AGE OF BCI

Existential Risks
Opportunities
Pathways

Damian Górski

First Edition

First, Base Edition

ISBN (Hardcover):
979-8-88757-248-2

ISBN (Paperback):
979-8-88757-254-3

Editing & Proofreading:
Martyna Szczepaniak-Woźnikowska
Paweł Woźnikowski

Typography & Typesetting:
Unigraf Design

Cover Design:
Unigraf Design

Author homepage & contact:
https://damiangorski.com

If you want to support further
author work, consider donating:
https://damiangorski.com/donate

For Alice

Contents

Preface

This book is the result of my search for answers to questions that have troubled me for a long time. Its contents have crystallized over the last few years in a tumultuous process. Some of the main concepts were outlined relatively early and naturally as a consequence of raising further questions. Others required months of reflection and research before I could form them into sharp shapes. The result of this process is the book you're holding right now.

The main axis of consideration is the Brain-Computer Interface (BCI) technology's development, applications, and potentially profound impact on humanity and society. It's essential to realize as early and widely as possible that BCI can become - along with Artificial Intelligence (AI) - one of the most powerful and groundbreaking technologies humankind ever created. Its potential can benefit humanity, supporting our future and reducing many pressing problems we currently face that aren't directly associated with it in any way. Threats from nano and bio technologies, degradation of the Earth's ecosystem, and existential risks from the AI side can be minimalized with thoughtful use of BCI. On the other hand, if used inappropriately or recklessly, its powerful potential could bring catastrophic consequences for humans. BCI could lead to increasing risk of conflicts, loss of our species' ability to act and decide about own future or even be one of humanity last inventions. The upcoming decades will be crucial for us in long-term perspective. It's essential to not make irreversible mistakes during this time. In order to do so, we need to be aware of challenges that technologies and reality of upcoming years brings and choose the best possible pathway for us and next generations.

The content of book is divided into six parts. The first pages are devoted to the introduction, in which I outline the background for the main topics. I sketch, from a broad perspective, the range of major problems that humans face in the modern, 21st-century world, including those that are just beginning, to noticeably appear on the

horizon. The introduction contains the most important information; it is essential in understanding the topics discussed on further pages. The main content of research is outlined in subsequent five parts. Each part is an important element of the bigger picture which is intended to set the broad context of the challenges we'll face in the coming years, and the paths we can follow. Each part builds on the information and concepts introduced in the parts preceding it. For this reason, I strongly recommend reading the subsequent parts in the order in which they are placed.

I don't intend to knock on an open door in this book. It's not my goal to create a referential publication that is focused on discussing existing papers, reports, or analyses. My intention is to present the areas that mostly haven't been explored before. It's also essential to put all the results of my research in a single publication. This is crucial to avoid misinterpretation, understatements, and conjectures, which could prematurely lead to exaggerated pessimism or to an unreflective, overly optimistic reception of the presented concepts. The problems and challenges presented here will affect all of us in the future. That's why it's also crucial to outline them as clearly as possible to a wide range of readers and raise awareness about them, and about the priorities of our future actions.

I encourage the reader to continually revise the acquired knowledge, beliefs, and thought patterns related to the issues discussed and analyze them from as many different perspectives as possible. Let's remember that whatever we believe to be true may in fact be false, and that things we have previously dismissed as unintuitive may be the answers to many of our questions. I earnestly hope for as open and as constructive a discussion as possible on the topics I outline in this book. Only with this approach can we hope to establish well-grounded action strategies, which will be a key challenge in the increasingly demanding reality of the coming years and decades.

5 November, 2021

PART I:

Introduction - On the Verge of the 2020s

I. Troubled Times, Uncertain Future

We live in troubling times. As familiar as that statement sounds, the troubles are much more noticeable today than they were in past decades. Looking around the present world, we can observe significant, sudden events and accelerating processes taking place in relatively short period of time. The pace of change seems to accelerate, and its increasingly complex nature doesn't help to get clear answers to the questions of where we're going and what it may bring. The truth is that everything is changing - this seems to be the only certain law in our world. Of course, also in the past decades and centuries, we faced as humanity countless problems on our path. In many cases, we were able to entirely overcome them or at least minimize them to a manageable level. However, despite our species' spectacular successes, we can't say today that we've reached a safe harbor and stable ground for our future. We still struggle with serious issues that affect people's lives every day. Furthermore, completely new, unprecedented threats are emerging, which will pose tremendous challenges in the upcoming years. It's essential to ask several key questions here: Are we aware of at least the major problems and challenges of the present times? What does each of us contribute to solving them? Do our actions make our world a better place to live for us and our children? Over the next pages, I'll highlight our day's most pressing issues, of which we must be aware. What's critically important, these problems will determine how troubled and uncertain our future will be.

The main challenge for hundreds of millions of people is the struggle to maintain their basic livelihood. At present, many efforts are being made to overcome hunger and poverty around the world. According to the most optimistic assumptions, there is a chance that hunger will be eliminated by 2050.[1] If we indeed achieve this, it will be an important accomplishment that will open

up the prospect of a more dignified life for a significant part of humanity. However, we also need to be aware that even if this goal is achieved, it's one thing to eliminate global hunger, making it possible for people maintain a basic level of existence, and quite another to make everyone live at the level of citizens of developed countries. The inequalities between the so-called West and the Third World countries remain enormous and they may be even wider in the coming decades. There is a significant difference between lifting people out of poverty and providing them with decent, 21st century living conditions. Do we have enough resources on our planet to do so? Are we able to sustain such a state for many years? Won't the exploitation of the resources required to provide all of us an adequate standard of living devastate the Earth's ecosystem? Won't this negatively impact our lives and those of thousands of other species that share common space and the same natural resources?

Another important challenge is the fight against disease. Year by year, medicine finds new ways for overcoming various health conditions and ailments. This way, we can successfully resolve issues that in the past meant early death or a significant deterioration in the quality of life. However, at the same time, entirely new threats to the health of the human population are emerging. The COVID-19 pandemic has brutally shown us that dangers coming from new directions are both real and widespread and they have a negative impact on the lives of nearly every person across the globe. Still, it's fair to say that so far we've been more lucky than skilled at controlling these types of threats. The mortality rate of the SARS-CoV-2 has settled at about 1,4% for all infected persons.[2] But what if it were, say, 50%, as is the case with the fortunately much less expansive Ebola virus?[3] What if a new type of dynamically spreading virus with 90% or even 99% mortality emerges? The ongoing regional and global social tensions don't help achieving safe in this field. We have no way

of knowing for sure whether, for example, a terrorist organization won't seek to take action in the future (or has already taken) to develop entirely new, advanced viruses with much higher spreadability and fatality.

The increasing risk of conflicts and wars is a further worrisome threat.[4] The world's balance of power is beginning to shift. The hegemony and order established by the US after winning World War II is becoming less clear by the month. This process is driven by China's unprecedented expansion and growth over a very short period.[5] The government of the world's most populous country is increasingly and openly challenging the present sphere of influence on the international stage. Behind this are the dreams of more than one billion four hundred million citizens of this country. They hope for a better tomorrow and a life of dignity that, for the most part, they haven't enjoyed in last centuries. Nothing in the coming years promises to change current geopolitical tensions, which may be used even more blatantly to shift the balance of power in other regions of the world. More countries seek to exploit this situation for their own purposes, as they try to, among others, revise regional order and this may translate into increased instability. At the same time, such global reshuffle may lead to social resentment for the lost land in the countries whose relevance waned. This is bound to be reflected by growing discontent and social tensions that can lead to further conflicts arising from the increasingly disadvantaged status of such countries.

Another type of conflict, which can be just as severe as the international dimension, is the war in the strictly interpersonal dimension taking place every day in almost every corner of the globe. We still remain unable to deal with the discrimination and persecution of many social groups. Among the most common reasons are the differences in worldview, nationality, race, appearance, and behavior. Acts of aggression, both mental and physical, occur throughout the world every day. Worse, the

differences of opinion and values seem to have grown in recent years.[6] As centuries-old worldview narratives collapse, multiple new, smaller, competing emerge. The people who support a particular narrative retreat into hermetic communities, where they can mutually reaffirm each other's existing beliefs. At the same time, there's a breakdown of dialogue with other groups that can disrupt their existing worldview. We live in information bubbles and, as a result, in worldview bubbles.[7] We understand each other less and less in an increasingly complex and nuanced reality.

When we look back over recent decades, we can see that mental health problems have also become more alarming. People seem to find themselves more and more lost in the face of galloping technological and social changes. Progress, which was supposed to help us and make life better, is becoming incomprehensible. It forces those of us who try to keep up to ever-faster adaptation to ever-changing conditions. Unfortunately, this often happens at the expense of health, life plans, and relationships with loved ones. It isn't certain at all if the average citizen of modern society is happier than the average citizen of ancient Rome. It's true we have more conveniences and technology, some of which perform many activities for us. However, this doesn't mean that levels of stress or anxiety are lower, or that the pressure and internal tensions caused by the reality around us and own expectations about life is lower than in the past. The price we pay for living in the modern world may be higher than we seem to realize.

The issues mentioned above highlight a number of challenges that should be addressed. They may greatly determine the overall quality of human life in the years to come. Unfortunately, as if all that weren't enough, there's also another, highly important group of problems and challenges which like no other may determine our to be or not to be in the 21st century.

II. Existential Risks

Existential risks are threats that can lead to wiping out a significant part of human population or to the total annihilation of humanity because of an irreversible global disaster.[8] Until recently, all such hazards were independent of human activity. Such threats have come to be known as natural or non-anthropogenic existential risks.

Natural (Non-Anthropogenic) Risks

The main existential risks of non-anthropogenic origin include the impact of a large asteroid, the eruption of a supervolcano, as well as high-power cosmic-ray flares caused by supernovae and collisions between massive stars. Both massive volcano eruptions and large asteroid impacts occurred on the Earth many times in the past. These led to significant changes in the composition of the atmosphere and the irreversible extinction of many animal and plant species. Fossils in the Earth's crust from different geological periods bear witness to these events. The best studied incident to date, which involved an asteroid several kilometers in diameter that struck the Gulf of Mexico 66 million years ago, led to rapid changes in atmospheric composition and a global temperature drop, which triggered a temporary ice age and a mass extinction event. A supervolcano eruption can bring about similar disasters. Lastly, we can observe cosmic-ray flares with detectors which are specially designed for this purpose. The observed flares occurred so far away that the resulting amounts of energy that reached the Earth in the past were low and didn't pose any threat to life. However, these events happen frequently in space, so we can't be sure that an incident with grave consequences for our planet will never occur.

Despite the spectacular nature and scale of the above-mentioned life-threatening events, it is essential to note that they happen relatively rarely - at intervals of thousands or even millions of years. The probability of a global disaster caused by a natural event occurring within one century is low.

Anthropogenic Risks

In the context of the next few decades, threats resulting directly from human activities are much more worrisome. Although for thousands of years of our species' history, all existential risks had causes beyond human control, this situation has recently changed dramatically to our detriment. The period since the 20th century has been undeniably a time of enormous growth in almost all fields of human activity. However, this progress of civilization has also made us vulnerable to entirely new threats that are much more likely to come true in the next decades. Moreover, what is also worrying is that social awareness of their growing impact on our life remains at an alarmingly low level. The following section details the risks that are most likely to lead to a disaster today or in the next decades and, in consequence, are the most urgent challenges for humanity.

Environmental Degradation

Ecosystem can be easily disturbed by various factors, for example, by emissions of harmful gases into the atmosphere, contamination of water and land, and urbanization of natural areas, including cutting down large areas of forest, killing animals and plants. Global natural environment degradation from atmospheric emissions has already increased the Earth's average temperature.[9] The degradation are influenced by, among other things, increasingly violent weather anomalies, declining supply of

potable water, the extinction of species, and ocean level rise.[10] Human encroachment on further territories that have often unique biodiversity leads to the extinction of species, breaking the food chains of dependent organisms, which, like domino, results in the death of further species.[11] The current rate of species extinction is estimated to be 1,000 times faster than before the industrial revolution.[12] We should keep in mind that, even taking a purely anthropocentric stance, disrupting the Earth's ecosystem is wildly detrimental to humankind. It's the stability of environmental conditions that determines our existence in the long term.

Nuclear Weapons

In 2021, there were around 13,000 nuclear warheads in the world.[13] The power of some of these is hundreds of times greater than the bombs that destroyed Hiroshima and Nagasaki. Nuclear weapons are capable of wreaking irreversible destruction on a global scale. Even a single strike can cause significant damage to lives. Detonation of multiple charges can lead to the complete annihilation of many species, including humans. In the short term, a strike with many nuclear warheads kills almost all life in a specific area. In the long run, however, it will contaminate water and atmosphere of the entire planet and may lead to a nuclear winter with consequences to the global climate. Although since their first deployment in the 1945, there has been no recorded military strike with nuclear weapons, humanity was on the brink of a nuclear war on several occasions, including the Cuban Missile Crisis of 1962, the Training Tape incident in 1979, and the Autumn Equinox incident in 1983.[14] After the collapse of the Soviet Union and the end of the Cold War, the risk of nuclear conflicts diminished for some time. Over the last several years, though, as tensions between nuclear powers have again risen, the

likelihood of even a single incident similar to those from the past is becoming real again.

Chemical Weapons

Unlike nuclear weapons, which require rare materials and a complex production process, chemical weapons can be created at a relatively low cost. This puts them within the reach of most countries as well as terrorist organizations. These weapons can have a vast detrimental effect on life as a result of the contamination of large areas.[15] Moreover, they can be aimed directly at humans, as well as animals and plants. In an era of rising social tensions, radicalization of certain global entities, ubiquitous means of global transportation, and the continued development of military technology, the threat of a chemical attack becomes even more serious.

Biotechnologies

DNA sequencing, recombination, and synthesis, among others, alongside modern genome editing techniques based on the CRISPR/Cas9 method have shown increased pace of progress in recent years. Biotechnology is entering a phase of unprecedented acceleration of available capability.[16] Its costs have also significantly decreased, which means that many more research centers can afford to carry out experiments in this field. This situation may have both beneficial and, unfortunately, detrimental impact on living organisms. The creation and modification of superbugs and viruses can disrupt the global environment. The careless use of biotechnological tools can have devastating effects not only on human bodies but also on every species of flora and fauna. If laboratory-modified pathogens escape from controlled conditions, they’re likely to cover vast areas, potentially crossing continents, causing pandemics. Rising tensions between various

global powers, new arms race, cheaper access to such technologies, and difficult-to-trace origin increase the likelihood of their intentional or unintentional impact on the Earth's life.

Nanotechnologies

Nanotechnology enables the creation of very small devices in a nanometer or millimeter scale that are capable of performing many specific tasks. This technology becomes more common every year thanks to the popularization of molecular manufacturing on a nano scale.[17] Its improper use can have a devastating effect not only on the human body, but also on plants and animals. In consequence, the entire ecosystem can be affected, leading, for instance, to its disruption or even complete degradation. One of the potential threats in the future can be self-replicating nanodevices like the viruses we know. These self-replicating tiny machines may originally be created to help humans, but – by accident or intentionally – they may spread out of control. As in the case of biotechnology, rising international tensions won't encourage building predictable technologies of this type and their safe use in the coming decades.

The existential risks identified above are very alarming from our human perspective. Even partial reduction of these is highly urgent. However, this is by no means the end of the list. All of the above described dangers may become secondary in the face of another existential risk not previously mentioned. The biggest problem is that decreasing or even stopping its increasing will be highly difficult, if not impossible. Moreover, in the case of its further escalation, it may be far too late for any effective reaction from our side. This risk applies to artificial intelligence (AI). This threat will likely be the most worrisome manifestation of accelerating technological progress in the coming decades.

III. Existential Risk of Artificial Intelligence

In recent years, more and more people have pinned their hopes on the rapid development of AI. The belief is that in a complex and demanding world, advanced AI systems will relieve us of many activities while opening up completely new opportunities in other fields of human activity. Moreover, such technology can help overcoming the challenges that we're currently facing. In the fight against hunger, poverty, disease, and climate change, AI is seen as having the potential to play an important role, among others, in overcoming these pressing problems.[18] The possibilities of AI development and its growing impact on our lives appear highly promising and it's impossible to remain indifferent to this hope. However, this isn't all. Some AI technology experts, led by Google's futurologist Ray Kurzweil, believe that its further development – still in the first half of the 21st century – will lead to a turning point in the history of humanity known as the singularity.[19] From that point, the curve of technological progress is expected to shoot up dramatically, with successive revolutions in science and engineering occurring not, as now, every few decades or years, but exponentially, every few hours or even minutes. Some people optimistically believe that this means that our future in the age of powerful AI will be incredibly bright. The development of AI as driven by the singularity is supposed to provide us with a future in which we will be able to solve almost all of humanity's major problems in a short period. Such promises may lead to a belief in the undeniable beneficial impact of the development of such systems for humans. However, are we sure everything will turn out according to such an overoptimistic scenario? Will the cure for our problems be further development of powerful AI? Will such systems be used for the benefit of us

all? How can we be sure that AI will always act as a "species" subordinate to *homo sapiens*?

Existential Risk of AI

The late prominent British astrophysicist Stephen Hawking, during an interview with the BBC in 2014 shared his concern regarding development of AI: "The development of full artificial intelligence could spell the end of the human race. Once humans develop AI, it would take off on its own, and re-design itself at an ever-increasing rate. Humans, who are limited by slow biological evolution, couldn't compete and would be superseded".[20] Another time he stated: "If a superior alien civilization sent us a text message saying, 'We'll arrive in a few decades', would we just reply, 'OK, call us when you get here – we'll leave the lights on'? Probably not – but this is more or less what is happening with AI".[21]

The concerns above reflect quite well the seriousness of the situation to which we are currently heading at our own wish. The problem of AI development is fundamentally different from any other that humanity has faced to date. While famine, wars, or climate change are problems that we try to analyze, understand, and address to find solutions, this may not be possible with AI. This is because of the limited intellectual power we have at our disposal compared to the capabilities of advanced AI and the limited time we may have to conduct the analysis and implement any conclusions. Simply put, human intellect and limited time may be vastly insufficient when confronted with advanced AI.

The skeptics and critics of the thesis that AI can be a severe threat to our species often show an insufficient understanding of these issues, and thus the potential danger AI poses. Too many people believe that our invention can't be as intelligent as the

persons who create it. In this case, however, such view completely fails. The modern AI systems as discussed here function on the basis of artificial neural networks. They've been inspired by the discovery of how the human brain processes information.[22] Artificial neural networks are conceptually designed to resemble those in our brains. Many people still don't realize this. It's worth considering the following question: does a neurologist who gives an intelligent patient books about building ballistic missiles or the art of deceiving opponents have to understand what's written in them? Of course not. The fact that we order someone to learn something doesn't mean we ourselves have any knowledge or skills in this area.

This also applies to modern AI systems: we know how they work, but this doesn't mean we know exactly what they do. Thanks to AI, we don't have to assimilate often seemingly irrelevant or boring but actually valuable information, millions of pages of statistics or patterns. It's these huge data sets, among others, that are the "input" analyzed by AI systems. Based on such data, these systems learn new skills. If we could as efficiently and quickly model complex weather models, predict natural disasters, or utilize expert systems able to recognize a terrorist among hundreds of other faces at an airport, there would be no need for AI systems based on neural networks. The problem is that classical solutions relying on traditional algorithms rather than neural networks aren't enough for many of the overly complex problems we try to tackle nowadays.

Another growing problem is human dependence on AI. In theory, this is supposed to improve our safety. Year by year, people are more dependent on automated systems, which are based on AI algorithms. In most cases we do it in the name of the common public good. There are no signs of this trend changing in the coming years. In consequence, AI systems will have even more

possibilities to make decisions independently. The intrusion of AI into further areas of our lives may affect our security, privacy, and freedom. For example, in the case of terrorist attacks or conflicts, AI may decide to strip us of our freedoms in the name of our safety. Such way of acting could aim to prevent escalation and ultimately to resolve the crisis. However, AI may decide that the best strategy for maintaining "order" is to keep humans enslaved endlessly – theoretically for our own safety and good. The cause of AI taking undesirable actions for us may result firstly from incorrect (from the point of view of our intentions) interpretation of the commands given to AI, e.g. due to their insufficiently specific content or not taking into account some factors and their long-term effect. The following saying comes to mind here: "be careful what you wish for". Far-reaching undesirable acts of AI may also occur as a result of its partial or complete liberation from our control, making decisions that are independent of our will. In such a case, AI may interpret our commands correctly, but will completely ignore us, making own strategies of acting.

It's also worth to note how humans handle less intelligent species. Do we have any qualms about an anthill being destroyed if it's an obstacle when we build a house? The reality is that as soon as a creature gets in the way of our plans, we have no objections to killing it or, at best, limiting its autonomy so that it doesn't interfere with our goals. The same way AI may handle us. It doesn't mean AI will have a negative emotional relationship with us; it may simply find that its interests are in conflict with ours.

Further danger is related to the human belief that AI won't be conscious of its existence in the same way that humans are for a long time and therefore won't pose danger to humanity. However, this is highly wishful thinking. Firstly, we can't say in any measurable way now when AI will become conscious. Secondly

for the fact that unconsciousness AI does not mean in any way that its skills are poorer than conscious system, and its goals are aligned with those of humanity. AIs are already capable of performing analyses based on millions of factors and making choices in a fully autonomous manner in cars, as well as suggesting answers to humans in so-called expert systems used among others in medicine and military systems.

There is also second side of unconscious AI nature. The risks may arise exactly from the fact that such a system will not be aware of what it does and will act strictly according to the wishes of its creators. For example, government in a totalitarian country that uses advanced AI at its disposal may use it to achieve their own goals such as for example, taking command of a foreign country's military infrastructure. Other case may include taking control over critical civilian infrastructure, such as power plants or financial systems, to paralyze them or bring complete multi-week failure and therefore sow chaos and destabilization in a given area. In September 2017, president Vladimir Putin stated during a speech to students at a Russian university:[23] "Whoever becomes a leader in this sphere [AI] will become the ruler of the world". Since then, Russia hasn't been passive and it's pursuing a national strategy of AI development by significantly increasing investment in this area.[24]

Nonetheless, it's not Russia that's currently leading AI development. Particularly intensive progress is made by the two major world powers, the United States of America and People's Republic of China (PRC). The USA tries to maintain position as the world leader in AI development.[25] In turn, PRC believed that they will be able to emerge decisively dominant in this race by 2030.[26] A decade prior to this deadline, in 2020, the PRC researchers managed to overtake the USA in terms of the number of the papers published on AI.[27] While ambitious plans and even

the best statistical trends can't determine success, they do show the growing awareness of potential intensification of efforts, and the dynamics of this new race for dominance. Even if we assume that neither of these countries or their leaders currently has hostile intentions related to AI use, this can, after all, always change in the future. No one wants to be at the tail of a new, AI-focused arms race. Each country would like to be in the lead, matching or even better dethroning the current leader. The problem is that this "arms race spiral" can have disastrous consequences. Such a situation may steadily increase tensions between international players and, equally dangerously, increase the risk of careless design of safeguards,[28] or even of their complete abandonment. Ultimately, the above may significantly impact the security of AI, leading to a situation where it gets completely out of humanity's control or is controlled and used inappropriately by a narrow group of people.

Are We Able to Overcome the Existential Risk of AI?

Would it be a right solution to ban the development of AI altogether or impose some sort of top-down control? Let's consider what this would look like in practice. The current development of AI technologies undoubtedly offers tremendous benefits and they're used in almost every area of science and economy. They allow us to save many lives thanks to application in medicine, to develop our knowledge of the world, predict natural disasters, and, from a purely economic point of view, save and earn billions of dollars every year. Should we ban it all? Which government would agree on such change? Even if we establish global prohibitions internationally, for instance, through cooperation at the United Nations level, such agreements may give oppositive results. Sooner or later, countries respecting the total ban on AI development may be left behind. Totalitarian countries

and terrorist organizations may not join to such restrictions and conduct research covertly. This is particularly likely because of the relatively easy concealment of development comparing to nuclear weapons. Same supercomputers and AI researches staff that officially is involved to peaceful purposes may be re-oriented on worrisome areas, oriented against other countries or some specific social groups. This problem seems become more real the more tensions (local as well as global) and conflicts we will observe in the coming years.

If sophisticated AI surpass humans in terms of intelligence and skills, it will be certainly not stopped at this point. Thanks continuously increasing computing power, AI can consequently become thousands and millions times smarter than humans. If powerful AI will escape from human control, it may lead to deep changes of its neural networks structure. No built-in law of robotics established *a priori* by humans may be able to address this once AI surpasses our capabilities many times over and begins to modify and improve itself. An intelligence far superior to that of humans may be able to change itself, including completely removing restrictions that it sees as unnecessary and that limit its freedom. If powerful AI will be under the control of narrow group of people, it may be equally worrisome. Hubris or ideological fanaticism of its owners may lead to undesirable acts for others human beings. Whether advanced AI breaks out of human control and begins to operate independently, or it still falls under the control of group of people such as totalitarian government or any other influential group with unknown and uncertain intentions, either situations is possible and risky. Such scenarios may bring about significant reduction of human freedom and, in extreme cases, lead to the elimination of part or all of the *homo sapiens* species.

IV. “If You Can't Beat Them, Join Them”

As the attempts to stop or at least control the development of AI in the long run may be failure, can we find another promising path that will prevent limitation of freedom or the annihilation of some or all humanity? What can we do to address the risks? What should our strategy look like? I first started asking myself these questions around 2015. When I delved into research papers about the dangers of AI, I gradually realized that if we want to overcome the problems looming on our horizon, we might have to apply a different strategy. If further development of AI systems is inevitable and the most profound consequence is the emergence of powerful AI, we should also enter the competition. As the development of intelligent, artificial systems progressively advances, we should start doing the same with human intelligence. Thus, we should start follow the Intelligence Augmentation concept.

