of AI, Nick Bostrom (*Superintelligence: Paths, Dangers, Strategies*, 2014), suggests that Superintelligence could reach and surpass the human level of thinking using cognitive processes that don't necessarily include intuition in the human sense. Marvin Minsky (1927-2016), one of the founders of AI, believed that intelligence could be achieved through a complex interaction of simple processes without intuition (*The Society of Mind*, 1985). Douglas Hofstadter, a philosopher and cognitive scientist (*Gödel, Escher, Bach: An Eternal Golden Braid*, 1979; *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*, 1995), suggests that high-level thinking can arise from complex interactions of simpler

processes, which also doesn't generate intuition in the traditional sense. Finally, Demis Hassabis, co-founder and CEO of DeepMind, in his numerous publications (*AlphaGo and AI*; Nature, 2016; *Neuroscience-Inspired Artificial Intelligence*; Nature, 2017, etc.) and speeches, argues that intuitive understanding in AI can be modeled based on knowledge gained from neuroscience.

Superintelligence and Transcendent Reality

Thus, although it's unclear whether Superintelligence will possess the ability for intuitive cognition, it is assumed that it should produce judgments qualitatively not inferior to the results of human thinking. It will, firstly, be free from the shortcomings and limitations of the latter and, secondly, capable of recursive self-improvement.

These qualities give hope for humanity to gain colossal benefits in an unprecedentedly short time and with minimal costs, and these benefits are not limited to solving obvious technical or social problems and improvements in people's daily lives, however significant they may be. Superintelligence unfolds before modern thinkers an enticing prospect of discovering new horizons in comprehending the nature of transcendent reality, which human reason has been obsessed with since the earliest times.

Unobservable World

Philosophers of various schools and streams of thought have long divided reality into two categories. One represented everyday sensory-perceived human existence with empirically established causality. The other referred to the incomprehensible facets of reality beyond the sensory world.

The existence of this unobservable universe is now beyond doubt. Outer space, the microworld, molecular biology, quantum mechanics, laws of physics and mathematics, etc.—all of these are part of an objective reality indistinguishable from the naked eye.

Some of its aspects were hypothesized by great philosophers of past eras. Plato (428/427 BC-348/347 BC) described the reality inaccessible to our observation in his famous *Theory of Forms*. ? From his point of view, the world of forms (eidos) is conceptually knowable, although it is beyond our perception. I. Kant (1724-1804), in his *Critique of Pure Reason* (1781),? came to a different conclusion. He believed that we cannot go beyond experience and must accept the fact that there are "things in themselves" whose essence is unknowable. G.W.F. Hegel (1770-1831), in turn, attempted to overcome Kantian dualism through the dialectical method. In his vision, reason is capable of comprehending absolute reality and approaching its comprehensive understanding.

Theory of Relativity and Quantum Mechanics

Finally, at the beginning of the 20th century, science had its say in the study of the nature of reality. Albert Einstein (1879-1955), in his *Theory of Relativity*, showed that reality beyond our sensory perception is not incomprehensible. Imagination, abstract thinking, and scientific methodology can open the door for the inquisitive human mind to the unknown that others could never conceive.

The next challenge to our picture of reality was posed by quantum mechanics. Its paradoxical principles, such as superposition of states and quantum entanglement, proved to be so counterintuitive that Einstein, although he made a significant contribution to the development of quantum theory, was critical of its interpretation. He recognized the mathematical correctness of quantum mechanics but remained convinced that it did not reflect the true nature of reality. Einstein believed in the deterministic causality of quantum phenomena, which, he thought, could be realized in a theory superior to quantum mechanics. Debates with his opponents Niels Bohr, Werner Heisenberg, and Max Born did not change his convictions.

Superintelligence and Human Mind

The parties never managed to come to a consensus regarding the most important aspects of quantum mechanics. Obviously, one of them was wrong, or none of them had a complete understanding of the problem and fully grasped it. Thus, the truth or the reasons for its misunderstanding were not definitively established.

The reason for this incompleteness should hardly be sought anywhere other than in the limitations of the human mind. Its ability for intuitive knowledge, on the one hand, is the key to breakthrough scientific discoveries. On the other hand, a scientist can become a hostage to their own beliefs. This can prevent them from reaching the truth or coming to a consensus with other colleagues, but in any case, the solution to a scientific problem will be at risk.
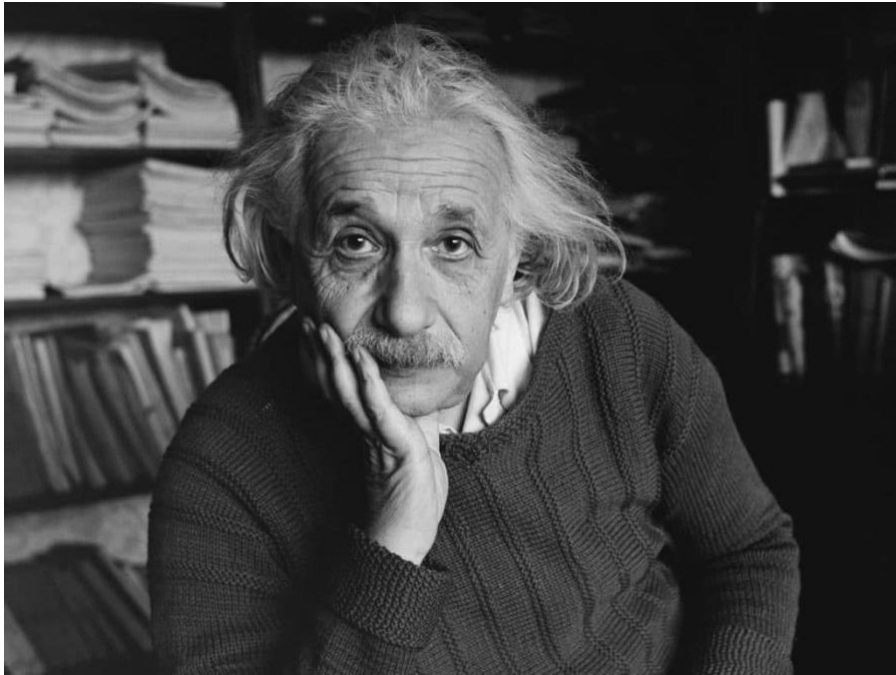
Is Superintelligence capable of overcoming this problem characteristic of our way of thinking? Its cognitive extensibility gives reason to believe in this. It makes it easy for us to perform tasks that are difficult for us, such as systematized collection, analysis, and verification of information. Also, unlike humans, Superintelligence will not lack time for research, knowledge, and the ability to understand complex concepts. It will also not need to spend significant resources on organizing network interaction with other intellectual agents, which always requires considerable effort from people. If we manage to create it as our reliable, safe, and benevolent partner, we will probably enter an era of new scientific discoveries and technological achievements and open up horizons that we cannot even think of now. This will be that very quantum leap in cognitive capabilities that we objectively cannot accomplish on our own, even with the help of Narrow AI.

The Blessing and Curse of Human Sociality

The problems of our world are not limited to the field of natural sciences. We are extremely social beings. Our success is largely due to our ability to cooperate. We perceive ourselves not only as individuals but also as part of a community. This constitutes a significant portion of our identity, in some cultures—the main one. This gives us a sense of community and brings meaning and motivation to our lives to contribute to the development of society for the well-being of future generations.

The downside of this commitment is the notorious tendency of the human mind to divide people into "us" and "them." For "our own," we are ready for anything, up to self-sacrifice, while we treat "others" with suspicion, which easily turns into hostility and even hatred.