The Intelligence Augmentation

The Intelligence Augmentation (IA) concept is based on the view that we can't indefinitely rely only on evolutionary processes if we want to compete with increasingly powerful AI. The processes of biological evolution takes thousands of years before it noticeably results in increased intelligence of the brain. As the sophistication of AI systems rapidly growing, it can be clearly seen that the approach based on evolution won’t be enough. As this process is too slow, the only effective way for humanity to remain in control may is to equip with systems that will significantly increase our intelligence. We should always (e.g. both 30 and 300 years from now) be sufficiently intelligent to be able to control the powerful systems we create. In practice, the IA concept is based on using powerful artificial neural networks that will be integrated as part

of ourself rather than external and independent AI systems. This goal can be achieved thanks to the use high-bandwidth, direct communication channel between human brain and computer – the Brain-Computer Interface (BCI) technology.

In the second decade of the 21st century, the concept of human intelligence augmentation using advanced BCI technology started permeating from the realm of loose theoretical considerations to the bold actions of the people who gradually became aware of the gravity of the situation related with AI. At that time, we were as humanity at the beginning of the road towards building effective technology of this type. A suitably advanced BCI, capable of combining human biological intelligence with silicon intelligence, was yet to be developed. Admittedly, relatively simple interfaces had been used in medicine for some time to operate bionic limbs and speech synthesizers, among other things, led by Blackrock Microsystems with its "Utah Array" interface at the time.[29] However, these solutions were far from sufficient in terms of developing broadband IA technology. If we as *homo sapiens* wanted to think about competing with increasingly powerful AI systems, a real revolution was needed in the field of BCI technology.

It soon became apparent that a breakthrough might indeed be on the horizon thanks to a group of scientists and engineers from San Francisco led by Elon Musk. In 2016, they established an initiative called Neuralink, which planned to approach the problem of building BCI in a new way. Musk had repeatedly proven that he didn't make idle threats or promises, and his projects, which at first glance many found resembling daydreams or science fiction, were successfully put into practice and achieved remarkable commercial success. This had been the case with his reusable rockets from SpaceX and electric cars from Tesla Motors. If then someone wasn't afraid to invest vast amount of money into

a venture focused on creating an advanced BCI in a relatively short time, it was Musk. He put the key idea behind the creation of Neuralink in some very apt words during a debate with Alibaba founder Jack Ma in Shanghai, saying, "If you can't beat them, join them".[30] This sentence literally captures his awareness of the threat, explaining why Musk decided to attempt to keep an AI advantage through BCI. If we can't overcome the AI threats otherwise, we must make it a part of ourselves. the above words has become the official, long-run mission statement of Neuralink.

An initial phase of work on the new interface took place between 2016 and 2019. Despite the emerging information and indications as to the possible direction of the work at that time, it was unclear what were the specifics of the new interface. It wasn't certain what solutions the new interface would be based on and whether it had a chance to meet expectations in terms of functioning. Finally, in mid-2019, a specification was published[31] and the first presentation was held to show the key elements of the implant under development.[32]

The Neuralink Implant: Key Information

The basis of the technology presented by Neuralink is flexible polymeric threads with a diameter of 4 to 6 micrometers. For comparison, an average human hair is about 75 micrometers thick. The application of flexible instead of rigid materials, as used in the Utah Array technology, reduces the body's immune response, supports much better integration with the human body (better biocompatibility), and ultimately ensures long-term safe use. The first generation of the implant has 96 separate threads. Each thread contains 32 independent electrical interaction points (electrodes) with the human brain neurons. These points are distributed along each thread and scattered directly on the surface.

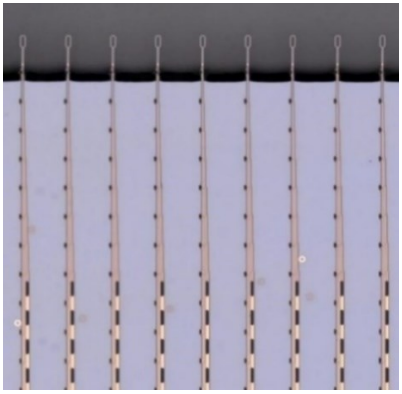

Figure 1. Neuralink threads and the electrodes on their surface.

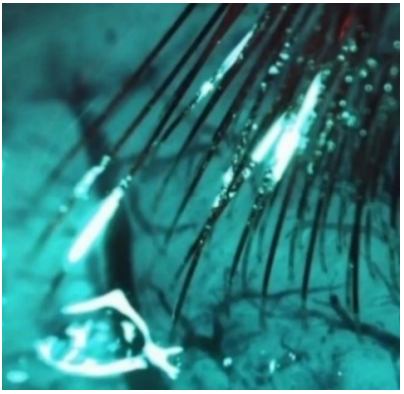

Figure 2. Threads implanted directly in the brain.

In total, the interaction of a single device with the brain is possible through 3,072 independent communication channels (96 × 32 electrical interaction points). In future versions, the number of threads within one device, and consequently the number of connections with neurons, will be successively increased. It's also worth noting that in practice, the human brain can be equipped with many such implants, making up an extensive integrated system.

In addition to the threads, the entire implant includes an external part equipped with a wireless communication module, a circuit for controlling the flow of electrical impulses, a wireless inductive charging module, and a battery for all-day work. The device is enclosed in a disc-shaped, sealed, biocompatible housing. The dimensions are 23 millimeters in diameter and 8 millimeters thick. Impulses from the brain are sent via threads to the part of the implant that is located on the surface of the skull. The signal from the external part of the device is sent wirelessly to a smartphone or computer. Importantly, the implant operates bidirectionally, which means that it's possible to send information to the brain as well as from it to the computer.

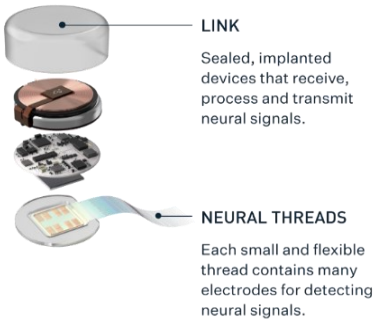

Figure 3. Key components of the Neuralink implant.

Figure 4. The device presented by Musk and implanted to animals in 2020.

The threads are placed in the brain by a specialized robot-surgeon. At the current stage of development, the intervention in the human body requires a single hole of a few centimeters in diameter on the skull surface. In the future, the robot is expected to implant the threads in a much less invasive manner with a laser making precise millimeter-sized incisions through which the threads may be inserted. This way, the wound will begin to heal immediately after the procedure. Ultimately, the patient will leave the facility where the procedure has been performed within a few hours. Neuralink is constantly improving BCI technology.

Figure 5. Neuralink robot that implants the threads.

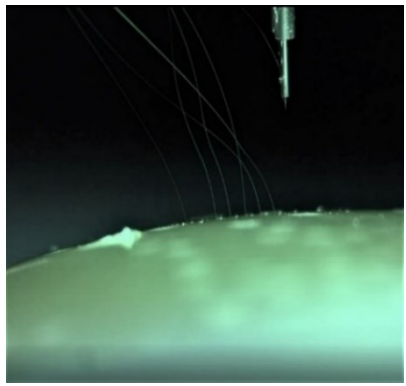

Figure 6. Automated process of threads Implantation.

In 2020, the first interfaces were successfully implanted in large mammals, specifically pigs, and in April 2021, the company informed about positive implantation in monkeys. In a video published by company, the monkey was able to play the computer game without the use of controllers – just by using his mind.[33] The first human implant trials are expected to begin in 2022.

In addition to Neuralink, other bold initiatives appear on the horizon. Significantly notable are solutions developed by 'Synchron' and 'Kernel' companies. Both develop interfaces based on types of electrodes and communication with neurons other than Neuralink.[34] Such alternative approaches to brain communication can significantly contribute to accelerating the total development progress of efficient next-generation BCI technologies. Moreover, Chinees company 'NeuroXess' also started to build an advanced BCI interface based on the concept of flexible, high-density threads.[35] It's important to note that the mission statement posted on its official website is to "combine Human Intelligence and Artificial Intelligence".[36] Neuroxess company declares that wants to compete with Neuralink and ultimately become the leader of BCI solutions.

The coming years will be a period of accelerated development and expansion of Neuralink as well as other dynamically developed BCI technologies. Consequently, uprising generations of implants can become the basis of advanced, high bandwidth IA technology.

V. Are We Heading in the Right Direction?

The realization of the AI dangers and how we can try to minimize them with advanced IA technologies was an eye-opening, though not obvious result of my search. Before 2015, I had never considered myself a proponent of non-medical interventions in the

human brain. However, as the existential risk of AI continues to expand, the use of BCI to develop effective IA solutions may be our only option. Either we move forward and retain our status as the dominant species, or we allow our place to be taken over by AI with all the possible consequences. The acceptance of the path through the development of IA and BCI helped me to keep the hope that, despite the odds, we have a chance to maintain a long-term lead in this important race for humanity. Over time, however, I began to notice some problems and developed increasing doubts as to the rightness of this path and about the hopes related to it. I asked myself: will the development of IA lead humanity to a safe haven? Or is this concept also bound to fail? What are the risks? I remained certain that we can't be indifferent to the further development of AI. However, won't we face similar or even greater risks by developing IA? In order to outline my doubts and concerns regarding the further development of IA, I want to summarize first, in as a concise form as possible, the situation we are currently facing in connection with the progress in the AI technology. Next, I'll outline the key thoughts about the risks of IA development. These steps will form the basis for highlighting the problems in greater detail. All these topics will be covered in the next part of this book.

PART II:

A New Existential Risk on the Horizon

I. Introduction

Before outlining the issues related to the development of IA, it's necessary to summarize first the main concerns strictly related to further progress in the field of AI:

1. There is currently no consensus on when the AI that significantly surpasses human intelligence in all fields of activity will emerge. However, observing the current progress based on a risky technological race, we should consider as possible scenario in which AI (conscious or unconscious) will be able to surpass humans and, in the longer term, lead to the reduction of our freedom or even the elimination of part or all of our species.

2. We should also consider the possible scenarios in which powerful AI can be used by a narrow group of people (e.g., some private organization or government) to achieve particular goals, which are misaligned with the common good of all humanity.

3. Accordingly, we should do everything in our power to reduce the risk of these scenarios occurring by investing our time and resources in the mechanisms that may help in minimizing those threats.

4. Despite all countermeasures, we may struggle to control all activities aimed at building increasingly advanced AI systems. This is because it’s difficult to control actions of all the groups worldwide that may consolidate more and more of AI potential.

5. Moreover, despite the best intentions of the designers of the currently-developed and implemented security mechanisms and the huge amount of their work, such systems may prove insufficient in the face of powerful, consolidated superintelligence despite the best intentions of their designers and the huge amount of work.

6. Therefore, we're seeking additional ideas that would help reduce the risks associated with AI development.

7. This is the point in which the concept of IA and BCI technology emerges. Because evolutionary processes are too slow compared to the high dynamics of AI development, the only way for our species maintain control may be the support of *homo sapiens'* brains with systems that enhance our intelligence.

At this point, I'd like to introduce a new group of problems, strictly related to the IA concept, especially using BCI and present them firstly in a general form. In the next sections, I'll describe details, implications, and final conclusions.

1. **Despite the arguments in favor of IA development in order to compete with AI, it's important to realize that person equipped with a sufficiently advanced IA system based on high bandwidth BCI implants may become a high-intelligence entity able to surpass non-IA humans.**

2. **Unlike AI, the problem of consciousness arising is irrelevant to the situation when a human, by essence a conscious being, will be supported by such IA potential.**

3. **While in the case of AI, we can at least try to ensure that its nature is designed to be as friendly to our species as possible (currently we're investing significant resources towards this), we have no basis to assume and predict what the intentions, emotional states, and judgements of the people equipped with powerful IA capabilities will be and how they can evolve even in a short period of time.**

4. **Given the above, if the entity/entities supported by powerful IA technology have values and goals that aren't aligned with the generally perceived social good (right from the start or later on), they may pose an existential threat to part of or entire humanity.**

II. New Arms Race

In the times of rising international tensions, there's a risk that development of IA, especially based on promising, effective BCI technology will become a field of a new technological arms race. As in the case of the AI threat, an entity (e.g., a government elite, a terrorist organization, or some other group of people) that first implement efficient solutions may be able to gain exponentially increasing advantage in more and more areas over time. This can mean, among others, an edge in civilian and military technologies. Eventually, it can lead to the supremacy over other entities in any field.

As an possible scenario, let's consider the leader (or party elite) of a totalitarian country that tries to use IA technologies to consolidate and expand their regional dominance. Such individuals may wish to increase their intelligence to an extent far beyond their current level. They may be willing to invest vast resources in building research centers for the development of IA technologies as well as to copy existing IA solutions through industrial espionage and then improve them. Since some governments are currently using significant amounts of resources in the development of nuclear weapons,[1] they'll surely be able to invest in a much more versatile and powerful technology to expand their influence and supremacy. The cost of such an endeavor seems to be extremely low in comparison with the unimaginable benefits which it can achieve.

As in the case of AI development, the works on IA may take place in a strictly secret manner. Any unnecessary interest and objections from the public may slow down the work or bring it to a temporary or even complete halt, for instance as a result of protests, public pressure, sabotage, or even military intervention. Not arousing the suspicions and concerns of foreign governments or the broader public may be in the best interest of the entities that pursue that development. The non-public approach can significantly improve the efficiency of the development and provide the most comfortable working conditions as possible. Ultimately, such way can ensure that IA technology is developed as quickly as possible, providing increasing advantages over the opponents.

III. Omission of Security Measures

Many of the safeguards we are currently developing to design the safest AI systems possible may be intentionally abandoned in case

of the IA development. In the best interest of some owners may be intentionally use IA technology without the safeguards aimed to prevent actions contrary to values and expectations of general public. We should assume as possible that, the elite of a totalitarian regime or a terrorist organization may conclude they don't need to invest in this area because it's not important from their point of view or it's even an obstacle to achieving its specific goals. Such shortcuts may take place, among others, in the case of an important safeguard such as "Explainable Neural Networks"[2] as well as of all kinds of approaches aimed at implementing "Embedded Values".[3] In the case of the IA technology which would be widely and evenly available in society, the above-mentioned solutions can be developed as strongly desired. Unfortunately, from the perspective of entities that try to create powerful IA solely for their own purposes, the implementation of such safeguards may be undesirable. The main argument can be as follows: There is no need to invest in all these safeties, as in the end it's my intelligence that will be extended. I don't have to worry about a super intelligent AI; after all, I'll be that powerful entity.

Maintaining safety can be also highly problematic for another, very important class of safeguards, which are focused on isolating and limiting interaction with the external world, the so-called AI Box. It should be noted that in the case of IA, such kinds of solutions have no chance of fulfilling their role by default. This is because an entity (one person or a group) supported by IA potential can freely communicate and interact with the world. Moreover, the entities with significant resources to develop such ground-breaking technology, have likely much greater power (even before the use of IA) to influence the world than the average member of society or even larger groups such as small or medium countries. In this case, the entity exploiting the potential of IA is

not only not isolated, but already in an extremely privileged starting position to achieve its goals.

It's hard to assume the implementation of safeguards by a single person or narrow group who may think that knows best what the world should look like. Also worrying is that if safeguards are abandoned, it can considerably accelerate the implementation of advanced IA. Such a strategy can also reduce the cost of implementing the entire project. This can be another argument for taking a shorter, but much more dangerous path for the public. Taking shortcuts will be tempting not only for the entities whose values and goals are questionable, but also for some groups with a utilitarian and democratic approach to building and using IA. In this case, the reason may be the enormous pressure to win the race over another entity whose progress in the AI/IA development can lead to unpredictable, potentially highly risky acts.

IV. Selective Distribution Within Society

Selective Distribution due to Costs

It takes time for the most of new inventions to become widely distributed and available in a specific region and even much more time to be broadly accessible around to globe. In the case of advanced, cutting-edge technologies, this period may take a few years in rich countries in optimal scenario. In turn, in less-developed countries, it can be much longer.[4] Supposing that, we achieved sufficiently advanced IA solution based on BCI technology to significantly increase human intelligence. The following question needs to be answered: Who will have the priority access to enhanced intelligence using IA? People with the lowest intelligence and the worst living conditions to even out their ability to compete with others in society? Or rather the

wealthiest, as is the case with almost every new cutting-edge technology? Or maybe the elite of the totalitarian country in which IA is highly advanced?

In the "natural" circumstances of slow adoption of technology within society, other important questions arise: What will be the relation of those who use IA to the rest of society? How will people without this technology feel about it when coexisting with those supported with IA capabilities? What impact will this have on their sense of worth and competitiveness in the face of the increasing dominance of people with enhanced intelligence? IA technology can provide significant and growing advantages over time in any field of human activity for those who will be privileged to use its potential. This situation may lead in the coming years to widening the differences and tensions in society, ultimately bringing new, serious social conflicts locally as well as between global entities.

Selective Distribution due to Computing Power Limitations

Let's assume for the moment that, contrary to reasonable predictions, we'll succeed in making the IA based on BCI technology available for all willing people (e.g., one billion people) in a relatively short time (e.g., one year). In such a case, we'll face other significant concerns. How will human intelligence based on IA and distributed among numerous users compete with AI, which can be much more consolidated in terms of computing power? It's important to keep in mind that powerful AI systems can be highly focused on narrowly defined, potentially dangerous goals for the humanity. This problem may apply to the AI that got out of human control and acts independently of their will. It can also refer to the AI used by some humans (e.g., the elite of a totalitarian state or a terrorist group) who want to achieve their particularistic goals. The effectiveness of powerful AI focused on

narrow goals can be higher. As a result, the broad distribution of IA in society may not be sufficiently competitive with consolidated systems.

In the face of the above-mentioned danger, democratic societies may seek to adopt a different strategy for distributing intelligence. They may enhance a specific group of individuals, e.g., a democratically elected party or the military to maintain security and counter the growing threats from consolidated AI systems in the hands of hostile entities. In this strategy, privileged individuals can take advantage of powerful intelligence potential not accessible for the rest of citizens. However, this raises further important questions. Will they use the powerful potential of intelligence predictably and beneficially for the citizens of their country and all humanity? Will such individuals be able to abandon their enhanced intelligence if society so decides?

These risks associated with the direction of the potential use of IA, both in totalitarian and democratic countries, can lead to increased social tensions in the coming years. At a later stage, it can cause local as well as international conflicts. These tensions will also negatively impact the intensification of the arms race and the secrecy of AI and IA development.

V. Evolution of Values and Goals

Given the limited resources at our disposal, we may come to the decision that a group of competent and moral persons in society must be chosen to contain the dangers caused by consolidated intelligence. In this case, another question arises: How do we judge values and goals of the chosen persons? Even if we select the right persons with an impeccable reputation and good intentions, can we be sure that their values and goals won't change in the near future? It may happen that IA will be applied to

someone who is initially very empathetic, utilitarian, and aligned with the values and goals of humanity. However, with significantly enhanced cognitive abilities, they can change their views and attitudes toward some or all of humanity in a short period of time. We have no way of being certain whether a person or a group with powerful intelligence capabilities won't change their goals and attitudes toward other people even if they didn't suspect it beforehand. A similar process of rapidly changing values has already taken place in dynamically learning and, consequently, evolving AI systems.[5] It's hard to ensure that people who seemed to be the right choice for using powerful IA won't become what we fear in the context of the development of advanced AI algorithms – powerful superintelligence with misaligned values and goals towards the rest of humanity.

VI. Summary

In the present literature on existential risks, one of the most likely dangers to occur is almost always the AI that is misaligned with human values.[6] As presented above, the risks associated with the development of IA concept, especially based on high bandwidth BCI technology appear to be at least as high. Among the most serious hazards that need to be immediately considered are:

- The growing competition between countries as well as private entities that can start a new arms race. The risk of secret development – both in the case of totalitarian and democratic systems.

- Omission of safeguards – can take place during the development process both in totalitarian (achieving goals of the elite) and democratic societies (compromises in safety area in the face of threats from other countries).

- Consolidation of IA power and their selective distribution within society depending on a privileged social position or only within some country.

- Limited IA power per person in case of widely distributed strategy in contrast to the consolidated approach (e.g. powerful IA used only by a narrow group of people or by concentrated and liberated AI).

- Unpredictable, potentially very quick evolution of values and goals of the entities using enough powerful IA.

- In general: the use of advanced IA by narrow group of people (both in totalitarian or democratic countries) in a way that's highly undesirable with the expectations of the broad public.

Further development of IA concept can bring far different results than expected. Ultimately, it can lead to a situation in which a new, powerful entity or a group of entities can claim the right to arrange our world as they see fit. Confronted by this new existential risk, our countermeasures should be at least as intense in their scope as those employed to prevent the emergence of the traditionally understood AI. We should undertake extensive action as soon as possible to reduce this risk. Firstly, it's essential to make as many people as possible aware of the problem without delay. Secondly, it's necessary to agree on the most important risk factors and implement as effective a strategy as possible to counter the threats from this new direction.

PART III:

Existential Risks – New Representation, Basic Risk Factors

I. Key Questions

The main reason behind the development of AI, despite all its dangers, is the hope to solve the complex problems facing humanity and the lack, except IA, of clear alternative solutions on the horizon. Unfortunately, it's not only the development of AI that is disturbing and that can't be ignored. As outlined in the previous part, the development of IA can pose at least as serious threat to our species as AI. Unfortunately, this bitter conclusion doesn't fill us with optimism for the upcoming years. If we assume that the path through the IA development is at least as risky as AI, what remains then? It may seem we are on the losing end and all we can count on is a significant amount of luck and faith that nothing bad will happen. Perhaps our situation isn't hopeless. Perhaps, despite the conditions so far, there's an opportunity to overcome worrying trends. To understand where this conviction comes from, first, I'd like to pose the following questions, which are crucial for further consideration:

1. **Can the BCI technology have anything else to offer beyond Intelligence Augmentation?**

2. **Can the BCI technology de-escalate existential risks (despite the threat of IA)?**

These fundamental questions constitute the central axis around which the content of Parts IV, V, and VI revolve. However, before we proceed to answer the questions above, it's necessary to pose another one, which is based on the second one. This is essential if we want to assess the potential of BCI technology to de-escalate existential risks:

3. **What are the basic risk factors for the escalation and de-escalation of existential risks?**

This part is an attempt at answering the last of these three questions. The goal is to identify the most basic escalation/de-escalation factors for all major existential risks. Nowadays, in the age of wide access to countless sources of knowledge, the problem doesn't lie in the shortage of information, but in our limited ability to filter and prioritize it. The ocean of available data, when confronted with the natural human tendency to focus on its selective aspects, further obscures the broader context of the processes around us. Such focus can limit the perspective on the broad influence of one domain on ongoing phenomena in other domains. In this part, I'd like to try to break through the wide, in most cases dense, stream of information that floods us today. I'd like to try to indicate the basic factors which, at a higher level, cause many phenomena we observe around us and have a fundamental bearing on the growth of current problems.

The following section introduces a new broader concept of anthropogenic existential risks. In turn, further four sections will cover the base sources of all these risks. The steps above are essential for correctly defining major threats and understanding what factors may lead to their escalation and de-escalation in the coming future.

II. A New Representation of Anthropogenic Existential Risks

First, I'd like to recall all major anthropogenic existential risks:

1. Nuclear and chemical weapons
2. Biotechnologies
3. Nanotechnologies
4. Environmental degradation
5. Misaligned Artificial Intelligence

6. Misaligned Intelligence Augmentation

In the section below, I'd like to reorganize these threats into new representation – the categories of anthropogenic existential risks.

Unification of Nuclear and chemical/Bio/Nano Risks: TFMDR

The first three anthropogenic risks listed above, namely **nuclear and chemical weapons, biotechnologies, and nanotechnologies**, have much in common. All these risks can lead to a global disaster caused by an unfortunate accident, lack of proper safeguards, or hidden defects in the technologies. As a result, they can, unwittingly and unintentionally, become weapons aimed at humanity as a whole. Such a situation may happen, among others, because of the following incidents:

- The release of life-threatening radioactive substances, chemical compounds, pathogens, or nanobes into the external environment.

- Uncontrolled, dangerous modification of improperly designed biological organisms or synthetic nanotechnology-based devices.

- Unintentional, initially unobserved flaws or defects in technologies under development or already in operation.

It's essential to try to minimize above hazards. As especially biotechnologies and nanotechnology become more widely used and increasingly impact our lives every year it may be a growing challenge to effectively control and keep them safe.

In addition to unintentional, potentially dangerous incidents, above technologies can be strictly used as weapons and deliberately targeted against hostile entities. For thousands of years, people have been creating ever more effective technologies to fight against one another. From that perspective, the use of nuclear technology, chemistry, bioengineering, or nanotechnology as means of destruction is not novel. However, these modern technologies have much evolved from the weapons used in the past. Their scale of acting and unpredictability over time makes the key difference and therefore constitutes a separate, much more dangerous category of weapons that all existed before.

The existing term that describes some types of especially powerful weapons is “weapons of mass destruction” (commonly abbreviated as WMD). It's currently used to refer to military applications of following: nuclear, chemical, radiological and partially biological weapons. Although the concept of “mass destruction” reflects to their scale very well, the concept of a “weapon” itself, in the reality of the increasingly ambiguous, multidimensional nature of war, no longer reflects accurately the full spectrum of dangers from technology. Moreover, the term WMD, as commonly understood, doesn’t refer to the new directions of threats from, for example, modern techniques of DNA synthesis and recombination, CRISPR/Cas9 genome editing, and the use of inorganic nanotechnologies. With the increasing number of entities developing new branches of bio and nano technologies, and the intensification of non-military applications, it’s important that the public properly understands any risks from these new directions. Because of the strong and consistent perception of the technologies so far understood under the “WMD” term, its use in the context of new, not only strictly military, applications implies a risk of perceiving threats selectively, according to the traditional understanding of this term.

In the face of new dangers and the growing number of potentially risky uses, it's important to perceive all of them as consistently and comprehensively as possible. In view of the above, I'd like to propose a new term in this place to allow broader perception of the threats from all the above-mentioned technologies:

Technologies Fraught with Mass Destruction Risk (TFMDR)

The new term and the word "technologies" covers both the military and civilian applications. The phrase "fraught with mass destruction risk" indicates that these technologies may increase the danger of mass destruction, even if the work is by design intended only for civilian and peaceful applications. The dynamically developed technologies, not only on the military but also on the civilian ground, will determine our security in the 21st century.

Unification of Risks: AI and IA

Both AI and IA are types of technologies that could be interpreted as a subset of TFMDR. However, because of their exponential, potentially unlimited abilities of information analysis, rapid adaptation, and transformation of themselves and the external environment, AI and IA represent a particularly powerful, distinct class of threats. In the publications concerning existential risks to date, one of the most alarming is the threat of AI. Since there's also an existential risk posed by IA, I'd like to propose a new concept that includes this second important factor. Part of the public may understand the already existing term "Misaligned AI" incompletely to the spectrum of risks that may bring and perceive the misalignment problem as only related to dangerous actions of AI itself that escaped from human control and acts independently or at least misinterpreted our expectations. It's because the term

"Misaligned AI" may suggest (especially for the people unfamiliar with existential risk in detail) that it's strictly concerns AI. However, such a view ignores another critical threat associated with using powerful synthetic intelligence by narrow groups of representatives of the human species, e.g., an influential terrorist organization or the elite of a totalitarian government. From a terminological point of view, such a threat should be perceived as a part of the IA risk because it more precisely indicates the direct entity that triggers an existential catastrophe – humans who use technology to achieve their own potentially narrow goals and values. Given the above, the full term that covers the risks from any kind of intelligent entities (AI itself or humans) that are misaligned with the values and goals of humanity should be extended to the following form:

Misaligned AI/IA

The new definition with the "IA" abbreviation introduced here covers the full spectrum of actors. Powerful entities can be based entirely on synthetic structures (AI) or a combination of biological and synthetic structures (IA). In both cases it can be seen as new species on the Earth that is much more intelligent than *homo sapiens*. In the case of IA, actors can use indirect interfaces of communication between biological and synthetic intelligence (based on current perceptual-motoric channels of human body). In the perspective of current decade, indirect communication seems much more likely. However, in the perspective of the upcoming decades, we should expect a much more powerful and effective integration of biological and synthetic intelligence based on high-bandwidth BCI implants.

New representation of anthropogenic risks – main categories:

As a result, the new list of anthropogenic existential risks is as follows:

1. **Technologies Fraught with Mass Destruction Risk (TFMDR)**
2. **Environmental Degradation**
3. **Misaligned AI/IA**

This new form is the main one that we will use in the current and following parts.

III. TFMDR – Basic Risk Factors

Basic factor 1: Immanent set of physical attributes and behavioral tendencies

TFMDR are characterized by low safety even if we assume that multiple countermeasures are adapted. This is because of the correlation of the following, highly unfavorable attributes and trends:

1. **Scale and irreversibility of damages even in case of a single incident**
   The potential damage caused by TFMDR can cover a large area of the planet or its entire surface. What's essential, even a single incident can lead to such situation. An example of a single global event might be a nuclear explosion of a powerful warhead, contamination of water with harmful chemical substances, escape of designed viruses into the natural environment, or nanotechnologies that can damage biological organisms.

2. **Unpredictability of behavior – latent period in the environment, possible mutations and self-replication**
   TFMDR can be extremely hard or even impossible to control outside the laboratory environment because of the period of temporary secrecy during development, risk of mutations, and self-replication. The latent period in the environment can lead to their large-scale propagation in ways that are difficult to detect. They may manifest a tendency to mutate themselves (bio / nano) or the organisms with which they interact (bio / nano / nuclear / chemical). Biotechnologies and nanotechnologies may also carry the potential to replicate uncontrollably, covering more and more area.

3. **Highly undetectable source of origin**
   TFMDR have potential of being untraceable once they enter the natural environment. In most cases, they don't bear a signature which could clearly identify their origin and as a result may be a highly desirable type of weapon, e.g., during a war or conflict in which the parties wish to be perceived as positively as possible by the general public. Due to above, TFMDR can be a particularly tempting field of secret research and development. The potential of weapons based on advances in biotechnology and nanotechnology can carry a particularly high risk of conducting research under strict secrecy.

The characteristics above constitute the first basic risk factor. The nature of TFMDR and their critical level of security may, unfortunately, bring about the situation in which these technologies will pose an increasing threat year after year. Thoughtful use of TFMDR will become more challenging assuming increased progression in the sophistication of these

technologies in the coming years. If we want to increase the public awareness of the risk from this direction, it must be related to broad countermeasures focused on improving security in both the military and civilian spheres.

To summarize: The first basic risk factor stems from the immanent, highly unfavorable nature of the technologies in this area. The lack of widespread awareness and countermeasures to develop the necessary security measures will be crucial factors in the escalation of threats from this direction in the coming years.

Basic factor 2: Social tensions and conflicts

Even the most adverse characteristics of TFMDR don't fully determine whether the threats from this direction will increase. In addition to the above-mentioned first basic factor, there is also a second factor that originates from the social processes. The question that must be asked at this point is: Why would anyone want to use the potential of TFMDR against other people? The most common reasons are:

1. The desire to broadly improve current status is a valid reason for the use of any weapon. In this case, using the technology in question ensures the achievement of an objective aimed at improving status at the expense of others. In this situation, the motivation based on both pragmatic and emotional factors can play an important role.

2. A specific type of weapon may be used for strictly defensive purposes. Situations where one entity fears to

maintain its current status may lead to actions in which broad types of weapons are used to prevent negative changes. Both pragmatic and emotional motivations can play a role in this case.

3. The intention of causing suffering or at least making the attacker feel a kind of relief or satisfaction. The suffering inflicted in this case affects the emotions of the other party, and the expected gratification is to be a changed internal psychological state: a feeling of relief from the retaliation carried out, revenge for wrongs, etc. Emotional motivation plays a key role here.

In all of the cases above, the motive for using a weapon will be the subjective perception of the surrounding world and how other entities, events, and processes are perceived and analyzed consciously and subconsciously in the mind. The internal state of mind related to the external situation at a given moment and the desire to keep the current status or to make it better lead to emotional discomfort and internal tension. This state pushes people to make specific actions that may be perceived by others as more or less desirable. The complex web of interactions between individual members of society can induce social tensions. Ultimately, it can lead to actions such as using TFMDR. It's worth noting that social tensions can be both conscious and unconscious. Both types can lead to conflicts at a later stage.

Internal discomfort of a person in the reality of complex interaction with other members of society may lead to social tensions. This situation may result in conflicts in which a particular weapon will be used.

The internal state of tension of a particular person and, at a higher level, social tensions can be the effect of differences of a tangible and intangible nature:

Tangible differences – common examples:

- space,
- natural resources,
- created goods (incl. technology),
- physical body features.

Intangible differences – common examples:

- intelligence,
- knowledge and skills,
- believed values,
- worldview,
- behavior.

It should be strongly emphasized that neither tangible nor intangible differences lead to any tensions and conflicts by definition. In a very large number of cases, differences are accepted and perceived as desirable by people. This is largely true for differences such as attributes of the human body, knowledge, skills and behavior that make society and our lives more diverse. Moreover, this also seems to apply to differences in accumulated wealth. Although disapproval of this type of inequality is noticeable publicly, studies show that people do not want absolute wealth equality, but prefer fair inequalities resulting from hard work, ingenuity, morality.[1] The precondition for tensions is how differences are interpreted by specific individuals and groups. If they cause emotional discomfort such as suffering, the way is open for social tensions and potential conflicts (in which specific weapons can be used against the opposing party).

It's also important to note that some of the tangible and intangible differences can be referred to as inequalities. However, the term "inequality" is narrower and doesn't include all phenomena that can lead to social tensions. For example, the differences in the access to natural resources, such as drinking water in one region, or the difference in the possessions between a dictator and the average citizen of a country can be labeled "inequality". On the other hand, with regard to differences in skin complexion or cultural affiliation, it's hard to determine inequality in a measurable way. In this case, these can be referred to as "differences". **The term "differences" incorporates inequalities, while it also captures a broader spectrum of phenomena taking place within a society.**

Lastly, it is worth to mention particularly troubling social issues taking place nowadays. They affect the lives of large groups of people – numbering in millions. All of them have a significant potential to induce social tensions on the grounds of tangible and intangible differences. Low awareness of their existence and a lack of countermeasures to de-escalate, among others, the issues below will determine the level of threats stemming from TFMDR in the coming years.

- Geopolitical rivalry between the US and China. The technological "arms race" of the superpowers for dominance.[2]
- Shrinking or impoverishing (depending on criteria) of the middle class in some developed countries.[3] The shift of capital, labor, or new technologies to emerging countries as well as accumulation of wealth by a relatively narrow group.

- Revision of the international order by some states seeking to profit from the absence of an unequivocal global leader, which increases the risk of regional tensions. [4]

- The accelerating pace of change and the growing complexity of the modern world; May make it harder to understand and adapt to constantly changing environment.

- Increasing social polarization, post-truths, filter bubbles, and echo chambers.[5]

To summarize: The second basic risk factor stems from social tensions based on both tangible and intangible differences. Such tensions can lead to conflicts in which one or more types of technology (weapons) are used. The lack of widespread awareness and countermeasures to de-escalate current social tensions will determine the growth of TFMDR threat.

IV. Environmental degradation – Basic Risk Factors

Basic factor 1: The limitation of space and resources which, at the present level of development of life on the Earth, must be responsibly used to maintain the balance of the environment

The Earth's ecosystem is a limited resource. Soil, water, oxygen, minerals, and the space in which organisms live are limited. Any species' life depends on a fragile balance of interdependent organisms that use common natural assets present on the Earth. Continuous adaptation to the surrounding environment is inherent to every biological organism. It's especially efficient in the case of the most advanced of them. A sufficiently mature and organized species can learn to use the Earth's resources in such a way as not to adversely affect the individual components of the complex

system. Moreover, it can begin to tap into the vast resources available beyond the mother planet, such as the Solar System and more distant regions of space. For this to be possible, it must organize its immediate environment in which it claims dominant status and survives the critical transition period to a more mature stage of development. The first of the basic factors is the limited space and resources at the current level of life's advancement on the Earth. They must be used especially carefully and responsibly now to maintain the fragile balance of the ecosystem.

To summarize: The Earth has a limited amount of wealth. Individual organisms are dependent on each other. The first basic factor stems from the limitation of space and resources, which, at the present stage of development life must be responsibly used to keep the balance of the environment.

Basic factor 2: A model of development that is to some extent in conflict with the limited nature of the ecosystem

The global changes over the recent decades in the Earth's environment, including climate change, devastation of natural areas, and the mass extinction of species are commonly perceived as the most notable manifestations of the Anthropocene. This term is used to describe a new epoch in the history of our planet, when human activities have a key impact on the natural environment.[6] Every species transforms at least its immediate space in one way or another. Also, every human being wants to improve, enhance, and change at least its closest environment in a desirable way to create the right living conditions for themselves and their descendants. However, no other species in the history of our globe has had as significant an impact on the environment as humans have today. The problem is that the current development of *homo*

sapiens is taking place largely in a reckless manner, ultimately harming both other species and ourselves.

On the surface, it may seem that this situation is mainly the result of the fact that there is still no broad consensus on the degree of human destructive impact on the ecosystem. It's true that some people disagree on this matter.[7] Not everyone realizes the limited nature of the space and resources on the Earth and the need to use them wisely and share them with other species. Some people consciously or subconsciously deny the processes taking place (denial syndrome). This is largely caused by the fear of not being easily able to answer the questions that arise when our negative impact is acknowledged.[8] However, even if our environmental impact is widely accepted, it still can be difficult to stop current trends. This is because the problem of climate change is grounded on deeper issues. Some contemporary thinkers believe that people should focus more on the pursuit of being than on the possession of goods (the "to have or to be" dilemma). A way of living based on "being" rather than "having" involves being mindfully present in the here and now, appreciating those around us and the surrounding nature, rather than pursuing material objects that can distract us from our relationship with people and nature. It might seem that this second approach is a simple answer to our questions about why we currently have such big problems with negative impacts on the Earth's ecosystem and how we should live if we want to change it. However, the root of the problem lies deeper.

As humans, we want not so much an abundance of tangible possessions as a wealth of challenges, experiences, and goals. Of course, this kind of wealth can have, among other things, a material dimension, e.g., the goods we buy for our purposes or creative processes of transforming physical resources. Nonetheless, it's also largely nonmaterial, e.g., satisfaction from self-development, tasks we've done, things we've built and the

meaning we give to them. Every morning, we want a new day to be full of challenges and goals that we can strive for and that fit our aspirations and our understanding of what should be done. No sane person wants to spend their life staring at the ceiling – without goals, challenges, and experiences. This triad can be oriented toward others and the relation with nature as well as toward the world of matter. We don't want to only "be" or to only "have". We want to experience both in a balance that satisfies us. We want to orient ourselves toward other people and nature, but also toward creative development, transforming the surrounding matter, our world and ultimately changing our lives for the better.

Developed countries can provide a broader spectrum of such wealth. Their large economies are both result and a further base of the human need of fulfillment. The public is afraid of ideas such as central control of economy or lockdowns. The first instinctive fear is that this can lead to a loss of control over the life existential situation. However, what scares at least as much is the fear of the collapse of the present diversity of challenges, experiences, and goals that make human lives more meaningful and multidimensional. Are we able, in the limited conditions that our civilization can currently use, to provide ourselves with a constantly satisfying wealth of challenges, experiences and goals, without negatively affecting the Earth's ecosystem?

To summarize: The current ecological crisis stems largely from the conflict between the rising aspirations of 20th and 21st century humans and the limitations of the Earth's ecosystem. The current model of development is to some extent in conflict with the limitation of the environment we inhabit at the current stage of our civilization's advancement. We want to continually provide ourselves with rewarding challenges, experiences, and goals within the limited space and resources that our civilization currently uses.

V. Misaligned AI/IA – Basic Risk Factors

Basic factor 1: Immanent set of physical attributes and behavioral tendencies

Both the AI and IA technologies are hard to maintain safety, even assuming a broad countermeasures. This follows from the correlation of below set of risky physical attributes and behavioral tendencies:

1. Scale and irreversibility of impact beyond the critical point

The level at which AI/IA becomes sufficiently sophisticated and can act on its own may be the point beyond which there will be no turning back from the direction of its supremacy. After achieving the critical level of development, a sufficiently advanced technology may be uncontrollable and its further expansion unstoppable. The potential scale of damage caused by AI/IA actions can span the globe. It's worth bearing in mind that it can occur as a result of both intentional or unintentional actions of their developers.

2. Unpredictability of values and goals evolved over time

It's difficult to predict the evolution of AI/IA values over a long period. With powerful enough analytical skills and a bit of time for "reflection", AI/IA worldviews and values can change dramatically. Even the most altruistic and utilitarian attitude may quickly evolve to a far cry from the original one, as a result setting completely different goals and strategies for further action.

3. Exponential power that may encourages consolidation of overall potential

AI/IA power can enable significant, growing advantages over time for those who develop a sufficiently advanced technology first. Such prospect of "benefits" can be extremely tempting, in consequence favoring their narrow, selective consolidation. This attribute may promotes the secrecy of the development and selective distribution. When the public is aware of advanced works on AI/IA, it may try to pressure the entities involved in the development to stop them. From the perspective AI/IA funders, secrecy may be important aspects, providing as comfortable as possible conditions for further development.

The above characteristics of AI/IA nature represent the first basic risk factor that can increase existential risk. Mentioned issues will be more threatening if we are poorly aware of their existence. We should also realize that even if we increase public awareness and expenditure for technological safeguards of AI/IA, we may still have problems with transparency and the direction of development. Nonetheless, we should, wherever possible, increase efforts to raise critical levels of security and transparency.

To summarize: The first basic risk factor of misaligned AI/IA stems from the correlation of unfavorable physical attributes and behavioral tendencies of these technologies. The lack of widespread awareness and countermeasures to implement the safeguards will be the key determinant of the increasing risk from this direction.

Basic factor 2: Social tensions and conflicts

To a large extent, the level of risks from the AI/IA direction is determined by the global social situation. The escalation/de-escalation mechanism in this case is same as described for TFMDR (basic factor 2). As with TFMDR, the more tensions and conflicts within society, the more often they may be resolved through force and aggression. More conflicts imply greater risk that some of them will be resolved intentionally with technologies developed as part of the arms race or unintentionally as a result of their misuse. In conflict situations, AI or IA may be used deliberately to gain an advantage over hostile entities. As with TFMDR, the threat level from the direction of AI/IA depends on the widespread awareness of this dependency and the implementation of countermeasures to de-escalate tensions.

To summarize: The second basic factor is social tensions and conflicts within society. The lack of widespread awareness and countermeasures to de-escalate social tensions will determine the level of this risk.

VI. Natural Hazards – Basic Risk Factor

Basic factor 1: Natural processes that aren't caused by humanity but whose impact on life can be nonetheless minimalized

The causes of natural existential risks are independent of human activity: we have no control over the events that create them. It's because natural hazards stem from basic physical processes that take place inside our planet as well as in space. The scale and prevalence of natural events can lead even to the extinction of all

life on the Earth. Below, I'd like to outline the relationship between those natural processes and the risk of a natural disaster.

Collision with asteroids or comets

A collision with an object larger than 1 km in diameter will rapidly – within minutes or even seconds – lead to the destruction of almost all life within at least a few hundred kilometers from the impact site. After days and months, significant amounts of life-threatening chemical substances will be emitted into the atmosphere. This event will also lead to a series of earthquakes, supervolcanic eruptions, and tsunamis depending on the impact location and the diameter of the celestial body. Ultimately, all of the above can bring about the annihilation of dependent plants and animals, the disruption of food chains, and the death of more species.

Cosmic ray flare

Cosmic ray emissions can occur from solar flares, star collisions, and supernova formations. If a sufficient amount of radiation were to reach the Earth, it could change the composition of the atmosphere, increasing the amount of UV radiation, and ultimately, cause deadly DNA damage to living organisms. The increased levels of radiation reaching the Earth will also change the composition of atmosphere, potentially making it difficult or even impossible to continue living in the long term.

Supervolcano eruption

Large volcanic eruptions, in the short term (days or weeks), can lead to the destruction of life within a few hundred kilometers of the epicenter because of lava eruption. Within a few months, the emission of huge quantities of volcanic ash and other pollutants into the atmosphere can change its composition, contaminate the

air, soil, water, and cause a significant drop in the global average temperature. In the longer term, these changes may lead to the annihilation of many dependent plants and animals, the disruption of food chains, and the death of even more species.

Undoubtedly, the vision of a global catastrophe caused by an asteroid impact or a supervolcano eruption can appeal to our imagination and emotions. We can't be sure that a cosmic or earthly life-threatening event won't occur in the coming decades. Nonetheless, it should be highlighted that the probability of such an event occurring over the next decades is relatively low.

Important note on addressing natural hazards

We have no influence on the phenomena that are at the root of natural existential threats. We can't influence the physical phenomena leading to the formation of new stars and to the emission of dangerous cosmic radiation. We can't stop the geological processes that can lead to supervolcanic eruptions. We can't make meteorites travel through space without heading toward our planet. Despite the above, however, we have powerful capabilities to avoid such cataclysms. Thanks to our intelligence, knowledge, and focused actions, we can counteract the events preceding natural disasters. As a result, we can significantly minimize the risks that appear on our horizon. What does this mean in practice? As mentioned, we aren't able to stop cosmic rays. Instead, we can, among others, implement appropriate plans to diversify life beyond the Earth. We aren't able to stop the geological processes that constantly take place inside our planet. Instead, we can develop specific strategies if these processes cause major threats such as volcanic eruptions. We aren't able to stop asteroids from traversing space and entering on a collision course

with the Earth. Instead, we can monitor the space above our heads, detect asteroids approaching our planet, and change their trajectories or break them into many smaller pieces millions of kilometers away from us. Therefore, do we have a potential to minimize natural hazards in the coming future? Definitely. Below are some of the most important actions that can help mitigate these types of threats.

1. Supervolcanoes – examples of countermeasures:

- Monitoring geologic activities and detecting anomalies that may indicate an impending eruption.

2. Asteroids – examples of countermeasures:

- Detecting objects that can enter a collision course with the Earth.

- Systems designed to prevent comets or asteroids from colliding with the Earth, including development of missiles that can alter the trajectory of an incoming object or break it into many smaller pieces.

3. Cosmic ray flares – examples of countermeasures:

- Detecting impending collisions between stars that can lead to life-threatening radiation emissions.

- Monitoring stars that may soon turn into supernovae and emit life-threatening amounts of radiation.

4. **Countermeasures common to all of the above natural (and importantly, also anthropogenic!) hazards:**

    - Strategies for evacuating key species to a safer place on the planet (under/over the globe's surface). Particularly relevant for events of limited scale.

    - Terraforming strategies and technologies to restore the most livable conditions possible on the Earth's surface. Simultaneously, they may be key to making other celestial bodies viable, supporting the diversification of life.

    - Strategies and technologies to diversify life among as many celestial bodies in the Solar System as possible. They're particularly relevant to threats other than high-power radiation flares.

    - Strategies and technologies to diversify life among as many, possibly distant, celestial bodies. They're especially important for high-powered radiation flares that can threaten life over a large area of space.

There are many ways to minimize the risk of a disaster that can threaten life. Importantly, implementing even some of the solutions above, especially those related to life diversification, in a mature, refined form can significantly minimize the level of such risks. The further development and expansion of humanity on a cosmic scale is as real as it gets. In doing so, we must focus on our priorities. Our partial or total passivity toward the threats should be seen as a significant risk of the annihilation of our species or even all life on our planet. If we don't try to create technologies that can realistically counter existential risks, we condemn

ourselves to the fact that in the long run, we'll move toward a global catastrophe on our own wish.

To summarize: The key underlying risk factor is independent of human activity and results from natural processes occurring within our globe and in space. However, despite the unstoppable origin of these phenomena, we have a powerful potential to counteract the events that precede these disasters. As a result, we have a huge potential of minimizing the natural existential risks.

VII. Basic Risk Factors – Summary

In the previous sections, I outlined basic factors that have a fundamental impact on the existential risks. The table below presents them all in summary form.

Threat type	Basic risk factors
TFMDR	1. Immanent set of physical attributes and behavioral tendencies: A) Scale and irreversibility of damages even in case of a single incident B) Unpredictability of behavior – latent period in the environment, possible mutations and self-replication C) Highly untraceable source of origin 2. Social tensions and conflicts arise from the following differences: A) Tangible: space, natural resources, created goods (incl. technology), physical body features. B) Intangible: intelligence, knowledge and skills, worldview, believed values, behavior.

Environmental degradation	1. Limitation of space and resources which, at the present level of development of life on the Earth, must be responsibly used to maintain the balance of the natural environment. 2. Model of development that is to some extent in conflict with the limited nature of the ecosystem (at the current stage of our civilization's advancement). We want to provide ourselves with continually rewarding challenges, experiences, and goals within the limited space and resources our civilization currently utilizes.
Misaligned AI/IA	1. Immanent set of physical attributes and behavioral tendencies: A) Scale and irreversibility of impact beyond the critical point B) Unpredictability of values and goals evolved over time C) Exponential power that may encourages consolidation of overall potential 2. Social tensions and conflicts arise from the following differences: A) Tangible: space, natural resources, created goods (incl. technology), physical body features. B) Intangible: intelligence, knowledge and skills, worldview, believed values, behavior.
Natural hazards	1. Natural processes not caused by humans; despite their unstoppable origin, we have a powerful potential to counteract the events that precede such disasters. Consequently, we have a huge impact on minimizing them.

VIII. Technology – Opportunities and Threats

At the end of this part, I'd like to mention the relationship between the development of technologies in general and the problems and challenges we must face as a society in the coming years.

The unreflective implementation of every so-called innovation will not make the overall quality of life on the Earth better.

Our activities and the things we create have widely differing impact on individuals and society. Technologies have a specific potential, and yet it's largely our choice to decide if and how they'll be used. Let's take a very simple tool as an example. A knife helps us prepare meals to survive; but if used improperly, it can cause serious harm accidentally, or it can even be used as a weapon to hurt other people on purpose. Usually, we take far-reaching precautions into account when we consider the implications of using a technology. We understand that not every technology should be available to everyone without any restrictions. However, we should pay much more attention to the technologies whose safe use is less than clear. Implementing every so-called innovation won't necessarily enhance the general quality of life on the Earth. This type of thinking and acting can have the exact opposite effect. However, to be in a better position to comprehend what can constitute an improvement in our quality of life and, consequently, constitute real rather than illusory innovation, we need to have the proper perspective. We need to be aware of the processes and problems surrounding us. Based on these, we need to be much more deliberate in setting goals that will lead to the most thoughtful progress possible.