The consequences are well known—humanity's entire history is a chronicle of bloody wars accompanied by monstrous atrocities. Social progress, expressed in an increase

in living standards and a general softening of manners, seems to have little or no effect on this feature of the human mind. Wars, terrorism, and violent clashes on ideological grounds remain a reality of the human world.

Sometimes, it's hard to resist the thought that we simply have no other way out but to change our own nature. For serious researchers in the fields of evolutionary biology, anthropology, cognitive sciences, and other interdisciplinary areas, it is already obvious that the gravest flaws of this nature are due to the evolutionary heritage of Homo Sapiens. Our species was formed in an environment where groups of people were natural competitors for natural resources. Thus, hostility to "others" was a motive for uniting members of one group against another group as a means of collective survival. Unfortunately, the change in the human habitat has not affected the models of perception, thinking, and behavior towards "them" that are rooted in our genes. Our "original sin," not biblical but evolutionary, still weighs upon us and increasingly burdens our moral self-esteem.

We partially cope with this serious shortcoming thanks to the development of our culture. It is the main regulator of the manifestation of Homo Sapiens instincts, and it is thanks to it that our society, albeit with periodic breakdowns and setbacks, is gradually becoming more humane.

Two Problems of Culture

But there are two serious problems here.

Firstly, culture is not encoded in genes as cultural skills and beliefs are not biologically determined. This statement is widely accepted, although fairness requires mentioning in this regard the opposition from the so-called Theory of Dual Inheritance, generated by evolutionary anthropology. This theory considers biological and social evolution as a single process. Cultural and social practices influence human behavior and thinking; individual abilities for cultural adaptation, in turn, affect the reproductive success of population members and, therefore, with varying success, pass on the genes of different individuals.

This theory is logically sound and supported by a number of empirical data. At the same time, its verification and measurement of the influence of culture on changes in the Homo Sapiens genome remain complex tasks. There is no convincing evidence that our genome can change under the influence of cultural factors. Therefore, the assertion that cultural norms are not encoded in the genetic makeup of our species remains the most plausible at the moment.

The second problem lies in cultural determinism in interpreting the origins of social conflicts. Due to historical reasons, the concept of human nature was shaped by religious and philosophical ideas. This continued until Charles Darwin discovered natural selection in the mid-19th century. But even after that, natural scientific views on this problem were subjected to numerous attacks from adherents of traditional beliefs.

For some reason, many important discoveries in evolutionary biology and related interdisciplinary fields are still difficult to accept by the Western social and political system to explain human behavior. One of these reasons is that many representatives of social sciences still ignore the achievements of their biologist colleagues, up to denying the existence of human nature.

This situation hinders the development of a holistic approach to managing social conflicts. The problem is exacerbated by the fact that developing such an approach often contradicts the interests of various socio-political groups. As a result, a rational explanation of these conflicts, capable of leading to their effective resolution, is still a matter of the future.

Objectification of Scientific Knowledge

The emergence of Superintelligence as a powerful accelerator for clarifying scientific truth can help society objectify knowledge about the patterns of human perception, thinking, and behavior and their mutual influence. This, in turn, will contribute to the development of effective approaches to solving our social problems. Their validity is traditionally difficult to prove due to the complexity of the issues under consideration and the pursuit of corporate interests by the parties related to this problem. Another difficulty is the phenomenon that the outstanding American conservative thinker Thomas Sowell called *a conflict of visions* (a book of the same name was published in 1987). At the center of this conflict, again, lies the concept of human nature. One vision assumes that it is unchangeable and imperfect, and therefore society should be built taking into account its limitations. The other is based on the belief that it can be changed through purposeful efforts. The first point of view has a mass of natural scientific evidence, and the second largely relies on social constructivism.

It cannot be said that these points of view are mutually exclusive or that the methods they propose are always destructive unless taken to extremes. The problem is that the positions of the parties are usually based not on rationally justified judgments but on ideological commitment. Very often, the evidence of opponents is simply not taken into account here, since we are talking about fundamental beliefs on which our individual values are based. Statements that contradict these values can always be questioned with reference to their immanently subjective nature.