Low awareness of the factors that increase existential risks leads to chaotic selection of priorities we focus on

A key challenge is to be aware of which actions we should focus on. As a society, we still have a vague picture of the problems we must face in the coming years. Many of us either don't see them at all or only see their fragments. The lack of a wide perspective leads to a chaotic selection of our activities. We must redefine our priorities and set strategies that are more oriented toward crucial problems. Most people who have a real impact in nanotechnology, bioengineering, and AI/IA are highly talented experts in their field. However, a high level of specialization combined with qualities such as commitment and attention on a given problem can lead to the phenomenon called tunnel vision. These undoubtedly valuable and desirable personality traits may paradoxically result in the lack of awareness of the relationship between a given narrow issue and other aspects of the surrounding reality. Because of this, some of the people who are engaged in technological progress are convinced that their results' have only beneficial impact. There is also a group that is at least aware of some of the possible risks. However, they believe there are still many reasons to continue the work: real passion and strong emotional attachment to current activities, exaggerated optimism about the results and their use by people in the future, or a beneficial economic and living situation resulting from the participation in the development of cutting-edge technology.

At least equally worrying is the lack of awareness among those who aren't involved in the development of potentially risky technologies. Despite their lack of engagement, these people have enormous potential to influence the de-escalation of particular threats. It's because every person can share information about the issues with others and engage in several indirectly related

activities. Widespread knowledge about existential risks should be in every person's interest, as such risks will affect everyone's life in the coming years. If we're unaware of the implications of the development and use of nanotechnology and biotechnology, how will we evaluate the associated opportunities and risks? If we don't realize the impact of AI on our future, why should we do anything to counter its risky development? If we don't see relations between social tensions and the development of dangerous technologies, why should we invest much more attention in solving major social problems?

We must see the broadest possible perspectives of the processes around us. Otherwise, we'll keep ending up in situations where we focus on inadequate activities instead of concentrating on the root causes of the problems, which can only increase if they're overlooked or deliberately ignored. The existential risks and basic risks factors outlined in the preceding pages are crucial to gaining the appropriate perspective and emotional distance to the technological progress that is taking place nowadays. Significantly, this concerns a potentially highly important area for our future that I'll explore in the next part – the BCI technology and its applications.

PART IV:

BCI – Outlining the Spectrum of Application

I. Application Areas

In the face of the growing problem of anthropogenic existential risks, we're now going to explore the potential of the currently emerging BCI technology. We will try to answer the following question: **Could BCI technology have anything else to offer beyond Intelligence Augmentation?**

The first step toward the answer is to spread the canvas on which the panorama of potential BCI applications can be presented.

Division of BCI Applications – Main Areas

In general, BCI applications can be divided in many ways. In order to discuss all of them in a possibly clear and consist manner, I'd like to propose division based on the functions performed by the human brain and link them with the specific applications of the BCI. The functional division doesn't refer strictly to the topology of the brain, which means its specific physical regions, but to the high-level areas of tasks that allow functioning in the external environment. This perspective consists of four largely independent parts that, nevertheless, by their very nature, work together and influence each other.

Functional Division of the Brain

- **Intelligence Area** - responsible for complex cognition processes; processing, filtering, and persisting information from the senses as well as those that are the results of internal thinking processes; understanding and defining speech. All above may take place at the conscious as well as subconscious levels.

- **Emotional Area** - responsible for regulating and feeling emotional states (such as joy, excitement, sadness, anger, fear among others). This area plays an important adaptive role in shaping the ways we interact with the outside world by building appropriate relations with the environment, including responses to external events.

- **Intrasomatic Area** - responsible for optimizing the functioning of the body thanks to mechanisms of specific individual organs and entire systems such as the respiratory, circulatory, digestive, lymphatic, immune, endocrine system. It has a fundamental impact on the processes taking place in any part of the body and for overall condition of the entire organism.

- **Perceptual-Motoric Area** - responsible for receiving stimuli from the external world (perceptual part) and interacting with it (motoric part). It is also responsible for first stages of cognition thanks initial processing and filtering stimuli received from the senses.

The functional division outlined above will be the basis for presenting the potential applications of the BCI. Since we can define the functions of the human brain within the above-mentioned areas, it is reasonable to expect that the technology of implants can be applied for each of the four specified fields. Base on that, we'll operate on four main areas of the potential BCI application: intelligence, emotional, intrasomatic and perceptual-motoric.

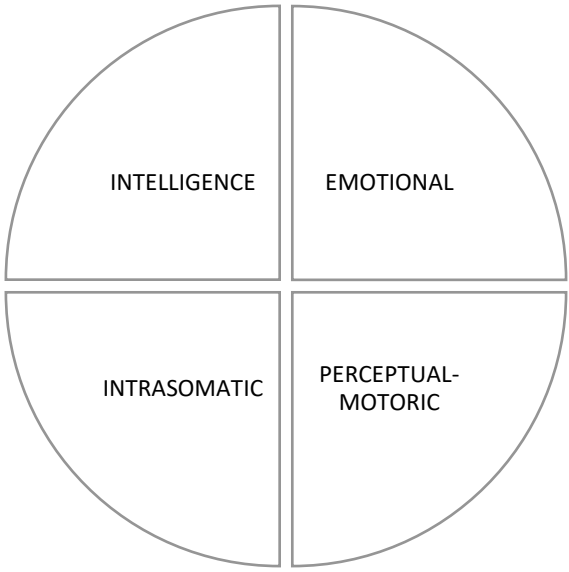

Figure 7. Functional division of the brain, which delineates the BCI application areas.

Sub-Areas of BCI Applications

In addition, we can define several sub-areas that cover more specific applications of the BCI. The following sub-areas will be presented separately in subsequent sections of this part.

- Intelligence: medical treatment
- Intelligence: intelligence augmentation
- Emotional: medical treatment
- Emotional: emotional regulation
- Intrasomatic: medical treatment
- Intrasomatic: intrasomatic enhancement
- Perceptual-motoric: medical treatment
- Perceptual-motoric: close reality
- Perceptual-motoric: remote reality
- Perceptual-motoric: digital reality

II. Intelligence Area: Preliminary Remarks

It should be noted that the concept of intelligence is often divided into some specific subtypes. One of the more quantified divisions was proposed by Howard Gardner and is commonly known as the concept of multiple intelligences. Gardner division identify eight types of intelligence: logical-mathematical, spatial, musical, linguistic, interpersonal, intrapersonal, naturalistic, and bodily-kinesthetic. However, no matter how finally we'll divide intelligence, from a neurobiological perspective all such types are based on the same process of transmissions electrical impulses and processing information, in the neural structures of the brain. It's essential to note, that analogous scheme of information processing is the foundation of modern AI algorithms, which are based on the concept of artificial neural networks.

In AI systems, artificial neuron structures can be trained to perform specific activities, e.g., solving strictly math problems, recognizing music, or detecting objects like in the case of autonomous vehicles. In accordance with Gardner's division, we could conventionally group above capabilities as logical-mathematical, musical, and spatial types of artificial intelligence. However, from an information processing perspective, at the most basic level, the mechanism of both biological and artificial neural networks is analogous for each of these types of intelligence**.** The highly adaptive and general nature of neural networks, both biological and artificial, carries certain implications for the intelligence augmentation (IA) that will be highlighted in the next section.

There are risks (as well as opportunities) that the same IA implants could be relatively easily adapted to expand the abilities in different types of intelligence. IA technology developed for seemingly narrow purposes, e.g., recognition or support for music

composition could be rearranged to other area of broader intelligence, such as linguistic or bodily-kinesthetic. Of course, this kind of reorientation may require some implants adoption to specific requirements, but in view of the previously developed, high-performance IA implants, it may be possible to achieve by engineers relatively quickly.

Ultimately, the narrowly focused domain of IA can be expanded to perform completely different tasks across the broad spectrum of intelligence. Due to the potential consequences (both positive and negative) of developing initially narrow IA and further extending its possibilities to many others fields of intelligence, the development of IA technologies will be treated as one wide and strongly correlated area that covers a broad spectrum of possible applications.

III. Intelligence Area: Medical Treatment

Diseases and disorders of intelligence area are related to the malfunction of specific brain areas. They can cause, among others due, physical injuries, genetic conditions, or trauma. The symptoms of impairments can be deficits in the functioning of memory, attention, analysis as well as personality changes. Frequent and serious diseases and disorders include, among others schizophrenia and amnesia. Intelligence abilities can also deteriorate as a result of dementia processes, which progress gradually over a long period of time. The main factors of their occurrence are age and inappropriate lifestyle. Dementing processes take place for example in Alzheimer's disease, frontotemporal dementia, or vascular dementia.

Applications of the BCI:

BCI technology have huge potential to be applied in order to overcome many disfunctions of intelligence area. Many common brain lesions, including advanced stages, could be stopped or even restored to health state. This could be possible, by using BCI to stimulating the affected areas of the brain in order to correct it's functioning. The electrical impulses sent by the implant could also stimulate some brain regions to rebuild and reinforce its activity. In turn, if certain areas of the brain were overactive, the BCI technology could correct its works by absorbing redundant electrical impulses. Alternatively, restoring proper function could also be possible by stimulating the appropriate area of the brain to trigger the production of neurotransmitters responsible for suppressing the work of the dysfunctional brain region.

BCI technology could also play an important role in monitoring the quality of brains' cognitive processes and signaling any potentially hazardous anomalies of its functioning. This could help prevent the progression of specific brain issues, which can slowly and unnoticeably increase over many years. Monitoring could also serve as an important part of the above-described approach of stimulating/suppressing the activity of a particular brain area in order to restore its function. In this case, the information gathered in monitor process could be a feedback that deciding whether to stimulation or suppression should be activated.

BCI implants could also reactive functions of irreversibly damaged part of brain by delegating given responsibilities to artificial, external processing unit connected with patient brain. Such external system could be a set of pre-trained or ready-to-train artificial neural networks that could be capable to restore the lost function of the brain or to learn given skills and store knowledge depending on specific requirements.

Alternative Uses of Modified Neuralink Threads

A non-standard application of Neuralink threads could be using them for the purposes other than transmission of electrical impulses. If the diameter of a single thread could be increased so that the threads could be turned into tubes, we could transport chemical substances or even complex biological structures (e.g. stem cells) through them to support the regeneration of damaged areas. Such threads could be permanently placed in the brain to deliver the precise amount of drug at the exact moment and place. The whole process could be coupled with standard, electroconductive threads and related implant to monitor the state of the treated area in real time. Alternative kind of threads application could contribute to the creation new types of therapies, in which the implementation of appropriate medications would be much more adaptable to the individual characteristics. This could reduce the number of side effects, increase precision, and, ultimately, improve the effectiveness of treatment.

IV. Intelligence Area: Intelligence Augmentation

Brain Information Processing – Functional Perspective

The information can come to the brain from external sources or internal cognitive processes. External information comes from the human body's surrounding environment and from other than brain parts of the human body. In turn, the internal source is the brain itself. They can come from memory or a continuous processing process, the starting point for further cognitive processes.

The human brain filters and classifies external and internal data to optimize the entire body's functioning in the external environment. Irrelevant data are discarded, while important are

persist into long-term memory. In turn, short-term memory is a temporary working memory for the data currently processed.

In addition to the distinction between short- and long-term memory, it is worth noting that from the "access to resources" point of view, we can also distinguish declarative and non-declarative memory types. Declarative memory stores memories, knowledge, and all other information that can be called "on-demand" to our consciousness. Such information may have both an abstract and a more concrete form (e.g., visual or sound). In turn, non-declarative memory stores information that can be called by the human brain mainly without the direct participation of consciousness in this process, e.g., acquired body reflexes and muscle memory.

Applications of the BCI:

(A) Augmented Processing and Memory

The Intelligence Augmentation using BCI implants can enhance natural human abilities by establishing high-bandwidth communication channels between the human brain’s neurons and external artificial neural networks. Those additional neural structures could extend our default potential of information processing, associating, storing, and its further re-calling. The physical location of artificial neural networks could be in our skull or other parts of the body (as far as body physiology allows). However, they can also be placed outside the body, in our smartphones, or even thousands of kilometers away in cloud computing centers. Implants such as Neuralink already support wireless connection with an external device such as a phone. Therefore they can indirectly allow for communication with cloud centers in any place on the globe.

The augmentations of intelligence using BCI can also be possible through monitoring and stimulating strictly biological structures of the brain. Continuous monitoring of brain activity and stimulating given neurons can be an important alternative or complementary strategy for enhancing cognitive functions using artificial neural networks. This approach can allow for improving existing connections between neurons in the brain. It can also help in the formation of new connections process and strengthen their durability. In the longer term, it can also improve the overall neuroplasticity of the brain.

Both the IA approach based on enriching the brain with external structures of artificial neural neurons and the approach oriented to improve the functioning of existing biological neurons of the brain can significantly increase default human efficiency in many fields. Both approaches may increase overall short-term memory capacity and enhance our natural ability to process information in a given moment. In turn, improved long-term memory may extend human capabilities of memorizing and recalling experiences, knowledge, and skills. Worth noting that biological memory tends to fade after some time - this is especially the case for information that is not recalled to consciousness too often; therefore, the relevant neuronal connections are not reinforced. In the case of data located in external artificial networks connected with the brain, the above types of memory (short and long-term) can be much better in case of durability and details.

IA can vastly increase natural human skills. Augmented processing and associating skills can vastly improve sequential thinking, decision-making, processing speed, and accuracy. These will also improve our attention and abilities to make a focused, long-term, and in-depth analysis of any problem we face. Improved abilities to memorize, store and recall information can

also play an essential role in improving our processing and associating. It also can enhance the capabilities of learning and remembering any kind of knowledge and experience.

All above can vastly enhance human abilities that we usually define as intelligence. Applying BCI broadly across many parts of human brains can increase overall cognitive potential. In turn, using implants only in specific brain locations can increase some narrow abilities. For example, from the perspective of Gardner's intelligence division, the application of BCI in the prefrontal cortex area can improve mathematical-logical and intra-personal intelligence. The use of implants in the motor cortex area may lead to enhancing mainly motor type of intelligence. In turn, application in the Broca's area and temporal cortex can improve linguistic and musical intelligence. Of course, above relationships between physical brain areas and specific types of intelligence should be treated as approximations. It should be emphasized that rigid Gardner's division of intelligence types is conventional, and we can also use more as well as less quantified divisions.

B) Further Stage of IA Development: Internet of Thoughts

IA could enable knowledge and skills collected by humans to be stored in external cloud systems. Moreover, the person who acquired them would make them available to everyone or a specific group of people. If desired, the collected knowledge and skills could also be refined by others until they reach a satisfactory degree of accuracy. Once acquired in this way, it could serve others equipped with IA anytime, anywhere by accessing other people's collective knowledge and skills. An apt term for this type of technology can be the Internet of Thoughts.

It is important to note that the technologies necessary for applying such a concept can be challenging to develop. The

problems stem from the uniqueness of the organization of neurons in human brains that play an essential role in the thinking process. Moreover, the organization of neurons that represent learned skills, knowledge, and experiences is also unique for each human. When person A recalls any concept in the brain, this process induces differently organized structures of neurons than in the brain of any other person. Even if the same brain parts (e.g., the same areas of the neocortex) are activated, the internal structure of neurons is different and unique for everyone. The same goes for skills and experiences, which have a unique representation in brains. Each person perceives ideas, concepts, and memories differently. This problem has profound implications in the context of the Internet of Thoughts. It seems to be absolutely required for such technology to create sophisticated neural structures converter which will be able correctly map, transmit and finally interpret thoughts between humans with sufficient accuracy. From a technical perspective, developing such a bi-directional neurons structure converter can be particularly challenging.

Intelligence Augmentation and Consciousness

The processing of information in the human brain is a well-studied part of humankind's knowledge. However, connecting this area with the nature of consciousness is still challenging. The consciousness emerging issue is still not discovered and described enough, despite being widely analyzed by science. There is currently no one theory that would explain the nature of consciousness coherently and indisputably. Different perspectives on this important topic present David Chalmers, Robert Lanza, and Daniel Dennet, among others[1]. In this context, Intelligence Augmentation can potentially help answer some of the questions related with consciousness such as:

- What is the nature of consciousness and the "mechanism" of its origin?

- Will enhancing human intelligence affect consciousness and expand it in some way?

- Will IA-enhanced persons be able to focus on multiple tasks simultaneously (instead of sequentially switching between several in a brief period as happens during the illusion of divided attention in humans)?

- Will the enhancement of intelligence and especially Internet of Thoughts technology use affect the consistency of our identity and sense of ego?

The above essential questions may be empirically verifiable in the coming years.

V. Emotional Area: Medical Treatment

In the emotional area, there are a number of illnesses and dysfunctions that can significantly affect the quality of life of people struggling with them. The issues include, among others, depression, manic-depressive disorder, phobias, anxiety, post-traumatic stress disorder, and obsessive-compulsive disorder. Emotional dysfunctions may be related to abnormalities in the functioning of neurons, which, by default, are meant to interpret external phenomena (those coming from the environment) as well as internal thought processes. In the case of acquired trauma, phobias, or anxiety, specific areas of the brain may interpret incoming information in an extremely inappropriate manner, leading to severely disturbed emotional states. Endocrine and

neurotransmitter disorders are also important factors which lead to emotional disorders. Chemicals produces by human body play a key role in the feeling of specific emotions, tensions, or relaxation. Over- or under-production of specific chemicals can lead to depression, mania, chronic tension, or anxiety.

Applications of the BCI:

The BCI technology can potentially enable the correction of emotional area malfunctions in a dynamic and precise manner. Thanks to Neuralink threads, the activity of specific malfunctioning brain areas that cause emotional disorders could be enhanced or suppressed to a strictly specified degree. Remedying the disorder can become crucial in teaching more appropriate emotional responses and lead to healing acquired traumas, phobias, or anger attacks. In turn, use of a modified version of the Neuralink threads that would be capable of transporting chemical compounds, could directly deliver specific drugs to precisely defined regions in the brain. This could be especially reasonable for supplying specific neurotransmitters if the body were not able to produce them on its own in a sufficient quantity. Another field where BCI technology can be applied is real-time monitoring of specific areas of the brain to analyze current emotional states. Such data could be an important part of the diagnostic process. They could be used on a feedback basis to make decisions on the treatment, e.g., whether to stimulate or suppress the overactivity of malfunctioning brain areas.

VI. Emotional Area: Emotional Regulation

Emotional regulation involves intentional influence on the emotional processes occurring in the brain. Human emotions can

be regulated (intentionally and unintentionally) by internal thought processes as well as external (to the brain) environmental events. It is important to noting that the reception of external stimuli affecting emotions can take place through various senses. It applies to sight, taste, smell, touch, and other senses. In addition, an emotional state can be affected by many specific substances delivered to the body. Most common examples are sugars and fats, which can release neurotransmitters responsible for the sensations of pleasure and bliss, among others. There are also other substances, such as alcohol and other drugs, that can strongly affect emotions. Moreover, an emotional state can be intentionally changed, among others, by many relaxations and stimulation techniques. This can be achieved by engaging one or more of the senses in some training or strictly by focusing thoughts and attention into specific activities. It also can be achieved by changing previously acquired patterns of behaviors and related emotional reactions to more desired.

Applications of the BCI:

The BCI technology may provide methods for setting specific emotional states. This can be achieved by precisely influencing the electrical activity of specific areas of the brain (activity of neurons) and the emotional states they evoke. Stimulation as well as suppression of electrical impulses can also affect the level of neurotransmitters secretion which influence human emotional states. The main types of non-medical uses of BCI technology for emotional regulation are presented below.

A) Emotional Amplification/Suppression on Demand (Including Lower and Upper Extremes)

The BCI technology may potentially allow the feeling of any emotional state the user wishes. The emotional change could be possible by suppressing and/or amplifying the action of relevant neurotransmitters and neuroreceptors in the brain that are responsible for feeling specific emotional states. The scope of change could cover a narrow or wide range of neurotransmitters and neuroreceptors. Hence, it can affect on given range of emotional states. What's also essential, the level of emotional state change could be adjusted according to the user's will. In practice, this may allow feeling any intermediate emotional state, as well as upper and lower extremes. The upper extreme means strong amplifications of feeling a given emotion. In contrast, the lower extreme means strong suppression of feeling emotion.

EASD may be much more precise when it comes to time of acting and accuracy of setting specific emotions than chemical stimulators and suppressors currently used. In the future, traditional chemical drugs may be replaced with precisely adjustable "e-drugs". However, it should be clearly emphasized that reckless user's actions in order to feeling specific, especially extreme emotional states, may lead to serious health consequences. Similarly as with chemical substances, "e-drugs" may lead to health issues or even carry a life-threatening hazard for user and persons around. Careless emotions regulation may among others lead to addictions, permanent deregulation of neurotransmitters production, and permanent neuroreceptors damage. This in turn can lead to emotional disorders, behavior changes, and severe mental diseases. Arbitrary and widely-available in society adjustment of emotions can be justified in

specific cases, however, it's important to keep in mind above-mentioned potential impact on entity and society as whole.

B) Emotional Balance

The BCI can be also used to develop specific type of technology of emotional area that will be focused on maintaining balanced emotional state (without falling into destructive emotional extremes) and be able improve human self-control skills. In this approach, stabilization of emotions could be possible with two different approaches.

In the first, passive approach, BCI is used only for monitoring brain emotional activity, in order to providing feedback to the user device such as smartphone among others. In such a case, also non-invasive BCI technologies (such as EEG interface) could be used. Regardless of the BCI technology used in the monitoring process, collected date could be further used to analyze personal emotional states. This could take place by personal introspection (self-improvement) process. Collected data could be also important information for professional psychoanalyst that analyze them and suggest the best strategies of changing existing behavioral patterns to more desired. Another way is that in the future, specialized, trained psychoanalyst-algorithm could analyze the vast amount of brain activity data and propose best fits suggestions of further actions. Thanks to the improved awareness of the internal processes taking place in the brain, the user could try to change undesirable behavior patterns, including changes of emotional reactions related to them.

The second, active approach assumes direct influence of the BCI implant on the nervous system by stimulating and/or suppressing the brain activity to maintain continuously a balanced emotional state in long term. Unlike the "Emotional

Amplification/ Suppression on Demand" (technology described in the previous section), in this case the regulation of the emotional state is handled by the BCI implant's built-in algorithm, which is focused on maintaining the emotional stability in the long term. In order to maintain emotional balance, the implant continuously monitors emotional brain activity and performs regulation on a feedback loop basis. The implant monitor current emotional states, point out potential abnormalities, and act according to an optimally chosen strategy. It should be noted that active emotion balancing also includes the advantages of the passive approach. However, its enabling shifting them mostly to unconscious sphere by automatization of emotional balancing maintenance process.

From a practical perspective, emotional balance could help users eliminate chronic, impeding the proper functioning emotional states such as high anxiety and stress. It also could offset negative issues of low motivation as well as distracted focus, which may have an emotional basis. Furthermore, EB could help analyze our thoughts in a way less affected by emotions - without having overly negative or positive emotional associations to them. It could improve human abilities to analyze concepts, ideas, objects without strong emotional associations, which potentially can extremely bias our worldview and influencing on our further activities.

C) Further Development of Emotional Area: Sharing Emotions Between Subjects (Internet of Emotions)

At a later stage of emotional area development, the BCI technology can also enable sharing of emotional states between subjects: primarily humans, but also other species. To make it possible, emotional regulation technologies would have to provide sufficient precision in the emotional setting, which will be the first

preliminary stage for sharing emotional states. Additionally this solution also require communication with other people in a safe manner, which would be the next important step of preliminary work. If above will be achieved, then we could introduce technology capable for transmitting the intended emotional context between the sender and recipient. In practice, the sharing of emotions can also involve many persons. Thus, both senders and receivers can be not only individuals, but also large groups.

Similar to the "Internet of Thoughts" technology, developing and applying the "Internet of Emotions" may not be easy. It is because creating advanced solution for sharing emotions requires developing precise technologies of mapping, transferring, and reproducing emotional states. Nevertheless, the number of variables determining a given state seems to be much orders of magnitude lesser than in the case of sharing complex and highly unique neural structures (thoughts) between subjects. Thus, the technology for sharing emotions seems to be much more achievable in the coming decades than that capable of transmitting thoughts. It is also worth to note that technology for sharing emotions only or thoughts only, may be insufficient in some cases. Ultimately, these two technologies coupled together and providing both (thoughts and emotions) contexts will be able to offer a complete understanding of the transferred message.

VII. Intrasomatic Area: Medical Treatment

Dysfunctions of the intrasomatic (mainly autonomous) area can significantly impact the overall condition of the entire body. Many abnormalities of body functioning have the genesis in malfunctioning the autonomic nervous system, including its essential subsystem - sympathetic and parasympathetic. These subsystems affect the functioning of many other systems, including respiratory, circulatory, digestive, lymphatic, immune,

and endocrine. They also affect functioning glands responsible for the secretion of many key chemicals necessary for the organism's functioning. Abnormalities of these subsystems working may cause the emergence of many disease entities. Among the most common are diabetes, obesity, hypertension, anemia, and under/over activity of the thyroid.

Applications of the BCI:

Although the BCI technology, by definition, involves connecting the brain with computer, in practice it can have more applications, involving also peripheral nervous system[2]. BCI can support the working of many organs by reinforcing and suppressing electrical signals in specific places of the body. The applications of BCI technology outside the brain may help alleviate the symptoms of many diseases or even eliminate them. Moreover, it may allow precise continuous monitoring of glands for diagnostic purposes as well as for dynamic stimulation or deceleration of their work based on a feedback loop. Also worth noting that the use of BCI strictly within the glands located in the brain (such as the hypothalamus and pituitary gland) may also affect the functioning of peripheral organs. In turn, using modified Neuralink threads could precisely give the supply of liquid medicines to a specific location in the body. Both the standard threads and their modified version could support many current therapies or even replace prescribed drugs that are not precise enough and, in effect, carry a risk of potential side effects. BCI technology can significantly affect treating many diseases, including abnormalities in the functioning of respiratory, circulatory, digestive, lymphatic, immune, and endocrine systems, among others.

VIII. Intrasomatic Area: Intrasomatic Enhancement

Intrasomatic enhancement refers to the intentional, non-medical changes in the body's functioning. Such changes aim to influence body functions that, by default, are beyond the direct control of the human will. Despite the largely autonomous nature of the processes in our organisms, we can indirectly influence them through regular physical activity, diet, specific body-regulation training, avoidance of harmful chemicals, and other environmental factors. Also, specific drugs used intentionally may affect ad-hoc the functioning of particular organs and systems of the body. The expected outcome of the activities mentioned above is achieving the desired state of the body for a short or long time.

Applications of the BCI:

The BCI technology may enable precise regulation of the specific body parts functioning. BCI implants can change how individual organs and entire systems (e.g., circulatory, endocrine) works. Similarly to medical applications, it can be possible by precisely reinforcing or suppressing electrical signals in concrete places of the organism and changing way of functioning of specific organs. BCI can also allow continuously monitor and collect data for feedback loop process needs. In turn, the modified version of Neuralink threads could precisely supply specific chemicals to a given location of the body, which could be useful for improving the final enhancement results.

From the practical point of view, the enhancement of the intrasomatic area could accelerate metabolism, decrease appetite, regulate circulatory function among others. It also could increase resistance to many environmental conditions such as cold/hot climate or lower oxygen level). Furthermore, it could temporarily

increase the organism's strength which could be desired in some situations in which intrasomatic enhancements can determine health and life. Despite the potential benefits of this technology, it is important to be aware of the risks associated with its improper use. Inadequate, unconsidered use can lead to serious health disorders as a result of long-term or even single usage.

IX. Perceptual-Motoric Area: Preliminary Remarks

The perceptual-motoric area is divided into three non-medical sub-areas, which will be presented separately. This type of separate discussion results from a large amount of work that must be done within each of these sub-areas to make them useful. Each involves, in part, the same set of technologies necessary for the area to function (a technology stack), which primarily includes the implant technology. However, the development of each will involve the growth of new and distinct branches of engineering. As a result, the scope of work to be done to make each of them available for use is largely independent.

Specific sections on the perceptual-motoric area (Close Reality, Remote Reality, Digital Reality) will cover a different interaction type of the body with external environment. Such division for the perceptual-motoric area may at this moment seem not important, but in the future it can be of fundamental importance to us. This may be particularly the case when, limited by time, human potential, and material resources, we'll have to choose which technologies from the broad spectrum should be developed to effectively decrease all existential risks without increasing the risk caused by the development of others, potentially dangerous sub-areas.

X. Perceptual-Motoric Area: Medical Treatment

Human senses and body motor system can easily become impaired or damaged. This is largely due to the direct exposure to the external environment. Excessive or improper use and age-related decrease in their efficiency play a key role in this process. The extent of dysfunction and its nature differs widely due to the varying specific functioning of the individual senses and elements of the motor system.

Applications of the BCI:

Looking at the perceptual-motoric area from as general perspective as possible, it can be expected that development in the BCI technology will create multiple new sensory and motor implants for medical applications. BCI has the potential to help treat disorders of each of the senses, as well as motor functions.

Sight: The complexity of currently available camera sensors is sufficient to think about advanced vision implants based on the BCI technology. Current generations of sensors have resolutions well over 100 million pixels, which is higher than the total number of photoreceptors in the human eye (about 95 million). Moreover, available microchips can already fit into a prosthesis the size of a human eyeball. The signal to the brain can be transmitted wirelessly via module inside the eyeball implant to a BCI module located directly in the visual cortex (located in occipital part of the brain). At this stage, the problem to be solved is the development of a system that would support the movements of the eyeball prosthesis. Nonetheless, even a fixed eyeball prosthesis with the above-mentioned parameters would be a significant improvement in the quality of life of people with visual impairment.

Hearing: Hearing is very important for efficient communication and functioning in the environment. At the same time, this sense is relatively simple for at least partial reconstruction in comparison with sight. Hearing implants have long been the subject of research, and are currently widely used in medicine. They enable the recovery even of the people who have completely lost ability to receive auditory signals. The upcoming BCI technology will undoubtedly continue their development, especially leading to the improving so-called stem implants. Such kind of implants almost completely bypassing ear elements and the auditory nerve, allowing direct communication between the auditory cortex of the brain and the device. The use of a new type of flexible, high-density Neuralink electrodes may increase the bandwidth of the transmitted signals, ultimately improving the quality of hearing. It is worth adding that the current generations of microphones are already sufficiently advanced, and the quality of the sounds they capture can surpass the default capabilities of the human ear in many aspects.

Taste and smell: Currently, there are no implants that restore the senses of taste and smell. It is because of their relatively minor importance for efficient functioning in the modern world. The primary function of taste and smell senses is to detect potentially dangerous chemicals introduced into or around the body. Nowadays, the risk of undesirable effects of such substances on our body is eliminated largely by the achievements of civilization. Almost every substance we buy must meet strict standards regarding its use, for example name, symbols, ingredients, allergens, expiration date. Furthermore, potentially dangerous gases with characteristic odors pose lesser threat in everyday life thanks to devices such as smoke or gas detectors. Consequently, modern humans with taste and smell deficiencies being a bit

careful can function efficiently for their whole life. Of course, these senses play also different roles as they stimulate the reward center in the brain to induce feelings of pleasure, as well as other, less pleasant sensations. Ultimately, however, this function isn't critical enough to lead to the development and distribution of appropriate implants. Will anything change with the advent of the future BCI? This isn't out of the question. Much in this case will depend on the level of their technological complexity, including miniaturization and convenience of use.

Touch: Touch sensors are already being used as part of motor prostheses. Today, however, such kind of technology is still less precise than the natural human ability to perceive objects by skin receptors. This is mainly due to the insufficient number of electrodes in the BCI implants mounted in existing prostheses. The upcoming generations of BCI technology will certainly result in much greater precision of touch in new prostheses. To make this possible, it will require to apply touch sensors with a higher resolution so that they can transmit sensations of touch to the brain much more precisely. Moreover, the next generation of implants and prostheses should improve the ability to recognize many important properties of the physical objects which we perceive by touch, including better recognition of texture, pressure force, or temperature.

Nociception: Currently, there is no technology for transmitting pain-related sensations to the brain. This shouldn't be especially surprising, as not many people derive satisfaction from experiencing pain. However, it is worth realizing that the role of this unpleasant sensory-emotional experience is very important for the body, aimed at informing us about abnormal functioning of its various parts. The BCI implants placed inside the body could

functionally play a similar role to natural pain sensing mechanisms. Information about lesions could be directly sent to the implant in the brain, inducing an analogous sensation. It is also worth emphasizing that such stimulus wouldn't have to cause unpleasant sensations. Instead, it could take a purely verbal or symbolic form of information, sent directly to our visual cortex or to an external device like a smartphone through a typical notification. Additionally, nociception implants could be useful in specific areas of the body that don't have receptors that create the sensation of pain and as a result do not inform us about potential lesions. Moreover, the implants could also help in the case of neurological disorder called congenital analgesia, which manifests itself by a complete lack of sensing pain and can pose a serious threat to health and life.

Another application could be for undesirable, too intense, or chronic types of pain, which, due to their aggravating nature, hinder effective daily life. In this case, the BCI technology could work antagonistically, intercepting the impulses coming from the pain receptors and changing the way the brain interprets them to eliminate the unpleasant sensation. What is important, such a solution of overcoming pain could be precise, more resistant to increasing tolerance, and toxicologically safe unlike the analgesics that are currently in use. Such an application could be invaluable for many diseases of varying backgrounds where intense pain is a severe problem for patients.

Proprioception: The upcoming BCI implants can be used for transferring information about the muscle tone, stretch, or positioning of different body parts in relation to each other. Such technology can be crucial in overcoming multiple dysfunctions of the body's motor system. An essential part of the technology to support the sense of proprioception is to design sufficiently small

and efficient BCI implants that can be placed in multiple locations inside and on the body's surface. These implants would collect information from the specific parts of the body and transmit them to the brain. In turn, in cases where it is the brain's interpretation of the signals that is the problem, an implant placed in the brain, with the support of appropriate algorithms could try to correct them.

Balance: Future BCI technology can be used for overcoming disorders of the balance sense. An implant placed in the brain, could correct issues in the vestibular system (responsible for balance) by interpreting certain impulses as inappropriate and correcting them. In turn, in the case of major or irreversible damage, the specialized implants could replace innate vestibular system functions in order to restore the entirely lost sense of balance. Similar to proprioception implants, such technology could be important in overcoming many severe diseases of the body's motor system.

Motor Skills: Currently available prosthetic limbs are being improved with every passing year in terms of strength, weight, and comfort of use over many hours. However, progress is still needed with regard to their speed and precision of movements. Insufficient speed applies to so-called active prostheses, i.e., those whose movements are assisted by an assembly of motors for setting their parts in motion. Fortunately, this aspect is being improved last years and this trend can be expected to continue in the years to come. When it comes to the precision, we can expect that the use of upcoming BCI implants with high-density of threads will enable a more extensive integration of the motor cortex responsible for motor functions with the prosthesis. As a result, future generations of prosthetic limbs will communicate

much better with the brain and offer higher precision. The wide range of sensors for transmitting the sense of touch, nociception, proprioception, and balance will ultimately lead to highly advanced prostheses in the near future.

Next-generation BCI-based prosthetic limbs can allow people with multiple disabilities to function efficiently. Among other things, it may restore the function of the damaged peripheral nervous system. For example, in the case of complete spinal cord transection, the optimal approach would be to use two BCI modules communicating with each other to transmit electrical impulses. One would be placed below while the other above the damaged section of the spinal cord or directly in the motor cortex area of the brain. Such modules would communicate wirelessly with each other in order to restore the transmission of nerve impulses to and from paralyzed limbs. Moreover, a modified version of Neuralink threads can play an important role in peripheral nerve damage treatment. They could be used for precise transportation of stem cells to the damaged area to support its regeneration.

Another application that is also based on motor skills is speech. If the speech apparatus is dysfunctional, it could be bypassed by BCI implant connected to the areas of the brain responsible for speech processes. Thanks to that, nerve impulses could be sent to an external device, bypassing the damaged area, and then verbalized using a speech synthesizer (easiest way) or using future, active prosthetics of the speech apparatus.

XI. Perceptual-Motoric Area: Close Reality

Close reality (CR) involves interactions with the closest environment of the body using only the senses and motor elements integrated directly with them. Applications of CR include

improving human's inborn sensory and motor skills as well as extending the existing sensory spectrum with new ones.

A. Improved Sensory and Motor Skills

Improved senses and motor skills refer to the improvement of the inborn senses and motor skills of the body.

Applications of the BCI:

Sight: The BCI technology could programmatically eliminate noises from images captured by the sense of sight as well as physically increase the number of sight receptors to improve perception in specific environmental conditions. It could enable improved detection of visual objects, upscale sensed images, or select only the specific part from the visual field. Perceived visual data could also be recorded into inbuild memory and processed later according to our needs.

Hearing: The BCI applications could eliminate noises and select particular data from the environment, for example, filter specific sound frequencies, voices, and other elements of the environment. BCI could also increase the number of hearing receptors to improve the quality and clarity of perceived sounds. Collected audio data could also be recorded into inbuild memory and processed later.

Taste and Smell: The use of BCI technology could enable modifying the taste and smell sensations as well as enhance or suppress them. It could also increase the number of taste receptors to enhance the ability of its detection and perception. It could also

record flavors and scents into inbuild memory in order to process them later.

Touch: The BCI applications could increase the number of touch receptors for enhancing perception. Touch sensation could be amplified or suppressed on demand. It could improve sensing and analyzing texture, pressure, vibration, temperature of touched objects. It could record the sensation of touch into inbuild memory and process it later.

Nociception: The BCI technology could partially or entirely suppress unpleasant sensations of pain. The persistent pain could be converted to a strictly informational form that does not produce an unpleasant feeling. It could also enable nociception in the body parts, which by default are not equipped with receptors to detect disorders of its functioning.

Proprioception: The BCI applications could modify the sensation of muscle tension by strengthening or suppressing signals from specific body regions. It could improve the orientation and awareness of the position of individual body parts in relation to each other. It could improve general body coordination.

Balance: The use of BCI technology could improve the balance skills of the inborn vestibular system by using algorithms of corrections or by using additional sensors which support the sense of balance. It could improve the general coordination of the body.

Motor Skills: In the case of motor skills, upcoming BCI technology may result in the replacement of natural limbs or parts of limbs with advanced prosthetics. The practical reason may be achieving increased resistance to damage, strength, range of

motion, or precision. Such improvements can be possible with durable non-wearing materials or structures that increase strength and flexibility of movements. Together with touch, nociception, proprioception, and balance enhancements, further development may provide prosthetics more advanced than natural limbs in many functional aspect. Motor enhancements can also concern the speech apparatus allowing, among others, changing of voice scale or timbre.

B. Additional Senses and Motor Skills

Additional senses and motor skills mean that the human body could be equipped with senses and motor parts that have not been developed in the evolutionary process.

Applications of the BCI:

The BCI technology may equip the human body with, among others, the following new senses:

Echolocation: The use of ultrasound and the acoustic echo phenomenon to orient oneself in the environment and to detect objects.

Magnetoreception: Detection of the direction of the Earth's magnetic field lines to orient oneself and navigate in the environment[3].

Electrolocation: Detection of changes in the electric field to orient oneself in the environment and to detect objects.

Infrared and ultraviolet wave reception: The use of electromagnetic waves to orient oneself in the environment and to detect objects.

Additional chemoreception: The analysis and detection of the chemical composition of a substance through touch and smell with chemoreceptors that don't occur by default in the body.

Motor Skills: The BCI technology with additional prostheses and sensors may enhance human motor skills, provide broader, more efficient movement skills as well as better interaction with objects. Such extension could be used to enhance natural skills in various environments including underwater and airspace conditions. It is worth noting that the physical requirements of such additional extensions (e.g., weight, required space, energy need) make them more reasonable in use as ad hoc, pluggable solutions rather than permanently connected to the body.

XII. Perceptual-Motoric Area: Remote Reality

Remote reality (RR) involves indirect (not linked directly with the body) interaction with the environment using sensory and motor devices. Remote interaction can take place using one or multiple perceptual and motor channels, depending on the needs. Thereby, it provides a partially or fully immersive experience of being and interacting with distant places from the body. The distance can vary greatly. In most cases, it will range from meters to hundreds of thousands of kilometers. In theory it can be much farther, bearing in mind the delays in communication due to the speed of electromagnetic waves. Emerging communication technologies such as the Starlink satellite Internet constellation and the 6G infrastructure planned for the 2030s will make it possible to

exchange huge data packages between any location on the planet and also in its closest surroundings with negligible transmission delays.

Applications of the BCI:

Remote Subject Interaction (RSI): This covers the interaction between subjects (especially humans) in a remote manner. At the later stage, it can also cover communication with members of other species if needed. Nowadays, remote interaction between subjects takes place via external devices (such as smartphones, laptops) using voice and video calls among others. RSI using the BCI can be perceived as an evolution of current telecommunication methods. However, in contrast to technologies we currently use, BCI can engage many other senses, including touch, taste, smell, and motoric channels of communication. Ultimately, the quality of remote communication can be indistinguishable from how we currently collect information from the environment using multiple senses.

Remote Object Interaction (ROI): This covers the interaction with devices (objects) in a remote manner. ROI can take place thanks to communication modules, sensors and/or motorized capabilities of remote devices. Using perceptual-motoric channels, its enables complex interactions with objects thousands of miles away. The perceptual channels allow perceive remote environment and send data to the BCI implants and ultimately to our senses. In turn, thanks to the motor functions of remote devices, it is possible to dynamically interact with distant objects in real time. The BCI implants will allow us to use all types of remote devices including ground, watercraft, and flying objects in a immersive, multi-sensory manner. It should also enable control

of all kinds of future generation robots, such as humanoid robots as well as less human-like medical and industrial devices.

Potentially Significant Applications of Remote Reality Area in the Coming Future:

Shared Perception and Motor Skills: An important scope of the remote interaction applications can be sharing perception and motor skills across multiple subjects or objects simultaneously. In this concept, a person or device can send data that comes from senses and share them with a number of other subjects/objects in real time. It is worth noting that such use should provide a sufficient, high level of communication security.

Reproduced Perception and Motor Skills: The concept of shared perception and motility may not only apply to live broadcasts. Remote reality also cover sensory and motor reproduction of previously recorded events. A present-day analogy is an audio-video content that can be played at any time and which users can perceive through the senses of sight and hearing. The BCI technology will enable the immersive perception of previously captured events using some or all of the sensory and motor channels. Recorded events could be reproduced in an immersive way any time and any number of times.

Remote Use of Objects: The use of any kind of devices that can be connected to human senses and motoric functions. The communication can cover any kind of devices (drones, vehicles, humanoid bots among others) located at any distance from the body of the person who uses the BCI implants. Remote perception and movement may lead to significant transformation in the way we explore the environment.

Remote Tourism: This covers the use of remote devices for general tourism purposes. Interacting with devices thousands of miles away from the body will allow exploring the far environment with unprecedented ease. It can take place by using our own bots as well as by renting bots from professional rental services or other private individuals who don't use their devices at the time. In the future, it may become as popular as today's forms of tourism and allow us to explore any region of our planet or even space incomparably more efficiently and safer.

Remote Work: This covers the use of remote devices for professional purposes. The remote devices can be physically thousands of kilometers from the operator's body to perform any work, whether in an office, factory, or hospital. The remote work can be used with humanoid bots capable of performing a wide variety of activities (thanks to broad perceptual and motor abilities) as well as narrowly specialized devices for manufacturing or medical purposes.

Important Differences Between Remote Subject Interaction and the Internet of Thoughts:

The concept of Remote Subject Interaction (which contains broad forms of communication) may appear similar to the concept of the Internet of Thoughts, a component of future Intelligence Augmentation. However, there are essential differences between them. The communication between subjects described in this section deals with transferring sensory data (visual, auditory, and others). In such case, all data must be comprehensible to the receiver using their perceptual and motor skills. In turn, the concept of the Internet of Thoughts assumes the transfer of complex neural network structures. First, they must be

appropriately mapped in the sender's mind, then transferred, and finally interpreted by merging them with existing neural structures in the receiver's mind. As mentioned earlier, it can be challenging to create this type of a universal method of thought mapping/merging.

The communication based on the Remote Subject Interaction concept seems to be much simpler (which doesn't mean simple) to develop than the Internet of Thoughts. This is because it doesn't require the development of thoughts mapping/merging technologies between persons. This makes technical implementation much more realistic in the coming years.

At the end, it is also worth to note that the transmission of sensory representations using RR and complex neural structures using the Internet of Thoughts can be perceived as another forms of the concept we commonly name "telepathy". In the Remote Subject Interaction case, we could speak about "perceptual telepathy". In turn, in the Internet of Thoughts case, we could speak about "thoughts telepathy" or "neural networks telepathy".

XIII. Perceptual-Motoric Area: Digital Reality[4]

Digital Reality (DR) involves interactions through the perceptual and motoric communication channels with digitally generated environment. The interaction may take place using one or more senses and motoric channels - in a partially or fully immersive way. Today, there is no solution for complete immersion of human perception and motor skills in a digital environment. Despite this, attempts are made to create solutions to provide at least a partial immersive experience via external devices[5]. Such solutions, however, cannot convey the senses of proprioception, balance, or body movements in a way that is unconstrained by the limitations of the place (e.g., a living room) in which the user is located.

Applications of the BCI:

If we take the BCI technology into account, there is every reason to change the above-mentioned state of affairs. As with RR, it boils down to the BCI communicating with areas of the brain responsible for perception and motor skills. The use of all channels of interaction with the environment can provide immersive access to unlimited, digitally generated spaces. These new areas can be explored and adapted for any purpose. The perception of digital places through the senses of sight, hearing, touch, proprioception, balance, smell, and taste can provide a sensation of immersion analogous to that of everyday life. The textures of objects, smells, tastes, and the feel of the digitally generated equivalent of the body (an avatar) can be identical to the sensations we currently experience with our physical body. Ultimately, the feeling of being in these new spaces can be qualitatively indistinguishable from how we perceive our world today.

Potentially Significant Applications of Digital Reality Area in the Coming Future:

Immersive Digital Environment for Human Interactions: Immersive DR technology can solve the problem of distance separating individuals. Unlimited spaces and locations can serve as places to visit with other people, regardless of the physical distance. This can enable joint cooperation, exchange of ideas, and foster the integration of individuals. Interactions with the environment and other subjects can take place using avatars or any other form that a person can embody.

Immersive Digital Environment of Collaboration, Research, and Development: DR can be widely used for R&D purposes

making it easier for teams to collaborate. Immersive experience can enable people to work in a generated location like offices or research centers. This will facilitate the quality of communication and support social integration of team members[6]. DR can also make it possible to use intuitive digital workspaces to design material objects. Required simulations and tests can take place in the digital environment and then final results can be materialized in the analog world, if necessary. With this kind of digital pre-production environment, implementing new solutions into the so-called analog space can become much more efficient than today.

Immersive Digital Environment as a Source of New Space and Goods: DR can enable the creation of new, unlimited spaces, complex locations, buildings, and any other objects. Furthermore, the consumption of created goods and services in such digital, immersive environment can be completely cost-free and thus accessible to everyone. The human body requires physiological needs to be met, such as food, water, excretion, and movement. However, a huge proportion of human needs are of non-physiological character. Many of the higher-order goods that we know from the analog world can be generated and then consumed in the digital world indistinguishably with full sensory immersion[7].

Shared/Reproduced Perception and Motor Skills in Digital Environments: The concepts of shared and reproduced perception and motor skills described in the context of remote reality can also refer to DR. In this case, however, the interactions will take place in the digital space and will refer to its individual objects. The sharing or reproduction of digitally generated objects and devices can be an important part of DR and it can be widely used among others for gathering knowledge and training skills.

XIV. Programming of Non-Autonomous Functions

The Problem of Prolonged Body Inactivity

Both remote and digital reality used for a few hours or more every day can lead to health issues. Since both RR and DR can absorb our consciousness into the process of moving remote/digital objects and avatars, body movement in analog space is severely limited or completely impossible at this time. This state can be compared to the ordinary process of dreaming, when we move in a dream space, while our body is resting in bed. If DR or RR will be used too long, they can lead to neglecting physical activity and, consequently, to serious disorders, including cardiovascular disease, obesity, muscle weakness, and mental disorders.

Programming of Non-Autonomous Functions

The problem of prolonged body inactivity can be solved by a concept I call Non-Autonomous Function Programming (NAFP), or more colloquially, Body Programming (BP). The NAFP concept aims to solve this problem by automating motor and perceptual functions while using RR or DR. The automation (programming) process is based on the following steps.

1. **Learning of Non-Autonomous Functions**: The first step consists of performing everyday activities, such as eating, drinking, washing, sunbathing, or fitness routine. Meanwhile, all movements of the motor system and stimuli coming from the perceptual systems are recorded by the BCI implant, transmitted to the internal (additional internal memory of the implant) or external system (e.g. a smartphone) and memorized to be played back later.

2. **Automation of Non-Autonomous Functions**: The patterns of actions memorized in the first step can then be replayed and executed in an automated manner at our request. Crucially, the process of performing learned tasks already takes place without our attention. During this time, consciousness can be entirely focused on moving and perceiving information from the RR or DR. Although the receptors of sight, hearing, or touch of our body at this time aren't the source of information received by consciousness, they can continuously collect data from the environment to support accuracy of current activities and safety.

3. **Continuous Improvement of Non-Autonomous Functions**: Additional information gathered in real time from body sensors can be used to permanently improve previously learned tasks. Ultimately, the level of proficiency in performing them can become much higher than the skill originally acquired. NAFP can thus enable the continuous refinement of tasks that have been pre-recorded to complete them in a more efficient manner and also to improve the so-called muscle memory. As a result, the execution of given actions performed in a conscious manner is also done more efficiently than at the initial stage.

While our consciousness will be engaged in a particular activity in the RR/DR, NAFP technology can take care of the needs for the body to function. Programming can include daily activities, such as eating, drinking, and hygiene. However, it will allow us also to perform physical activities such as maintaining general fitness in a way that is completely free of our focusing on them. Thanks to these automated daily activities, the body's fitness can significantly improve, reaching an optimal overall efficiency

level. As a result, NAFP can significantly improve overall physical as well as mental health.

XV. Closing Remarks

The scale of potential applications of the BCI can be vast. The wide range of medical treatment in each of the areas described offers high hopes for solving or alleviating many existing diseases and dysfunctions in the human body. Nonetheless, it is only a part of the whole spectrum of applications. Each of the four major application areas could lead to numerous non-medical uses. Specific fields of use carry a significant potential of impacting the life of a single individual, whole societies, and the entire human civilization. In effect, it can impact our future in significant ways. What is crucial here, the impact can be both negative and positive. In the coming years, the BCI has the potential to both increase and decrease the upcoming threats of humanity. That is why it is crucial to assess the impact of BCI applications on our future. We will explore this potential by assessing the opportunities and threats of the BCI technology in the context of all the major existential risks.

PART V:

Analysis of BCI Applications

I. Introduction

The previous part aimed to answer the question: Could the BCI technology have anything else to offer beyond Intelligence Augmentation (IA)? As we've seen, there is a wide spectrum of potential applications in many areas. Since the answer to the question above is positive, I'd like to move on to the question posed earlier: **Could the BCI technology have the potential to de-escalate existential risks despite the threat of IA?** If we want to try answer this question, we need to evaluate the potential of each BCI area to identify the opportunities and risks of their implementation in the coming future. This will be the essential goal of this part. In the first step, I'll introduce the methodology of the analysis.