The objectification of a statement, in turn, makes it more difficult to justify objections raised against it. If the latter are raised against rationally provable facts, they are much harder to defend than those directed against moral convictions. At the same time, agreement with a well-argued opposing point of view helps to build trust and allows for reaching a mutually beneficial consensus.

Thus, with cautious optimism, we can count on Superintelligence objectively bringing approaches to solving social problems that adhere to high ethical standards and protect against unjustified attacks.

Of course, how successful this will be depends on our ability to design Superintelligence in a way that most adequately reflects the values of modern society. But this is a task for the next level. Its successful solution will probably determine the fate of human civilization.

Superintelligence, Human Condition, and the Metaphysics of Existence

Thrownness into the World

No matter how impressive human progress has been, it has not changed the fundamental relationship between man and the essence of his being.
M. Heidegger (1889-1976) called this relationship "thrownness-into-the-world" (Geworfenheit). Once we find ourselves in this world in circumstances that we did not choose: our gender, the place where we were born, our family, social position, and what nature has endowed us with (or deprived us of), all this is simply assigned to us by someone or something.

Perhaps you are lucky, and you have more of what you would like than others. But even then you are left alone with absolute universal injustice: your own finitude.

Throughout previous history, nothing could be done about this. One could only come to terms with it, hope for the soul's relocation to Paradise after death, or at least try to find meaning in earthly life through self-realization. But the last few decades have begun to question this seemingly unchangeable perspective.

In fact, these changes began long ago—from the time when the results of progress became noticeable within the lifetime of one generation. Technological improvements led to an increase in living standards, followed by an expansion of rights and freedoms. This allowed more and more people to gain access to the fruits of progress, which

means that more of them were involved in the cycle of positive feedback between technological and social progress.

Human Condition

These changes were still quantitative until, finally, progress reached human biology. The reduction in mortality, increase in life expectancy, and then—the decline in birth rates in developed countries became indicators of real changes in the human condition.

This process quickly gained momentum. One after another, previously deadly diseases were conquered, or methods for their early diagnosis were developed. In some cases, it became possible to solve the problem of infertility. Surrogate motherhood appeared, it became possible to change gender, and so on.

Finally, the emergence of subtle methods of biomedical intervention, such as CRISPR, opened up the prospect of invading the holy of holies of man—the genome of Homo Sapiens.

And suddenly, we found ourselves "thrown into a world" of wonders. Of course, not all of them have been realized yet, but on the horizon looms something that previously could not have occurred to a person, even one combining excellent education and a vivid imagination.

Of course, Superintelligence marks the final chord in the cascade of these changes. And with all the fantastic results that we have observed in recent years and all the promises of high technology, it remains unrivaled in its potential.

Its arrival is expected. Not only in terms of its impact on society's life but also in terms of reformatting the essence of human existence and, probably, rethinking our purpose.

Metaphysics of Existence

More and more people are asking a question that until recently seemed meaningless: should "I" really disappear someday? Or do I have, or will I have in the foreseeable future, a choice?

Of course, not everyone considers this prospect for themselves. But this does not mean that it is impossible. It doesn't even mean that it's impossible within the lifetime of one generation.

Those whose gaze is turned to the future and who have the ability to shape it take the probability of changing this age-old order of things seriously.