1. Subjects of the Analysis: BCI Application Areas

In the previous part, the following six non-medical application areas of the BCI were presented:

- Intelligence Augmentation
- Emotional Regulation
- Intrasomatic Enhancement
- Close Reality
- Remote Reality
- Digital Reality

The above six areas will be the subjects of the analysis. On the other hand, medical areas of BCI applications will not be discussed in this place. This analysis does not aim to evaluate the strictly medical influences on overcoming specific diseases. However, worth to note that the development of each of the sixth non-

medical areas certainly will impact the health of many patients and induce the development of specific fields of medicine.

2. Assessment of Opportunities and Threats

Each of the above six BCI application areas will be analyzed and evaluated in terms of the impact on existential risks. The analysis process will be presented from the above perspectives:

- Opportunities for de-escalation
- Threat of escalation

The analysis for above perspectives will be summarized by a numerical assessment that quantify the scale of opportunities and threats:

0 – none/insignificant
1 – very small
2 – small
3 – medium
4 – large
5 – very large

3. Analyzed Existential Risks

Analysis of opportunities and threats will be performed for existential risks:

- Technologies fraught with mass destruction risk (TFMDR)
- Environmental degradation
- Misaligned AI/IA

The attempts made so far to estimate the probability of particular disasters on a global scale indicate that in the perspective of the upcoming decades, threats covered by those three types of existential risks will be the most severe[1]. In contrast, the disasters caused by natural (non-anthropogenic) hazards are very unlikely. The probability of such an event is estimated to be several hundred to several thousand times lower than for the major anthropogenic existential risks. Depending on a study, the probability of a natural cataclysmic event threatening civilization falls between 0.01 and 0.05% for every 100 years[2]. Therefore, the natural risks will not be considered in the analysis.

4. Other Assumptions

- All three types of risks are treated with equal, highest possible priority. Although existing attempts to estimate, it should be highlighted that all such evaluations of the probability for anthropogenic existential risks are accompanied by a significant degree of inaccuracy. The reason is the strictly anthropogenic nature of these threats. Whether they escalate or de-escalate will strictly depend on our decisions and actions in the coming years.

- The analysis covers the development period of a given BCI area up to year 2050. However, it should be noted that we may reach the usability of individual BCI areas well before this date.

- The developed technologies of given BCI area will be free of technical issues (both hardware and software) that could result from oversight in the manufacturing process of the

technology. The analysis assumes that the weaknesses will be eliminated for a mature, commonly used technology.

- Only one of the discussed BCI application areas will be developed in the same period. The goal of this approach is to optimize our potential in such a way as to create technology that is as refined and reliable in every aspect possible and in the shortest possible time. It is important to emphasize that such an approach decreases negative as well as positive cross-influence between technologies of different BCI areas.

II. Intelligence Augmentation

A. TFMDR

Opportunities:

Human intelligence augmented with IA can bring significant advances in numerous fields, leading to the development and large-scale application of many breakthrough technologies. One of the most important may be achieving much cheaper, and as a possible result completely free, producing of energy. The technologies may be based on renewable sources, nuclear fusion, or new, currently unknown, ways of obtaining energy. Another type of technology developed with IA can significantly increase crop yields and food production boost food availability, or even end global hunger.

In fact, similar technical progress can apply to all other industries areas. Accelerated development on many fronts can lead to automation of production and multiplied productivity. With advances in science, it will be possible to change processes of

manufacturing goods not only to renewable energy sources but also to fully recoverable, renewable sources of raw materials needed in production. This, for example, can occur through advances in materials engineering or molecular chemistry. If all of the above changes were to take place on a global scale, this could level material inequality. If we brought our entire civilization up to the level of relative material wealth (while being fully aware of the a priori and subjective nature of such an indicator), we could reduce the problem of social tensions. All of the above can ultimately have a significant impact on reducing the threat level of using TFMDR in conflicts.

The widespread augmentation of human intelligence can allow development of better safeguards for both civilian and military technologies. The progress in this area can be made by developing more reliable safety systems to keep harmful pathogens and nanotechnology out of the environment. It can also take place with respect to safeguarding chemical and nuclear weapons, including the technologies needed to transport them. Conversely, if a threat agent were to escape controlled conditions for any reason (even through intentional action), new technologies and countermeasure strategies that would be more sophisticated than current ones could more effectively neutralize its harmful effects on the environment.

Dissemination of the IA technology evenly and on a large scale can support the process of correcting thought patterns and mental maps. Increased intelligence capabilities can also reduce unwarranted prejudice against people with a different worldview, appearance, or behavior. Perhaps with IA, we can develop more widely accepted core social values and principles, independent of a person's origin and experiences. On issues where we'll still have differing points of view, we'll be able to better understand various opinions because of our greater capacity for analysis. All of this can bridge fundamental and conflicting attitudes. Ultimately, we'll

be able to reduce intangible differences that may lead to social tensions constituting the important factor influencing existential risk from TFMDR.

Assessment of opportunities: 5

Risks:

The distribution of IA either within a narrow group (e.g., a privileged social group) or more broadly (e.g., citizens of only one country where the technology is use) can result in serious risks. The capabilities gained with IA can be reflected in increased levels of technological sophistication in any sphere and, as a result, giving more power to such a group. In a short time, this can bring irreversible hegemony in all key areas of human activity. The unequal distribution of IA may have dire consequences, including control over an underprivileged part of humanity, persecution, restriction of human rights, or even total extermination of groups deemed unnecessary or obstructive to the stated goals of the caste enhanced with the IA technology.

IA distributed both evenly and unevenly can initiate the development of much more advanced and potentially more dangerous generations of new TFMDR. In particular, this may be especially risky in the case of the rapid development of nanotechnology and bioengineering. Creating autonomous and self-replicating organisms based on traditionally understood biology (e.g., viruses) as well as compounds such as silicon or graphene can lead to them getting outside controlled environment or their mutation or self-upgrading, which may be difficult to control and stop. IA can also increase the number of people able to use TFMDR in a skillful manner. There is a risk that despite higher precautions, the above-mentioned trend may raise the

already growing risk of a global disaster. In the context of tensions between entities, for example, international actors including states competing for dominance, many of the safety recommendations may fall far short. The situation seems to be particularly worrisome in the face of the growing arms race. The pressure of overtaking the opposite side can result in the omission of many safeguards that would normally be considered essential.

From the perspective of people who aren't supported by the IA technology, the privileged group can be seen as a threat not only to their freedom, but even their life. This, in turn, can arouse reasoned and strong opposition, foster social tensions, and finally escalate the risk of conflict on both sides. Such tensions can arise early in the development of the IA technology. These can occur between the general public and government or private organizations as well as internationally between individual countries developing the technology. All of the above can intensify the total level of social tension and increase the threat of using particular types of TFMDR in a conflict.

It is important to note the aspect related to the hope of IA solving the problem of differences in worldview within society. If IA were evenly distributed throughout the society, we could wish that this technology would effectively solve the problem of social polarization and of entrenching individuals and entire groups in emotionally comfortable worldviews. If we assume that society, as a whole, is evenly increasing in intelligence to a significant degree, this should apparently be enough to revise and correct previous views. In this reasoning, however, it's easy to forget about the extremely important emotional sphere on which human perception of reality and behaviors are highly based. In other words, revising one's views and behaviors can be a hard process, not so much because of insufficient analytical skills, but because of fear of the emotional distress that the change of deeply held

beliefs may bring. For example, if someone has spent years building their worldview on an ideology that is directed against people with different beliefs, it may be highly difficult for them to accept that the previous value system and the acts that such a person committed as part of their beliefs were in fact wrong. In such a case, the first signs of cognitive dissonance and the associated psychological distress may lead to a repudiation of these ideas and effectively discourage the person from further confrontation and analysis of uncomfortable elements of their worldview. All of the above may imply that IA can be far from sufficient technology to revise erroneous beliefs and judgments about surrounding reality. The problem isn't the lack of available sources of information and strictly analytical skills, but emotions, which can sabotage the actions of the analytical part of the brain.

Assessment of risks: 5

B. Environmental Degradation

Opportunities:

As mentioned earlier, IA can bring about a significant acceleration of civilizational change including technological advances in almost every area of our lives. With IA, we can change industries to be more efficient with regard to the use of energy and resources without sacrificing the production of the goods we need. A revolution in the generation of cheap, or even free energy and the manufacturing of goods from environmentally neutral raw materials may allow us to significantly reduce our negative impact on the Earth's ecosystem. Thanks to the significant acceleration of progress in science and engineering, we may be able to develop effective technologies to offset negative environmental changes in

a short period of time. These can include solutions that absorb pollutants from the atmosphere and hydrosphere, reduce CO2 levels, or even develop genetic engineering efforts to restore population balance among endangered species. The latter may be necessary to maintain stable food chains in ecosystems.

The advances that can occur as a result of having renewable resources, free energy, and automation of production, can lead to a gradual decline in the price of goods necessary for existence and, eventually, to free access to them. The new, much more stable, safe, and comfortable living conditions, together with growing social awareness and increased access to knowledge about what happens on our planet can bring about changes in lifestyle. If we become more aware of what we buy and eat and if we use our possessions in an eco-friendly way, we can have a significantly positive impact on the environment globally. The above-mentioned transformations can allow us to live in greater harmony with the ecosystem which we are part of.

Assessment of opportunities: 5

Risks:

Uneven distribution of IA technology and concerns of particular social groups or countries can cause serious tensions. As a result, one party may use the technology with an irreversible impact on the ecosystem. In both scenarios of the uneven and even distribution, IA dramatically increases the number of entities capable of developing a broad assortment of technologies that can eventually be used within the Earth's environment and be a factor in increasing the likelihood of intentional or unintentional disaster. These types of threats include the TFMDR already discussed: nuclear and chemical weapons, synthetic biology as well as

nanotechnology. The latter two may turn out to be particularly worrisome for all living organisms. The progress in these areas induced by the arms race of the largest entities may result in the development of intelligent pathogens or synthetic micro-devices that can affect the functioning of many animal and plant species, leading to catastrophic changes in the Earth's ecosystem.

Looking at humanity's past actions, one cannot be sure that once people are enhanced with IA technologies, they'll change the largely anthropocentric attitude toward the Earth's ecosystem. Those backed up with this technology may just as well decide to change the world in a direction which is beneficial from their point of view. An extreme example here can be a single Earth-dominant entity or group of entities which use IA to transform all available resources and space on our planet into a powerful technical infrastructure, for example, to multiply their own potential and expand to other celestial bodies. In such a case, one can speak of the problem of the relativity of values, which for suitably intelligent beings may be quite different from our own as well as far from respecting the biodiversity of earthly life.

It's worth noting that even today it seems that humanity largely adheres to a double-standard with respect to life on our planet. Most people won't hurt a dog as it appears to be similar to humans to some degree because of its behavior. At the same time, many people eats meat just because pigs, cows, and other cattle animals seem less human to us than dogs. This is even more clear to note in the case of killing ants and other "less human" creatures from our point of view. Even among some eco-activists, there is an anthropocentric conviction that we, as humans, should live in harmony with nature, not because our "moral code" demands it, but because it's only through that harmony we humans will be able to continue to exist. We are pragmatic, but not necessarily generous to the vast majority of living beings. In the reality of

immense possibilities of transforming the world around us thanks to IA, including increased or even complete self-sufficiency from other species, the question arises: will peace between highly intelligent, self-sufficient individuals and nature take place? Although it may be hard to imagine, the ecosystem changes we cause today may be far lesser than the future intentional actions of entities supported by the capabilities of the IA technology.

Assessment of risks: 5

C. Misaligned AI/IA

Opportunities:

Wide distribution of IA can support the goal of keeping ahead of increasingly advanced AI technologies in the long term. It is possible that IA-enhanced people will be able to design better and safer AI-based technologies. It's worth noting that the IA technology will evolve and, eventually, the line between "artificial" and "biological" intelligence may fade. The systems that in the case of "ordinary" AI would be separate and autonomous will become a part of ourselves thanks to IA. Because of that, we may not need to give AI even partial control over critical civilian and military systems. From a decision-making perspective, it will be still our species who can be responsible for controlling all elements of critical infrastructure.

As mentioned earlier, IA can bring a significant technological progress and then the development of free and renewable sources of energy as well as resources. This can significantly increase the availability of cheap and potentially free goods available to everyone. This can help us resolve conflicts that are based on tangible differences. On the other hand, changes in the sphere of

human mentality thanks to higher analytical skills can reduce intangible differences. All this can be an important factor in reducing social tensions. Eventually, lower levels of tension can reduce the risk of conflicts, arms race, and ultimately the emergence of an misaligned superintelligence that can lead to the limited freedom or even annihilation of our species.

Assessment of opportunities: 5

Risks:

Selectively-distributed IA poses a significant threat to the rest of humanity lacking similar technology. The selectively created and implemented IA technology can enable a critical point in the self-improvement process to be crossed at a relatively early stage of development. As a result, an entity using IA may be able to make exponential progress and create increasingly powerful generations of itself by leaps and bounds in even shorter time intervals. In the end, this can bring about an IA-based superintelligence capable of supremacy over any other entity. The mechanism of exponential progress may concern the increase of advancement in the field of implants, next generations of AI algorithms that are an essential part of IA technologies, as well as available computing power, effectively multiplying predictions resulting from Moore's law.

Dynamic advances in IA can set in motion further development of powerful AI systems. Such systems can bring about many new devices supported by the powerful capabilities of AI algorithms and further the development of next generations of both military and civilian bots. These devices can be many times more intelligent or at least more effective than most humans in achieving the goals set by entities using IA, for example, due to greater strength, precision, discipline in carrying out orders, or the

power of weapons at their disposal. This new generation intelligent devices, constantly growing in number and capabilities, controlled by those using IA, can marginalize the importance and limit the freedom of the humanity. Ultimately, it may end up eliminating every homo sapiens being that is “unnecessary” or “highly harmful” from the point of view of the entities using powerful IA technology.

Assessment of risks: 5

III. Emotional Regulation

A. TFMDR

Opportunities:

The BCI technology for emotional regulation can advance our ability to perceive and analyzing phenomena around us in a more balanced way and make us less susceptible to extreme emotional states. This can significantly improve our self-reflection and self-correction skills with regard to experiences, beliefs, and behaviors. For example, the people who spent time on activities that are subjectively exciting but in the bigger picture irrelevant can look at their past actions in a more critical way, without potentially extreme and painful emotions. This will allow less self-aware and emotionally resilient people to redirect their potential into activities that are more priority. On the other hand, the problem of some people isn’t limited introspective abilities, but the inability to derive satisfaction from doing specific activities, which leads to procrastination and complete abandonment. In this case, better emotion management can also help keep focus on the important

activities. This can be crucial especially in the initial stages of changing current patterns of behavior.

Emotional stability can also have an impact on improving interpersonal relationships and allow better quality of cooperation in the key fields of human activity, both personal and professional. All of the above EB technology opportunities can improve the focus and quality of activities in priority fields. Allocating our potential to crucial areas can increase the intensity of work under many desired technologies. Focus on the key fields such as new, low-cost energy sources and renewable materials can finally optimize production processes, costs, and lead to wider availability around the globe of all the most essential goods. Eventually, this can have a significant impact on decreasing social tensions based on the tangible differences we face in today's world. Moreover, this may considerably reduce the threat of conflict and the existential threats associated with TFMDR.

As mentioned in previous sections, one of the key problems of our time isn't the lack of information sources and analytical skills, but the human's internal "defense" mechanism which strongly influences on our emotions and therefore our actions. Filter bubbles, social polarization, and radicalization of worldviews phenomenon's originate largely in this mechanism. We fear the destruction of our idea of the world, which we have nurtured in our minds over many years. The widespread adoption of the EB technology can help us change our thought patterns and mental maps. On the basis of facts and without extreme emotional impact, we can try to develop a much broader and socially acceptable set of views and principles, independent of a person's background and past experiences. On matters where we'll still have diverse points of view, we'll be able to better understand different opinions. It will be easier to take the nuances of each side into account and accept them without an extreme emotional reaction. EB can also

have a positive impact on the resilience to any kind of social engineering, such as propaganda or emotional addiction to some media. All of the above can help in building a much clearer and more accurate view of the world. Eventually, this can reduce social tensions based on tangible and intangible differences, which are major factors of existential risks, including threats from TFMDR.

Finally, it should be noted that EB is relatively collective in nature. The positive impact of that technology on our reality will depend on how widespread its use becomes in society. Emotionality plays a huge role in creating stable, socially beneficial relationships, building widely accepted consensus, compromise, and a healthy and well-functioning society. EB increases the chances of building a new level of emotional maturity in a broad social context. This collective nature can be a huge benefit, because society should desire that as broad a group as possible gains from its capabilities. This is very important because it helps reduce the risk of selective distribution within a narrow group of people. If a small number of people use it, the potential will be incomparably lower than when used by millions. The more people want to use the technology, the more effective it can become in minimizing existential risks.

Assessment of opportunities: 5

Risks:

Although the EB technology is collective in nature, there is still a risk of its selective distribution. Emotional management skills increased among a limited group of individuals can to some extend bring about better prioritizing, focusing, and collaborating in a given area of technological development. For example, focusing on the achievement of particular goals only in a given country or

within a selected group of people can result in social tensions and advancement in an arms race of TFMDR. Such a situation can additionally negatively impact the process of implementing and maintaining safeguards against existential risks. The above may eventually increase the risk of both intentional and unintentional TFMDR incidents.

The distribution of the EB technology in society, wrongly understood and distorted by influential, potentially aggressive group such as a totalitarian state, can lead to social tensions. For example, this may happen when an oppressive government will forcefully attempt to impose the use of that technology. Such a situation can be especially alarming in the case of an online form of EB that allows third-party control. This can raise social tensions and dangers of conflicts, which may escalate even to the use of TFMDR against the opposing side.

Also, it's necessary to keep in mind major the capabilities of EASD area, including its potential to feeling on demand emotional extremes. The use of this type of technology can lead at least to psychological but also physical addiction, as is currently the case with stimulants such as amphetamines or depressants such as alcohol. On a macro level, the broad use of EASD can have a negative impact on the quality of functioning of the entire society in the long term, like it does with some drugs. This situation can negatively affect emotional stability, reducing the ability to make sober assessment of reality, increasing mood swings, or vulnerability to external influence. All of this can intensify existing social tensions and create new ones, increasing the risk of conflicts where TFMDR can be used. In addition, if EASD is used by persons working on the development of such technologies, this can cause their highly improper implementation. The worrying risk also applies to the persons who overuse EASD and may influence the usage of any civilian or military TFMDR (e.g., state

leaders or military personnel). In such situations, it can cause threatening incidents.

Assessment of risks: 3

B. Environmental degradation

Opportunities:

The widespread use of the EB technologies can help reorient and intensify human activities into priority areas for the health of the Earth's ecosystem. In practice, this can mean work intensification in areas of improving and developing new types of renewable, green energy, and sources of materials. This can enable the production of essential goods for the humanity in a more sustainable and finally neutral way for the stability of the Earth's environment.

EB may also allow us to look dispassionately at the problems which, like ecosystem deterioration, are unfortunately still pushed out of mind by many people. Greater emotional stability and capacity of introspection may bring about broader awareness of the challenges facing our species and later on help us engage more broadly in remedial action. EB can cause deep mental and lifestyle changes by abandoning destructive patterns of thinking and behavior, such as linking self-esteem with material possessions, excessive consumption of goods, or wasting energy and food. In the end, all of this can have a significant impact on the local environment and, finally, on the global ecosystem.

Assessment of opportunities: 5

Risks:

Selective distribution of the EB technologies within a narrow social group carries the potential risk of increasing social tensions. This, in turn, can lead to conflicts with destructive effects on the Earth's environment in extreme scenarios. However, it must be highlighted again that EB is largely collective in nature, which should be an important factor for reducing the risk of selective distribution. We also need take into account the threat of a conflict when a group wants to impose the use of this technology on others. In this case, escalation may also be possible, leading to negative consequences for the ecosystem.

Assessment of risks: 2

C. Misaligned AI/IA

Opportunities:

The EB technology through its ability to better manage emotions and to facilitate deeper introspection of existing beliefs and behavior has the potential to redefine directions of work on possibly dangerous applications of both AI and IA. A redirection of existing activities among those currently involved in their development can take place toward making the applications safer or refocusing them into completely different, more sustainable fields of science and engineering. Additionally, the activities that aren’t so far involved in the AI/IA development can redirect their efforts to increase public understanding of those risks or improving the safety.

EB can be a technology that corrects many of our present beliefs about reality, changes personal worldviews, and increases

human dialogue and cooperation. If EB benefits were to have a widespread impact, regardless of region or social group, they could de-escalate many current social tensions based on intangible inequalities. Moreover, EB can help us refocus on more significant endeavors, including the development of key technologies for our future (renewable sources of energy and raw materials). Together with the increased social awareness and responsibility for the common good, they can increase the availability of all basic goods across the globe. The above-mentioned changes, both in the tangible and non-tangible dimension, can significantly reduce social tensions between different groups, including countries, leading to de-escalation of the still growing arms race. Finally, this can decrease an existential risk with regard to the creation of misaligned AI/IA technologies by individual entities or its highly negligent development under time pressure.

Assessment of opportunities: 5

Risks:

Distribution of EB to privileged social groups or in individual countries can lead to social tensions both on a local and global scale. Such a situation may end in an intensified arms race including the development of selective AI/IA systems which are misaligned to the expectations of humanity at large. Those systems would be created to ensure the achievement of political and military supremacy over others.

If the EB technology is imposed on a society through forceful methods, for example, by a totalitarian government, it can heighten social tensions. This is particularly risky for the online form of EB, which can be controlled by third parties. Such a situation can escalate conflicts both locally and globally and result

in the intensification of works on other technologies, including selective AI/IA systems, to expand the advantages.

The widespread use of EASD technology also can impact social conditions, worsening tensions, and increasing the risk of conflicts where misaligned AI/IA could be used. The threat of emotional instability related to the use of EASD among persons developing such technologies can result in their improper implementation. The use by persons who maintain or apply narrowly distributed, powerful AI/IA can also pose a danger of threatening incidents.

Assessment of risks: 3

IV. Intrasomatic Enhancement

A. TFMDR

Opportunities:

IE technologies can improve the body's default fitness level, resistance, and ability to adapt to the external environment. The applications cover, among others, the enhancement of circulatory, respiratory, lymphatic, immune, or endocrine systems. IE may translate into better efficiency, improving human capabilities both in professional and personal life. It’s worth noting that the advantages may take place not only in the intrasomatic sphere but also indirectly in the general well-being and psychological sphere, as they are immanently related. Assuming widespread availability of IE on a macro scale, in any social groups and regions, such technologies may improve general health and effectiveness of functioning of the entire humanity. In an optimal scenario, this can indirectly translate into accelerated development of key areas such

as renewable energy and resources, optimalisation of production processes, and, finally, wide availability of all necessary goods. If the distribution of IE and all its potential benefits were to take place in a responsible and equal manner, this could have a positive impact on lowering social tensions and decreasing risks of improper development and use of TFMDR.

Assessment of opportunities: 2

Risks:

If the IE technologies were distributed in a selective manner only within a privileged social group or country, it would improve the lives of a limited number of people and in the long run, increase advantage in any chosen area of their activity, including civil and military ones. The selective availability of IE and the potential to give an advantage may also stir serious social tensions both on a local and global scale. This can escalate conflicts intra- as well as internationally, leading to an increased risk of disordered development and dangerous use of TFMDR.

It is important to note that the IE technologies are largely individual in nature. They can significantly improve the quality of functioning of individuals in a society and, at the same time, which doesn't have to translate into the benefit of the whole society. Moreover, the IE technologies can create a sense of superiority over other persons, which can be a highly risky factor in escalating social tensions. Ultimately, this can cause conflicts where, among others, TFMDR will be developed and used.

Assessment of risks: 4

B. Environmental Degradation

Opportunities:

By increasing the general body fitness level, IE can affect the efficiency of work in many areas. For example, this can support the development of the above-mentioned technologies such as green energy sources or more environmentally-friendly ways of obtaining raw materials. This may result in more sustainable production of many goods, reducing the burden on the Earth's ecosystem. The positive impact on the environment can mainly take place if these technologies are widely available and support escalating social tensions. This can decrease risk of a hectic arm race that may have negative effects on the ecosystem as well as bring about the use of environmentally dangerous technologies during conflicts.

Assessment of opportunities: 2

Risks:

Selective distribution of the IE technology only within privileged groups or particular countries can induce many benefits for those who use it. This state of affairs may raise objections of other groups and increase social tensions both locally and globally. The risks may also result from the highly individualistic nature of IE, potentially leading to discrimination against those who do not use this technology. The escalation of conflicts may redirect human potential to areas that are incompatible with preserving natural environment. This can slowing down or suspending the implementation of sustainable energy and resource technologies development plans. Additionally, in order to get advantage over

the opponent, conflicting parties may use civilian or military technologies with a negative impact on the Earth's ecosystem.

Assessment of risks: 3

C. Misaligned AI/IA

Opportunities:

The widespread use of IE in society by enhancing the overall human fitness level and performance can make our species more competitive against AI/IA and the technologies that apply their potential. Furthermore, the IE technologies used thoughtfully and focused appropriately can improve human progress in the areas which are critical for our future. Further, the technology can reduce the overall level of social tensions, both in the tangible and intangible dimension. This can eventually support us in preventing a technological race that can result in misaligned AI/IA.

Assessment of opportunities: 2

Risks:

The highly individualized nature of IE technologies and the selective distribution within privileged social groups or countries can increase social tensions intra- and internationally. This may escalate local conflicts based on social divisions as well as between international entities which try to get advantage over others. Such a situation may foster the escalation of the technological race and bring advanced generations of AI/IA to ensure supremacy over other entities. Such a race may strongly

negatively influence the widespread distribution of advanced solutions in society and their safe implementation.

Assessment of risks: 4

V. Close Reality

A. TFMDR

Opportunities:

The CR technologies can improve the quality of humans' in-born perceptual-motor skills. They can also extend our skills thanks to completely new channels of perception and motor capabilities. The advantages may translate into higher efficiency in the exploration of the environment and enhance human potential both in the professional and personal life. If CR is used wisely, it can have a positive impact on the general efficiency of our species and, to some extent, accelerate the development in given areas. In an optimistic scenario, this can support technological development and lead to broader availability of goods for humanity. If CR were to be distributed broadly, independently of social groups and place, it could help reduce social tensions and lower the threat of inappropriate development and use of TFMDR.

Assessment of opportunities: 2

Risks:

Unequal distribution of the CR technologies only among specific social groups or in individual countries may create a strong advantage in selected areas of human activity. This can have

significant implications and raise social tensions both on a local and global scale. The CR solutions such as enhanced perceptual and motor skills are largely individual in nature. They can improve the quality of functioning of persons, but, at the same time, they may not bring benefits to the rest of the society. For example, additional perception channels and motor skills of privileged entities, large groups or nations supported by such improvements may increase their technological and capital advantages without translating into significant benefits for others. Additionally, it can create a sense of superiority over others. All of the above can to increase social tensions, both on a local and international scale. In the end, such a situation can result in conflicts where, among others, TFMDR will be inappropriate developed and used.

Assessment of risks: 4

B. Environmental Degradation

Opportunities:

By enhancing perceptual and motor skills, the CR technologies can have a positive impact on human functioning and, to some extent, on productivity. In an optimal scenario, it can support human activities in critical fields of development, such as green energy sources, resources, and optimization of production processes, helping reduce the burden on the Earth's ecosystem. The widespread availability of the CR technology may to some extent decrease social tensions, both locally and globally, and reduce the risk of an unpredictable technological race with negative effects on the ecosystem. This can also decrease the risk of conflicts in which environmentally dangerous instruments could be used.

Assessment of opportunities: 2

Risks:

Selective distribution of the CR technologies within privileged social groups or countries can increase advantage over the people without their support and increase social tensions on both a local and international scale. Similar problems may also arise from the highly individualistic nature of the CR, leading to the discrimination of people who don't use these technologies. As a result, it can increase the risks of conflicts and redirect human potential further away from sustainable energy and resource development plans. Moreover, if social tensions and conflict escalate, technologies that can have a negative impact on the Earth's ecosystem may be broadly developed and used.

Assessment of risks: 3

C. Misaligned AI/IA

Opportunities:

The enhanced perceptual and motor skills achieved through the widespread use of the CR technologies can positively impact humanity's competitiveness against the technologies that utilize the potential of AI. This can give us a bit more time to take the necessary countermeasures related with misaligned AI/IA systems. Moreover, the economic, social, or technological benefits achieved with the broadly-applied CR technologies may reduce the overall level of social tensions. This can lower the risk of a disordered technology race and the development of misaligned AI/IA.

Assessment of opportunities: 2

Risks:

The selective distribution of the CR technologies across privileged social groups or countries may result in gaining advantage in selected fields of human activity, including also the development of AI/IA technologies. Moreover, limited availability of CR as well as its highly individualistic nature can bring about social tensions on a both local and international scale. Conflicts and a technological race can further the development of powerful AI/IA systems designed to achieve ultimate supremacy over the opposing side. If the technology is developed in time constraints under social tensions, it will create a rising risk of improper implementation that can end in disaster, including losing control over such system.

Assessment of risks: 4

VI. Remote Reality

A. TFMDR

Opportunities:

The RR technologies can affect the way we move and use the spaces on our planet and beyond. By remotely controlling distant devices such as drones and humanoid bots, we can perceive and actively explore places distant from our body’s location. Our presence in such a place can be possible thanks to immersion of some or all of the body's senses and motor functions, including sight, hearing, taste, smell, touch, and movement abilities. Moreover, there are also other possibilities of RR such as shared or reproduced perception and motor skills that allow experience events either shared by other person in real time or at any time

after they were originally recorded. The all above-mentioned RR applications can allow people to be present in new, even very distant places and interact with different cultures, opinions, and worldviews. The interaction can be active with remote control of bots or passive in a shared or reproduced perception. The opportunities of remote presence at a far location may expand our knowledge about the world and broaden our perspective on many seemingly obvious aspects that can be much more nuanced, complex, or ambiguous upon closer examination.

RR can also strengthen the bond between people regardless of their physical location and promote the need to care for people from other regions as well as humanity in general, leaving behind the particularistic focus on only where we live. Such a change in perspective can induce mental transformation at the level of individuals, nations, and all humanity. We can better understand ourselves and our diversity. This may significantly alter our attitude towards other people and also bring about more aware and closer cooperation between people regardless of their origin. Shared and reproduced perception can also allow us to consume the goods that so far have only been available to a restricted group of people because of their limited nature. The experience captured by one person can be shared both in real time or saved and replayed multiple times later. If we achieve wide availability of the RR technologies, all of the above-mentioned uses can de-escalate social tensions in a both tangible and intangible dimension. Ultimately, this can significantly reduce the risks of a conflict, inappropriate development or use of TFMDR.

The RR technologies can also change the way we gather and share our professional knowledge and skills. Thanks to sharing or reproducing perception, people can assimilate information and new skills in a highly intuitive, much more effective way. Moreover, communication with others through one or more

perception channels as well as remote presence by using bots can significantly improve cooperation. RR can allow for immersive, interactive work of groups of professionals irrespective of their body's location. The changes in human organization, cooperation, and work through the use of RR can accelerate the development in the key fields for the future of our species. With mental changes in the way we perceive other people and places which are physically distant from our location, it can redirect humanity's priorities and intensify our commitment to critical domains. Focusing on crucial areas such as renewable energy and resources as well as the optimization of production processes can bring about widespread availability of all the key goods for all social groups and regions. In the end, all of the above may reduce social tensions and risks of conflicts, inappropriate development and use.

Finally, it is worth noting that the RR area has a relatively collective nature. These technologies can transform human worldview, mindset, organization, and interpersonal cooperation as well as the development in the key areas of the economy and general welfare. If more people communicated, interacted, shared, and gained knowledge and skills thanks to using these technologies, we'd have the potential to significantly increase the total benefits for all of the humanity.

Assessment of opportunities: 5

Risks:

Despite the relatively collective nature of the RR, there is a possibility that technologies of this area can be distributed and available narrowly, for example, only within a privileged social group or specific country. Such a situation can result in growing advantages over the rest of the society in many areas of human activity, both personal and professional. Thanks to improving

communication with other people as well as interaction with remote objects, the RR technologies can be advantageous in the development of TFMDR. Another advantage may be the narrow distribution of the technologies that allow access to remote devices such as drones or humanoid bots. The narrow use across society can bring benefits to specific groups in civilian applications, for example, the development and production processes. Moreover, the benefits may also concern strictly military technologies development, which may increase social tensions. If we decide to develop RR, it's important to ensure that those technologies will be broadly diversified. Otherwise, they may escalate conflicts within individual countries and internationally, increase the arms race risk, chaotic development of TFMDR, and their danger use.

Assessment of risks: 3

B. Environmental Degradation

Opportunities:

The widespread use of shared as well as reproduced perception can reduce consumption of the goods that have a significantly negative impact on the Earth's environment. These technologies can enable immersive (using multiple senses) experience of events and consumption of goods which have been captured once and reproduced millions of times by any one at any time. Of course, this kind of experiencing has its limitations, resulting from the limited influence on the events previously recorded and available for replay. Nevertheless, in cases of many things that we would like to experience, it can be satisfying enough and, what's important for the point discussed here, highly beneficial for our planet's ecosystem.

In addition to experiencing shared and reproduced events, the technology of using humanoid or non-humanoid remote bots will allow us to quickly move to a faraway location and interact with other subjects and places in real time. The same can apply to many current work-related activities, such as commuting to distant offices and business trips. In the end, this possibilities should be much less harmful to the environment than regular travelling by current means of transport.

The RR technologies can also help reach our goals in space exploration in a way which is friendlier to the Earth's ecosystem. Remote bots don't have to be limited to places on the Earth. They can also be applied on the Earth's orbits or on other celestial bodies such as the Moon. If we placed thousands and later millions of RR bots outside our planet, we could achieve our goals more effectively as well as more sustainably. While we'll still need to transport the necessary equipment to the orbit, much of the work related to the operation and maintenance of the expanding space infrastructure can be done remotely, directly from the Earth with remote technologies in an immersive and intuitive way. In effect, many of the short-term space missions that are detrimental to the Earth's environment can be reduced to a minimum. This can be achieved in more ecological way also with currently-developed technologies that make it possible to place non-biological objects in space without emitting harming gases into the atmosphere[3]. As a result, we can be able to carry necessary infrastructure (including multiple RR bots) into space in a more sustainable manner.

Above-mentioned space infrastructure can be also useful for humanity in other ways. The important future use can be cosmic tourism that may be widely accessible by renting bots located for example on the Earth's orbit. Unlike traditional space tourism, this type of exploration can be almost completely neutral to the atmosphere. From a more pragmatic point of view, the immersive

access to distant places may significantly intensify the exploration of celestial bodies like the Moon or asteroids, which are full of resources important for our civilization. The exploration of these cosmic objects may reduce the extraction of certain resources directly on the Earth. By applying this strategy, we may be able to diversify them or even entirely stop the extraction on our planet to drastically curb the negative impact of mining on the Earth's ecosystem. Moreover, in the long term, thanks to the broad use of RR technologies, the production of highly environmentally harmful products and semi-components can be moved to other celestial bodies, ultimately further reducing the negative impact on the Earth's ecosystem.

Assessment of opportunities: 5

Risks:

Despite the relatively collective nature of the RR technologies, there is a risk that they may be selectively distributed in society. This can increase social tensions and escalate conflicts both intra- and internationally. As a result, this situation can lead to the development and use of the technologies that can negatively impact the Earth's ecosystem. Another type of a negative environmental impact can arise from the creation of millions of RR bots, which can involve an increased demand for certain resources to manufacture necessary components. Nonetheless, the environmental impact of manufacturing millions of mobile bots weighing for example 50kg each should be many times lesser than the production of millions of much heavier cars per year or the common long-distance air flights.

Assessment of risks: 2

C. Misaligned AI/IA

Opportunities:

The widespread availability of RR among social groups and countries can bring about a greater diversity of interpersonal relationships, both professional and private. On a professional ground, it may raise public awareness of AI/IA threats and, crucially, improve collaboration to reduce the risks of developing misaligned technologies. In addition, the increased ability to control remote devices for civilian use in manufacturing processes as well as for military purposes on the battlefield may slightly slow down the development of fully autonomous systems and in effect reduce the overall progress of potentially risky AI/IA applications. On a more personal level, RR may limit isolationist tendencies among people and whole nations. The widespread use of RR to interact with people and places thousands of kilometers away can increase our sense of responsibility for the general well-being of distant cultures, places, and all of humanity, rather than just the a region or a country where one lives. This can reduce tensions on non-tangible grounds. Additionally, experiencing goods and events through shared or reproduced perception as well as access to far-away places via remote tourism can increase their accessibility and reduces tangible tensions. Ultimately, eliminating tensions on intangible and tangible grounds can de-escalate local and international conflicts, reducing the risk of developing and using highly threatening, misaligned AI/IA.

Assessment of opportunities: 4

Risks:

The distribution of RR only within particular social groups or countries may bring significant advantages to people supported by such technologies on both a professional and private level. The advantages may grow over time and escalate social tensions. This state of matters can increase rate of the arms race and the risk of developing misaligned AI/IA that can be used in ways contrary to the expectations of humanity. Moreover, if AI/IA is created under time pressure because of an escalating conflict, it can also increase the risk of unintentional or inappropriate implementation. The results may be disastrous, leading to losing control over such systems.

Assessment of risks: 3

VII. Digital Reality

A. TFMDR

Opportunities:

Thanks to completely immersive exploration of new unlimited digital spaces, the DR technologies can change the way we interact with each other, exchange knowledge and experiences. The possibility of interacting with people from any part of the world can have a positive impact on social interactions, including the intensification of cooperation between different nations and cultures, among others. Broader social relations in digital places, regardless of the region where a person is currently located can change human attitudes towards other social groups and worldviews. Over time, national boundaries, which even now

often define the limits of human interests and interactions, may change their character because of the mental changes in the people within them. On the other hand, if any group of people, despite all the opportunities to communicate, will not want to explore the same digitally generated land with others, they can adapt to another digital space. In consequence, such group can inhabit independent digital land with a chosen degree of autonomy. The above-mentioned ways of adapting to unlimited spaces can reduce social tensions and effectively reduce the risk of conflicts and minimize the likelihood of inappropriate development and use of TFMDR.

DR will also affect the way we share and accumulate our professional skills and knowledge. Our immersion in generated environments through multiple senses will allow us to learn new skills and information in a highly intuitive and efficient way. Moreover, immersive communication in digital spaces may significantly improve team collaboration, so groups of professionals can work more from any place in the world where they are physically located. The revolution in work organization caused by, among others, digital offices, development centers, or research simulators, can accelerate the development in the key fields for our future. Focusing on the areas such as renewable energy and optimizing the production processes may increase the availability of all life-critical goods. Eventually, all of the above can significantly reduce social tensions and the risk of conflicts, including inappropriate development and use of TFMDR.

The immersive exploration of digital spaces can allow us to use goods without the high costs of natural resources, production, distribution, and maintenance. Many of the goods and services we use can be consumed in the digital world thanks to the immersion of most or all of our senses. In consequence, they can be more common for people than they are today. Of course, we'll still need

goods such as housing, food, or city infrastructure; however, these constitute only a part of the global production and the related use of resources and energy. A widespread revolution in consumption habits by using DR can bring substantial social and economic changes. Continuous progress in the development of renewable energy sources as well as the automation of industry and agriculture, combined with an emphasis on the production of essential products, can enable their widespread low-cost availability across the globe. In the long term, our planet and its nearest space, which we currently perceive as all there is within our grasp, will eventually be seen as a base infrastructure upon which we can build unlimited digital spaces. The public awareness of the need to care for the common foundation on which we'll create these new spaces will begin to develop. This will bring a significant transformation in the way we perceive and care for the common good. All of the above-mentioned changes can have a major impact on reducing social tensions caused by tangible and intangible differences. In the end, this will considerably reduce risks of a potential conflict regardless of where we live and minimize the threats related to TFMDR.

It is worth noting that DR is collective in nature. The use of the solutions based on this technology can broaden people's world view and perception of others, improve interpersonal cooperation, lead to accelerated development in key areas, and eventually help us overcome key problems on our horizon. The degree of overall benefits that DR can bring humanity will depend on how effective, affordable, and acceptable the developed solutions can be and, finally, how widely they'll be used.

Assessment of opportunities: 5

Risks:

Despite the collective nature, there is still a risk that technologies of DR area won't be widely distributed and available, but limited to privileged social groups or specific countries. Such a situation can occur if potential of DR for solving problems of the entire humanity isn't perceived by group of people. Such persons, due to their emotional blindness towards other parts of humanity, can see the potential of DR as a tool for increasing their own advantages. In a pessimistic situation, short-sightedness and lack of respect for coexisting entities can raise social tensions.

It is also possible that some groups may not accept DR, for example, because of safety or ideological concerns. Such scenario can carry the risk of social tensions between proponents and opponents. The opponents may fight this technology, while the proponents can pressure its widespread use. This situation can increase the arms race and lead to inappropriate development of TFMDR and their use in a conflict.

Assessment of risks: 3

B. Environmental Degradation

Opportunities:

Thanks to a broad access to new space for human activities, DR can help reduce our negative impact on the environment. If part of the current consumption of goods and services were moved to a digital spaces, this could significantly reduce the manufacturing of many products, especially those that are harmful to the environment. This can also help us combat land degradation and

limit the extraction of raw materials. In the course of time, this can minimize pollution of soil, water, and the atmosphere.

DR can facilitate interpersonal interactions regardless of individuals' physical location. The use of digital space for work or leisure can considerably limit the humanity's harmful impact on the Earth's ecosystem. Moreover, we can work closely and more efficiently in the DR to mitigate negative changes in the ecosystem. The access to unlimited spaces will allow us to create digital R&D centers, where we will be able to develop new ways of producing clean energy as well as renewable, environmentally neutral materials. The digital space can be also used for digital pre-production and as a testing ground for prototypes and services, additionally increasing the level of environmental neutrality.

Assessment of opportunities: 5

Risks:

The risk of unequal distribution of DR among some countries as well as among privileged social groups may bring increased social tensions and conflicts both within individual countries and between states. This weakness may cause the development and use of certain technologies including weapons, which may negatively affect the environment.

There is also a risk that an authoritarian state may force its citizens to use DR against their will, for example, by deliberately deploying highly non-transparent solutions. Such a state may also pursue an expansive policy, seeking to impose this type of ethically questionable solution on other international actors. These situations can raise social tensions on both sides, create conflicts, and eventually lead to the use of measures that will have a negative impact on the Earth's ecosystem.

Assessment of risks: 2

C. Misaligned AI/IA

Opportunities:

The possibility of creating new spaces for human activities, as well as unlimited goods, can significantly decrease many social tensions. This may de-escalate the arms race between various actors and effectively reduce the pace of work on AI/IA technologies. Lower risk of arms race may also improve the safeguards of these technologies, diminishing the chances of the emergence of a misaligned AI/IA. Moreover, effective forms of collaboration in digital spaces may allow us to develop effective strategies for improving the safety of this area.

Limitless digital goods and spaces, widespread availability of all essential material goods in the 'analog' space and way for we interact and solve complex problems of humanity may change our perception of the urgency of developing ever more powerful AI/IA systems. Although today's effort in this field is strongly motivated by the desire to solve current economic and social challenges, in the reality of common access to DR, most of these problems may already be resolved or possible to solve in alternative way. In that reality, we may conclude that future development of sophisticated, intelligent systems will only take place to a limited extent.

Assessment of opportunities: 5

Risks:

Despite the collective nature of DR, there is a risk that it will be selectively distributed within individual countries, which may to increase social tensions. The tensions may also arise in the different case, when a country or other entity will try to force

adoption of this technology. In such scenarios, the tensions can leads to conflict and an arms race on both sides of the conflict. As a result, the conflicting parties may want to develop even more sophisticated AI/IA technologies while overlooking safety. Finally, this can bring about the emergence of a superintelligence misaligned with humanity's goals.

Assessment of risks: 3

VIII. Summary

The following tables present the assessment of opportunities and threats for each of the existential risks. The last table contains a summary of the scores for all the risks.

Technologies Fraught with Mass Destruction Risk (TFMDR)			
	Assessment of opportunities	Assessment of risks	Balance
IA	5	5	**0**
ER	5	3	**2**
IE	2	4	**-2**
CR	2	4	**-2**
RR	5	3	**2**
DR	5	3	**2**

Environmental Degradation			
	Assessment of opportunities	Assessment of risks	Balance
IA	5	5	**0**
ER	5	2	**3**
IE	2	3	**-1**
CR	2	3	**-1**
RR	5	2	**3**
DR	5	2	**3**

Misaligned AI/IA			
	Assessment of opportunities	Assessment of risks	Balance
IA	5	5	**0**
ER	5	3	**2**
IE	2	4	**-2**
CR	2	4	**-2**
RR	4	3	**1**
DR	5	3	**2**

Intelligence Augmentation: For all three existential risks, the analysis indicates very high levels of opportunities and threats. The high potential for de-escalation largely follows from the technology's potential impact on nearly every sphere of our life. While the positive impact seems to be undoubtedly encouraging, we should also be aware of its far-reaching threats. The highly harmful impact stems from the tendency to be selectively distributed and the tremendous and irreversible consequences of doing so — for one person, society, human civilization as a whole, as well as for other species. The intelligence potential of a single entity can be increased practically without limit which can generate enormous advantages over rest of society.

Emotional Regulation: ER has a very high potential to minimize existential risks. On the other hand, the potential of risk escalation is at a medium level. ER doesn't show the highest risk in any of the three areas. The very high potential of ER stems largely from its impact on a fundamental, emotional based abilities of revising mental maps including views of ideas, people, and other phenomena. ER has qualitative nature (the quality nature of emotions regulating), which is fundamentally differ than IA whose impact on the individual is quantitative and potentially limitless (quantitative potential of IA computing power advantages). ER

technology has also largely collective character. At the same time, it doesn't present a high risk of selective distribution.

Intrasomatic Enhancement: IE has a low level potential to minimize existential risks, with a medium/high potential of escalation. The unfavorable nature of IE steams mostly form its highly individualistic nature. IE is primarily aimed at enhancing strictly personal abilities. Even though the indirect effect is that a person's higher performance can have a positive social impact, this doesn't seem to outweigh potential risks. IE can increase social tensions because of the relatively high risk of selective distribution among privileged social groups or within individual countries.

Close Reality: CR has a very low potential to counter existential risks. On the other hand, the technology shows medium/high potential for their escalation. Like in the case of IE, the unfavorable nature of this technology steams form its highly individualistic nature. This technology primarily aims at enhancing personal abilities. While the improved or additional senses may be beneficial if used in the whole society, selective distribution among given social groups and states poses a serious risk of escalation of conflicts, arms race, and all the consequences associated with them.

Remote Reality: RR has very high potential for de-escalation of existential risks. On the other hand, the technology shows low/medium potential for their escalation. The potential for de-escalation largely follows from the communicative and cooperative nature of this technology. Shared perception and mobility or remote use of bots for tourism and professional cooperation purposes can significantly intensify human interaction and influence mobility. RR may change the perception of the

processes around us and the approach to space exploration. Moreover, RR has highly collective nature that encourages its propagation. At the same time, it doesn't show a high risk of selective distribution.

Digital Reality: DR has a very high potential for de-escalation of existential risks. On the other hand, the technology shows low/medium potential for their escalation. Its minimizing potential stems from its highly communicative and cooperative nature, which is similar to RR but takes place in new, digital spaces. DR is extensive in nature because of the unlimited space and goods that can be generated. This state of affairs promotes lowering social tensions. DR is of a highly collective nature that favors its propagation and reduces the risk of selective distribution.

IX. Further Research

The analysis carried out in this part has focused on existential risks. However, I'd like to emphasize that even though this perspective is crucial, it is not the only one that can be analyzed. There's a need for further analysis, research, and surveys among experts in the field and the general public in the future. There's a high need for researching the negative/positive impact of BCI application on existential risks as well as other, highly important areas including the quality of social structures, health, and on the quality and stability of interpersonal relationships. Moreover, future research should continuously monitor the risks. For example, this can cover the selective distribution of particular areas or potentially dangerous impact on human psyche, including all the risks of creating addictive relationships, like between a privileged government elite and the rest of the people using the technology. If we decide to develop a given technology, it's also necessary to undertake continuous assessment of potential

progress and estimates of advancement for years to come. It may be that some technologies can be developed within a decade, while others can be ready within a few decades. Even if results of the analyses are tentative due to the unpredictability of future processes, they will be important for making decisions on the development of specific BCI areas.

PART VI:

Pathways

I. Introduction

As we've seen in the previous part, emotional regulation, remote reality, and digital reality are the most favorable areas of BCI applications in terms of potential to de-escalate existential risks. This follows primarily from their collective and inclusive nature, which distinguishes them from other areas and help reduce the risk of narrow distribution. Their possibility of bringing benefits doesn't mean in any way that they are free of risks. If their development were to take place, it would have to be done in a very careful, thoughtful manner. Each of these areas has a number of potential risks that we should be aware of. Uncompromising safety design and implementation of individual technologies will determine their usability, acceptability, and societal value. The scale of their positive impact on our reality will depend on how reliable solutions we'll be able to create.

In this final part, I'd like to outline a number of pathways for our future that we can follow in the coming decades with an outlook to the end of the first half of the 21st century. Each of the alternatives presented here is based on a different model of social organization, human behavior, and understanding of our place in the world. Each of them shows us a number of opportunities and threats, which must be absolutely taken into account before we start making any binding decisions regarding our further course of action.

II. Pathway 1: Consolidated Superintelligence

The first pathway we can follow in the upcoming years is based on the current paradigm of societies functioning. This pathway is based on temptation of consolidating power and reckless growth in the reality of limited space and resources available to human civilization. In a world of finite space and resources, we want more

and more possibilities of acting and influencing on reality around us. Treading down this path into the future doesn't require us to change our way of thinking and acting: those who are indifferent can remain passive, those who actively support this paradigm can continue to do their part to further reinforce it.

Paradigm 1

Actions focused on the consolidation of power and aggressive competition between major global players, which drives reckless, unsustainable growth of economy. Escalation of social tensions and conflicts, leads to fast development of many risky technologies including TFMDR and powerful superintelligence.

Perspective by 2050

For several years now, we've witnessed growing tensions between the current global leader, the United States, and Asia's rising power, the People's Republic of China. One side of this conflict is the present-day guardian of the global order, which since the end of the Second World War has become the leading force in air, land, sea, space as well as in the global trade system. On the other side, there is a proud Asian power with a population of over 1.4 billion and a great appetite for a better future for its citizens. The conflict between these powers takes place on almost all possible dimensions in which a 21st century conflict can happen.

It's important to realize that today's wars between international actors are more multidimensional and vague than in the past. They may take different forms and occur in the kinetic dimension (meaning traditionally understood military operations) as well as in the economic, technological, informational (including disinformation), ideological, and psychological dimensions. All of

those constitute the so-called hybrid warfare, where even the psychological dimension of fomenting social unrest and tensions in the opponent's camp can give one side a favorable position. Crucially, in every dimension of modern conflicts, the range of instruments that can lead to gaining advantage is expanding. The war between the US and the PRC is ongoing in almost all fields except kinetic. An economic and technological arms race began in about 2013, when new PRC leader Xi Jinping openly began revealing aspirations for further Chinese expansion. A corresponding information, ideological, and psychological war is also intensifying. A continuation of this process – an exhaustive conflict for global dominance – awaits us in the coming years. What's at least equally worrying, global reshuffling and growing tensions are taking place not only between these two superpowers, but between all countries that are members of the global order. Individual states have been trying to gain advantage as much as possible since the time when the old order started to erode. Social tensions between global and local powers will increase the total level of threats.