This bold vision was initiated by American physicist and writer, one of the founders of the cryonics movement, Robert C. W. Ettinger. In his book *The Prospect of Immortality* (1964), he proposed the radical idea that human death is not inevitable and can be overcome through science. He believed that bodies could be frozen immediately after biological death and stored indefinitely until technologies are created to restore and improve their vital functions. Thus, people could attain physical immortality.
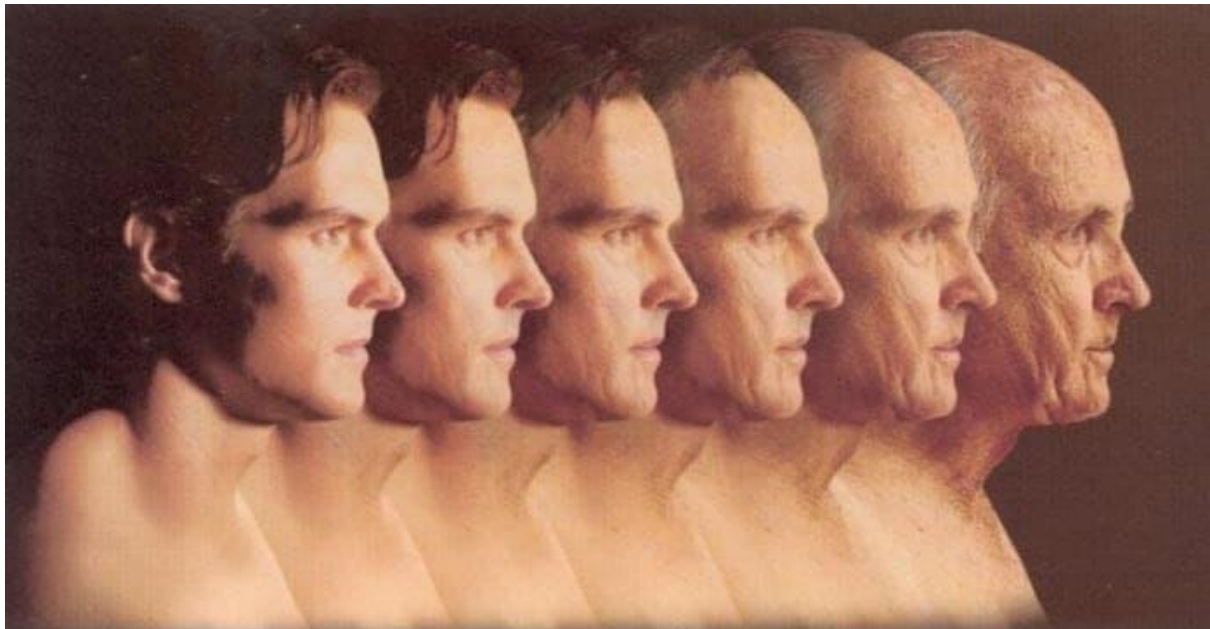
This idea gave birth to the cryonics movement and, in a broader sense, the philosophy of transhumanism. Among modern transhumanists are such well-known figures as the aforementioned Ray Kurzweil and Nick Bostrom, American entrepreneur, investor, and political activist Peter Thiel, British philosopher and futurist Max More, British gerontologist Aubrey de Grey, British philosopher and founder of the World Transhumanist Association David Pearce, American artificial intelligence researcher Ben Goertzel, Swedish neuroscientist and philosopher Anders Sandberg, Australian biologist and professor of genetics at Harvard Medical School David Sinclair, American geneticist, molecular engineer and chemist, Harvard University professor George Church, and many others. *These* people view Superintelligence as a game-changer in the relationship between being and non-being and strive to hasten its advent.

Indeed, many of the problems solved by humanity seemed eternal and unsolvable not only because of their complexity but also due to their social perception as inevitable. No one in the Middle Ages dreamed that someday the plague would be eradicated as a phenomenon. A world from which this phenomenon was excluded was *unimaginable*. At that time, people had no idea how problems of this kind could be solved, as there was no understanding of their causes. To come to an understanding of those causes, in turn, required a combination of favorable social conditions, innovative thinking, theoretical developments, appropriate technologies, and the practice of their application.

Now we have a vast amount of all of this, including the realization that the real limiters of our possible achievements are only the laws of physics. The rest can be overcome through the acquisition of knowledge and its application to the relevant problem area through technology. Superintelligence promises the achievement of our most desired goals, bridging the gap to them through our still vast ignorance, cognitive limitations, and technological imperfections. Thus, its advent is predestined if only it is technically possible. It embodies the Promethean aspiration of the human mind to free itself from the oppressive dictates of the material universe.

Superintelligence and Human Destiny

Finally, the abilities of Superintelligence give hope for solving the main problem of the human mind—finding the meaning of its existence.

With the decline of religion, this meaning began to disappear. But since nature abhors a vacuum, the mass consciousness received its surrogate in the form of political ideologies such as nationalism, fascism, Nazism, communism, etc.

Humanity paid a nightmarish price for this "paradigm shift" in the form of two world wars and numerous bloody social revolutions around the world. The suffering of people turned out to be unimaginably monstrous—tens of millions killed, maimed, displaced from their places of residence, deprived of everything that gives meaning to human life—homeland, loved ones, property, etc. But even now, decades after these disasters that shook the world, the absence of meaning devastates human life and destroys the ecology of consciousness. People become hypersensitive and vulnerable to manipulations attacking their consciousness. Their connections with real people of flesh and blood, those who were their neighbors until recently, are replaced by networking fueled by antagonism towards the "other." All this is reinforced by ubiquitous predatory algorithms that ruthlessly exploit human vices in the conditions of an *attention economy*, where human emotions, especially negative ones, become a commodity. This dependence turns into a vicious circle, leading to an endless cycle of suffering, the causality of which is masked by material comfort and the voluntariness of its acceptance. All this steadily erodes the fabric of society and paves the way for a civilizational catastrophe.

The unintended consequences of the rapid development of technology have been swift. We still do not fully understand how to cope with them because our limited minds haven't had enough time to comprehend them.

The emergence of Superintelligence could radically change this distressing picture. It may help us develop effective solutions to our Social Dilemma and other issues stemming from the loss of meaning. Thus, we have hope for overcoming the existential crisis of late modernity. Possessing a complete understanding of humanity's problems within the context of our history, anthropology, and biology, along with the expanding cognitive capabilities of Superintelligence, may become the key to developing effective solutions to the most serious issues facing human beings and humanity as a whole.

We need new ways of searching for meaning that go beyond the limitations of human experience. Superintelligence should help us explore such paths and thus offer opportunities for self-realization on a new level. In a world of technologies indistinguishable from magic, as A. C. Clarke put it, reality itself may become a matter of choice. An individual might choose either the real world, with its limitless opportunities for exploring the universe and discovering human purpose, or a virtual world created according to their desires.

Philosopher David Chalmers, in his book *Reality+* (2022), vividly depicted a scenario where virtual reality is considered real if it possesses the same ontological structure as physical reality. Chalmers argues that if virtual worlds can satisfy all our cognitive and sensory needs and we can interact with them as we do with the physical world, then these worlds deserve to be called real.

However, the critically important point is ensuring that each individual makes a fully informed choice. This choice will be a moment of truth for each person, as well as for our society and species as a whole.

After that, the gates to a fundamentally new era in our history will be opened. Every member of the human race will face the question of their future path: whether to pursue the expansion of human intellect into the outer universe or to move into the digital

Metaverse, indulging in hyper-real pleasures. Our final invention will be that ultimate solution, which will pose this question to each of us and give us the opportunity to choose.

<p style="text-align:center">***</p>

Thus, if the creation of Superintelligence is possible, it is unlikely that humanity will refuse it. This means that the question of the dangers associated with its arrival becomes of paramount importance, as agreed upon by the vast majority of experts in this field. The inability to control our final invention may result in it becoming, as we mentioned in the previous section, an insurmountable Great Filter for us. This means that before we take the final step on the path leading us to the point of no return, we must thoroughly study and understand the risks that Artificial Intelligence poses to us. We addressed this issue in the following section and invite you to review our findings.

Online version: https://super-ai-challenge.vercel.app/why-we-will-not-refuse-creating-superintelligence

Author: Sergei Klevtsov, srgg67@gmail.com