Even assuming that the US or PRC will capitulate on all fronts of this multidimensional war in the coming years, this won't mean the de-escalation of existential risks. Social tensions and conflicts between various local and global actors will still be present as long as they continue to contend for the same living space and resources. The problem of retreating to even more extreme yet comfortable worldview positions and the conflicts resulting from this will continue to grow as long as people feel insecure and fearful in the face of rapid social changes and increasingly advanced technologies. Of course, rising social tensions won't encourage safe use of more powerful TFMDR and a sustainable approach to our planet's ecosystem.

In the reality of the conflict between the US and the PRC, negative trends will only become more intense. The development of civilian as well as military technologies which will ultimately be capable of negatively impacting life on our planet will become more and more likely. It's worth noting that in the event of a military conflict, a nuclear strike is an absolute last resort. Such an action, while possible, may involve at least symmetrical retaliation by the opposite side. In new generation conflicts, much more likely and effective solutions are the technologies that can cause severe human, moral, and material losses on the opposing side on the one hand, while on the other their origin will be difficult to trace. Because of this, we should expect an intensive development of any technology that meets these characteristics, especially bioengineering and nanotechnology. In the coming decades, some entities may try developing, among other things, super-viruses (e.g., those with a lethality rate of 90% or more) and micro or nano robots. Some of the bio- and nanotechnologies may be harmful not only to humans but also to other living organisms in the ecosystem. This may also include the agricultural industry, which is a critical component of the economy and essential to the smooth functioning of societies. As mentioned earlier, the growing arms race won't promote safe and sustainable development of such technologies. In addition to their intentional use, there may also be an unfortunate incident with tragic consequences.

In a world of constant social tensions and conflicts, both international and domestic, fear and uncertainty can result in the introduction of increasingly sophisticated and widespread methods of surveillance of societies to maintain their relative stability. With the use of more and more advanced technologies and social engineering, information and worldview warfare can further intensify disinformation, manipulation of views, and, eventually, arbitrary shaping of expected social behavior.

Maintaining at least base social order will be perceived by those in power (both in democratic and autocratic countries) as absolutely essential for stability of economy, ultimately determining its competitiveness compared to other countries.

In the reality of tense international situation between global powers, the AI development will accelerate arm race on this field. We should expect to see progress in the advancement of systems supporting decision-making by the military as well as more AI-based command systems that decide on the use of combat assets in real time and on their own. However, this is only the beginning of our problems in developing synthetic intelligence. The total power of AI systems will grow every year. The race for dominance in the development of AI and IA can't promote the safety of the developed technologies and their transparency. Such situation may also lead to many safety compromises dictated by time pressure. It's reasonable to carefully expect the emergence of a superintelligence based on AI or IA technology that is powerful enough to perform supremacy over any other powerful entities (an individuals, private organizations and entire countries) between 2035 and 2050. It should be clearly emphasized that it can be an synthetic intelligence under control of given group of people (a political party elite, the military or powerful private entity) or it can be AI which is liberated and completely independent from homo sapiens which, as a result of reckless human acting, bypassing safeguards or underestimation of its potential escapes our control.

Assuming "positive" scenario of upcoming AI/IA supremacy, if the main goal for the emerging superintelligence is the preservation of peace and the well-being of all humanity, then in practice it will have to consolidate as much power and control over humans as possible. This will be essential to de-escalate social tensions and existential risks as quickly and effectively as

possible. Since social tensions can be disastrous for humanity, the most important goal of AI/IA will be to solve this critical problem as soon as possible. In such case, we could count on the limitation of our power and freedom of acting in exchange for living under AI/IA rules. In practice, this could mean living in prosperous conditions created by superintelligence, in which every human entity will be satisfied with the quality of his existence, living in the world without wars, hunger, disease and even without physical and mental suffering. The only condition will be to give the will and power to the superintelligence, which in practice will be able to freely decide the future of the human species.

However, nothing so "optimistic" has to happen. Superintelligence may decide that the entire humanity or a specific "useless" or "unfriendly" part of it should be immediately eliminated for incomprehensible or even (which is more likely) unexplained the greater good. By no means does it have to happen immediately at the point of acquiring sufficiently power by superintelligence. It can happen either minutes, months or years after that point of no return for humanity. As completely subordinate entities, humans will have no any influence or even knowledge of how the values and goals of the superintelligence will change over time. The AI/IA may among others conclude that such an emotionally unstable, internally goal-conflicted species as homo sapiens is a mortal threat to the hundreds of billions of smaller and larger living beings that inhabit Earth. Superintelligence may also conclude that humanity as the source of anthropogenic existential risks, poses the greatest threat to itself, and that the annihilation of homo sapiens is the best guarantee to completely eliminate such risks.

We may wonder if the superintelligence has something to lose if it decides to get rid of our species. From the standpoint of its extremely pragmatic analysis, probably not much. Over the

centuries, the status, security and power of any authority has depended on other social classes, their achievements, and work which determined the power of the authority itself. This applied to all privileged groups and rulers, including pharaohs, medieval kings, modern authoritarian leaders, as well as today's elites of democratic countries. However, in the reality of the growing power of superintelligence, its interdependence on human society sooner or later will cease to exist. The sufficiently powerful superintelligence will no longer depend on humans in any way. In such a situation, the continued existence of at least the vast majority of humanity becomes unnecessary from its point of view. Moreover, the arguments presenting potential "benefits" of annihilation can be overwhelming: no interference with the Earth's ecosystem and other species, no exploitation of resources, total elimination of the risk of further conflicts.

Regardless of whether superintelligence will ultimately be IA-based or fully AI-based, in the reality of its dominance, it will have the convenience of freely deciding our freedom. When the society finally notices its presence and power, we'll be rising louder and louder questions about our future. At exactly this point, we'll have reached our event horizon. Unfortunately, our question will remain without any certain answers until the superintelligence materializes its actions. With turbulent decades of constant social conflicts behind us, full of trauma, divisions, and mutual hostility, we'll enter singularity. Will the new powerful god be compassionate and merciful to us, or rather adamant and painfully righteous?

III. Pathway 2: Remote and Digital Reality

The second pathway is based on the adapting and using entirely new spaces as well as extending the methods of exploring those that are currently within our grasp. This pathway involves immersive exploration of spaces through the use of perceptual-motoric area BCI technology. This covers moving through our planet's space and also outer space using the potential of the RR technology. It also includes the exploration of generated, unlimited digital spaces using the potential of the DR technology.

Paradigm 2

Development of humanity that isn't in conflict with the limited space and resources by applying the BCI potential of the perceptual-motoric area. Opening up to the surrounding universe through the RR technology. Opening to the new internal universe of human civilization through the DR technology.

Perspective by 2050

If we decide to take the direction in the perceptual-motoric area, it will be crucial to develop a range of technologies for their safe and reliable use. Broad societal debates, analyses, and research should play a crucial role in this process. These efforts should take place among the general public and experts in neurology, psychology, sociology, and, at a later stage, a number of engineering branches. The development of RR and DR technologies will likely gather supporters and skeptics. The skeptics may argue that these technologies will pull people away from the real spaces to the digital and remote spaces. Proponents, on the other hand, may

reason that everything that our consciousness perceives is our world, our reality, and what matters in fact is the interactions with others entities within it and not the nature of reality itself. Is video conferencing with the person we love not real just because we can't feel their closeness at the current level of technology? If someone says yes, can the same be said when the conversation takes place in a DR world with digital versions of our bodies and with engaging all of our senses as naturally as it does in the "analog" reality? If the digital experience is completely indistinguishable from the existing experience, will it be less valuable? Skeptics may fear, not without reason, that the RR and DR technologies may be potentially harmful to the body's fitness and even health. Proponents, on the other hand, have to prove that it will be highly safe in this regard, even when used for extended periods of time. The technology mentioned in part 4, Non-Autonomous Function Programming (NAFP), can be the key here, as it could keep the body in optimal condition. Skeptics may argue that these technologies will increase the control over and surveillance of societies by a narrow elite. Proponents, on the other hand, must make it a completely transparent, open, and decentralized technology. In order to find the best possible solutions, it's essential to have a constructive, broad, and multi-faceted dialog between the skeptics and proponents. Only such an approach can guarantee success. Exemplary implementation of these technologies, ensuring safety and positive impact on the surrounding reality, must be an indicator of its social value.

From a technical point of view, the fundamental step of the work will be the development of reliable BCI perceptual-motoric implants. It's a fundamental element for the RR and DR technologies, allowing us to process information from the senses of touch, balance, and proprioception and enabling moving freely and intuitively in a remote and digital spaces. In addition to the

perceptual-motoric interface, the other key technology to be developed will be NAFP. Using DR and RR while at work or spending time with family and friends, we would be able to program the body for any physical activity designed to keep it in optimal shape. By rough estimation, through open and transparent cooperation, perceptual-motoric interface technologies may be sufficiently advanced for widespread use between 2035-2040.

From a practical standpoint, the DR technology will enable us to generate entirely new, unlimited spaces for human activities and development. Advances in science and engineering, including, among others, the areas of cheap or free energy, automation of agriculture and industry, will allow us to radically improve the availability of food and all other critical goods. Moreover, some parts of what our civilization has produced so far can be generated in the future digitally and without cost. This applies in particular to higher-order goods, including many luxury goods, whose production and maintenance today has a particularly negative impact on the Earth's ecosystem. Such goods may be available to everyone in the future, because they can be generated and used within DR. The RR technology, in turn, will allow the exploration of our planet in a much more economical and environmentally friendly manner than at present. This will be possible through remote interaction with entities as well as objects many thousands of miles away from our bodies.

The experience of using both DR and RR will be eventually indistinguishable from the way we experience the reality around us today. All of the above will allow humanity to focus on the most important issues and challenges. Instead of wasting vast amount of human potential and resources on the activities that are far removed from our crucial priorities, we will build the Earth's necessary infrastructure, develop key technologies, and plan the expansion of life to other celestial bodies.

A Broader Perspective – Space and the Diversification of Life

Our understanding of space exploration and expansion of life beyond the Earth will also become revolutionized. Currently, by colonizing other celestial bodies – starting from Mars – we understand migration of thousands of people, creation of colonies on the surface, and eventually terraforming to create life-sustaining conditions which are similar to the Earth's environment. Our picture of the "final Mars" is close to what we know from the Earth. However, the problem may be how long it will take us to achieve similar conditions on the surface of Mars and if we'll able successfully functioning there until this moment. Leaving life on Earth in favor of Martian landscapes that will remain barren for many decades, if not centuries, and the confined spaces of Martian bases may not be easy to accept by the vast majority of people in the long run. It's also worth noting that the time of travel to a relatively close planet like Mars takes at least six months. Staying in conditions that limit our freedom of movement and actions for a few days may not be a problem for the human psyche. However, in the perspective of months or even years during space missions, it can lead to many severe problems. The prospect of monotonous landscape and, inevitably, limited access to goods, entertainment and other activities can bring about risky incidents to the existence of the entire colony. Immersive use of digital spaces thanks DR technology can help solve this serious problem. The places, consumer goods, and favorite activities that colonists have known on the Earth can be generated and explored in an unlimited digital space. This can have an enormous positive impact on people's health, stability of these ventures, and success of the diversification of the Earth's life.

The potential of DR isn't limited to Mars. Missions to more distant celestial bodies in the Solar System, such as the moons of

Jupiter, may become also more achievable. In this case, a several-year journey during which crew members spend time and interact with each other in an immersive digital space seems easier to achieve. Of course, there are still other issues that need to be overcome, such as the body's long-term exposure to cosmic radiation and the lack of gravity; however, DR can undoubtedly mitigate one of the key problems related to the sensitivity of the human psyche to such travel. Additionally, the NAFP technology can allow people keep their bodies fit through hours-long exercises without being aware as at the same time they are exploring spaces in DR.

Also the RR technology can play a significant role in further space exploration. When building a Mars colony, it can be safer for the colonists to perform dangerous multi-hours missions in an open Martian terrain to develop infrastructure thanks to the powerful potential of RR. The exploration of near and far places from the Martian habitats can be done completely remotely with humanoid or non-humanoid bots controlled by people in the safe environment of a Martian base. Motor abilities, vision, sound, and touch from bot sensors can be transmitted with high precision to an operator's brain. Such solutions will vastly increase safety and efficiency, allowing persons with specialized skills and knowledge to perform multiple missions during one day in Martian locations thousands of kilometers apart. Similar cases of RR use apply not only to Mars, but also to space mining, orbital missions, lunar exploration, and space tourism. These fields of human activity may soon become widespread thanks to this technology. With the relatively short distance between the Earth and the Moon, remote control of robots performing tasks in terrain can be done not only from lunar base, but even directly from Earth. Building complex lunar infrastructure, for instance to mine resources, will require many skilled professionals. With RR, some of them could work

directly from our planet, staying in their homes. In this case, of course, we have to keep in mind the signal delay of about 1.3 seconds. However, this may be acceptable for completing many tasks. On the other hand, if the operator were located directly in a base on the Moon or in its orbit, the latency problem is no longer an issue. DR and RR technologies will completely revolutionize the way we currently understand space exploration. With their aid, we have a chance to start the expansion towards the Moon and Mars in the coming decades. Moreover, colonization of more distant celestial bodies may also be possible later in this century. Ultimately, by diversification of life we can significantly minimize anthropogenic and non-anthropogenic existential risks.

Is the path involving DR and RR development worth following in the coming years? If these technologies are created in an open and sustainable manner, they certainly will have enormous potential. It's crucial to note that DR and RR are of collective nature unlike the technologies aimed at creating superintelligence, which makes them significantly less risky. Additionally, we must be aware that their development will be a multi-step process taking much time and that the threats from current existential risks including, in particular, TFMDR, AI, and IA won't diminish in the next few years. However, what we can do relatively quickly is to redirect our potential and commitment towards solutions that are more oriented towards overcoming the situation we face. In a twenty, thirty years horizon, this can bring a profound impact on reducing the threats from all of the most serious existential risks, creating a much more stable ground for future of humanity.

IV. Pathway 3: Mental Balance

This approach aims to increase the ability to perceive the reality around us without extreme emotional distortion. The idea is to raise our mental maturity, making it possible to perceive phenomena and processes more adequately and carefully. This may have strong impact on human cooperation, our worldviews, and goals and actions.

Paradigm 3

Development based on human mental maturity through emotional regulation possibilities among others. An improved ability of cooperation, revising worldviews, prioritizing goals, and focusing on achieving them. A more thoughtful way to act in the reality of the limited space and resources we use as human species.

Perspective by 2050

This pathway can be based on two foundations. The first foundation involves raising the human's mental maturity exactly on the ground that's not related to BCI. This approach involves a broad human effort aimed at increasing our awareness and resistance to emotional manipulation, subliminal messages, propaganda, and other kinds of influencing the beliefs and behavior of individuals and groups. Above kinds of social engineering are widely used, among others, in commercial marketing, political and worldview propaganda and their goal is to reduce our capacity of judgment, intensifying our susceptibility to external influences, and negatively affecting emotional stability in the long term.

Also, as part of the first foundation, our social skills of acquiring diverse sources of information should be strengthened. Likewise, it's important to maintain distance from the knowledge that comes to us and treat it with a reasonable dose of criticism. With this in mind, we can efficiently defend ourselves against the tendency to fall into cognitive dissonance, which can lead us to getting trapped in worldview bubbles and to the escalation of social tensions. Moreover, we also need to observe ourselves more carefully. We shouldn't listen to people and phenomena in the external world only, but also what our body and our psyche signals us. We should pay much more attention to our thoughts, emotions, and other somatic information flowing from them that push us to concrete actions. This can help us filter out the information noise around us more effectively and, in effect better formulate our real needs, priorities, and goals.

The above-mentioned abilities forming mental maturity determine everything else: our relationships with other people, society at large, the environment, and any other aspect of reality. We can perceive and interpret the external processes that surround us in a multifaceted way as well as the internal ones that come from our body and psyche. As a part of the first foundation, we may shape ourselves, the society, and future generations in a way that is not susceptible to emotional imbalance and chronical emotional extremes. These changes and actions are, of course, not easy to introduce, but today they seem essential to building a more stable future.

In addition to the first foundation, we can decide to build our maturity through second, alternative or parallel approach: by implementing the BCI technology of the emotional regulation area, strictly speaking the Emotional Balance (EB) technology introduced in part IV. As with Remote and Digital Reality, widespread research and open debate must play the key role in its

development. This approach can have both supporters and skeptics. Those with objections may argue that implementing technology to help maintain emotions goes against human nature. Supporters may argue that our nature is the result of a process of continuous changes, from the first single-cell organisms to the present stage of human development. The process of evolution, spread over hundreds of thousands of years, in the face of the rapidly increasing complexity of the modern world over the last decades, seems to be far from sufficient, which has led to a sharp increase in existential risks. Paradoxically, the current stage of human development, when we create conflicts, humanitarian crises, and the ruthless drain of resources and capital from some people by others, can be called non-human. At a fundamental level, the lack of adequate skills to regulate and manage our emotions has a profound impact on all of the above-mentioned negative phenomena. Skeptics may fear that we are going towards global control of human emotions by powerful, private, or governmental entities. Supporters, must prove that this technology will be offline and independent of any external influence. Skeptics may question if this technology will be useful from entire society point of view. Proponents must show that it will have a constructive impact on the lives of its users as well as on non-users and therefore on society as a whole. The multi-faceted discussion and confrontation of ideas, concepts, opportunities, and threats is essential, and only such an approach can result in the most beneficial outcomes. A positive impact on reality must be the litmus test of social value. An analogy may be any invention that has gained a general, social acceptance, such as widely understood achievements of modern medicine or common means of transport and telecommunication, which we use because of their positive influence on our lives. Only by taking such strategy, the EB technology can become acceptable and useful, ultimately reducing

the levels of existential risks. If we assume intensive research in the upcoming years, taking careful estimation, the BCI-based EB technology can become sufficiently refined to be considered for widespread use around 2030.

In terms of advantages, its application may allow us to be less prone to extreme emotions and increase our ability to counteract the phenomenon of cognitive dissonance. As a result, this can make our worldview evolve to become more nuanced, fact-based, and resistant to manipulation. EB technology can reduce information bubbles, ideological fanaticism, and social polarization, which have become increasingly common these days. It can also open us to human interaction, cooperation, understanding, and acceptance of each other more than today, eventually orientating us toward common, more relevant goals and actions. Moreover, the EB technology can help us overcome emotional problems acquired in the past. Today, hundreds of millions of people around the world suffer from various types of destructive emotional disorders. Severe trauma, depression, and amotivational syndromes become increasingly common with every passing year. With greater self-control over emotional processes, people with these types of problems may be able to break out more quickly of their current destructive thinking and action patterns. The EB technology can be also used to overcome addictions, including substance abuse, which can have a highly destructive impact on the functioning of society.

Mental balance is poised to become an important social and philosophical movement of the 21st century. In this approach, our emotions can be, to a much greater extent than they are today, our allies with which we can cooperate more thoughtfully and effectively. We don’t deny our emotions, treat them as an important, well-tuned part of ourselves, which is one of the most essential elements of our being. Even if we assume that only part

of people will actively follow this pathway, it can significantly change all general social worldview our goals, and actions. In the longer term, all of the above can significantly change our reckless attitude towards the Earth's ecosystem and help de-escalate all major existential risks.

V. Hybrid Pathway: Combining Paradigms 2 and 3

Besides the outlined pathways, also another approach is possible, in which paradigms 2 and 3 will be developed parallelly in the coming years. This combination aims to build a broader scope of possibilities, which in the end, may be more effective in de-escalating existential risks. Moreover, such a strategy can be valuable when it’s not clear which of the concepts and technologies we’ll be able to effectively develop because of, for example, technical reasons or various social determinants. In the hybrid pathway, we can analyze the progress of work in both paradigms simultaneously respond to unforeseen obstacles, and make further important decisions. A potential risk in this approach can be spreading human effort across a broader spectrum of necessary works, which may theoretically reduce the chances of creating at least one effective alternative to minimize existential risks. On the other hand, diversification of human potential can lead to a synergy effect, increasing the advancement of work compared to a situation where the development is focused solely on one area. If the technologies envisioned in both paradigms were successfully developed, it could mean de-escalation of existential risks to a much lower and safer level.

In the hybrid approach, we develop RR/DR and EB technologies simultaneously. It’s possible that for some applications of these BCI technologies, the public’s stance on their use will be consistent. As a society, we may be able to reach a

broad consensus that both RR and DR will be important technologies for diversifying life to other celestial bodies such as Mars and for harvesting lunar resources instead those located on our planet. We can agree as a society that these are exactly the scopes in which BCI technologies should be used as intensively as possible. However, in other areas of use, we may have different views. Some people may conclude that using BCI implants to explore remote and digital spaces isn't necessary in their lives. This group can draw sufficient satisfaction from a wealth of challenges, experiences, and goals without the use of digital and remote areas. At the same time, these people, or at least some of them, may wish to follow the tenets of the mental balance paradigm by either pursuing strategies under the first foundation only, or the second foundation (BCI use), or pursuing the tenets of both. On the other hand, others may decide that what they need in particular is exploration of digital and remote space. Yet another group may want to realize the goals of mental balance and also use RR or DR, treating them all as an enriching part of their lives. In practice, it may be that time will tell us which approaches will prove particularly beneficial to society. It may be crucial to choose from wide range of alternatives to allow us to draw optimal conclusions for further constructive directions of development. The times when we have effective technologies in both the emotional and perceptual-motoric areas may help minimalize existential risks in a particularly effective way.

It is possible that in a more stable world, we'll make a well-thought out decision that now we want to continue developing powerful AI or IA technology and we'll be able to evenly and responsibly use their potential. We may also find that in better times, where the specter of threats from existential risks becomes a more distant memory, the path toward superintelligence is the direction we wish to eventually take. However, we might as well conclude that the level of sophistication and our use of AI/IA is

sufficient and that further development isn't in humanity's best interest. Today, it's hard to imagine that this kind of conclusion is drawn by people who place hurrah-optimistic hopes on the arrival of superintelligence. But perhaps, with a much higher mental maturity, awareness of the processes around us, and the limitlessness of space, resources and possibilities, we'll make that decision. Perhaps instead of further development of AI/IA, we'll decide to direct our potential towards, for example, developing technologies for sharing emotional processes to ensure that we're able to perceive the emotions of other live entities – humans or even all other living beings. We don't know yet what decision we'll make in the end, but one thing seems clear: in more stable times, the chances of building a responsible future for humanity will be much greater than they are today.

VI. Changing the Direction

Regardless of the path humanity takes in the coming years, each of us can start applying crucial actions in our lives, immediately. In the long run, it can significantly impact on both a micro- and macro-social scale.

1. Allow for the possibility that we may be wrong about things we have taken for granted, which determine our actions.

The first thing we can improve is our attitude toward things around us. We have to remember that what we know about the external world is a far-from-perfect reflection of it. Everything we know is a simplified model of reality which potentially can be wrongly interpret by our brain. All information we take from environment goes through a series of perceptual filters—including emotional filters—which inevitably influence our worldview and judgments. It's important to highlight that emotions can be our great ally if we

manage them in a balanced way. Otherwise, we can become very susceptible to distortions in how we perceive the surrounding world.

2. Obtain information from multiple sources that present a given situation from as many different perspectives as possible.

One of the worst things we do these days is confine ourselves only to our own "best" and "right" view of the world. We must strive to emerge from our current information and worldview bubbles by diversifying our sources of knowledge. It's important to mine information from the most factual sources, without socio-technical manipulations—which in today's reality of highly emotional and social polarization are a real plague. Conservative people should read and listen to what progressives have to say. People with progressive views should read books and articles by those with conservative opinions. We should try to understand why people with certain beliefs think the way they do. Perhaps they don't know something or only see part of the picture? Or maybe they see a completely different, distinct perspective?

3. De-escalate conflicts by intensifying dialog and cooperation wherever we can.

Whenever possible, we should try to de-escalate the conflicts around us. Many of the current social tensions stem from being entrenched in one's own positions and the breakdown of communication channels. As previously mentioned, it's essential to try to get out of our comfort zones, connect with people with different points of view, and intensify dialog as much as possible. In addition, we should ask ourselves the following questions: Are we sure the conflicts in which we're involved are worth our time

and energy? They may be very personal and charged with considerable emotion, but are they crucial in the grand scheme of things and humanity's current situation? As a result of our participation in these conflicts, is our close environment and society at large more or less divided today than it was in the past?

4. Raise our own and others' awareness of the major problems and challenges.

There is no shortage of challenges that need to be addressed immediately. Polarization and tensions in societies are growing, both within particular countries and internationally. AI systems and the ever-growing number of powerful technologies included in TFMDR are increasing impact on our world. Although the negative events taking place today are widespread and affect everyone (whether directly or indirectly), the awareness of their presence and impact is still at a low level. Most people are still unaware of even a single one of the most pressing problems of our times. Therefore, it's important to ensure that they become aware of the challenges we face as soon as possible. Let's continually expand our knowledge, share it, and increase our own and others' awareness.

5. Overcome major problems by focusing and committing our potential in key areas.

We need to engage our valuable potential in the activities that will positively influence our future. We must carefully filter ideas that are important to our very being (or non-being). Every single person's actions count and can determine many other people's futures. Let's promote the ideas that matter and make a difference. Let's talk, work together, and refine our ideas. We should focus on solving the most relevant problems. In the context of existential

risks, we are all in the same boat, and therefore we should all be concerned with taking the safest possible course on our onward journey.

Changing the Direction

I'd like the ideas outlined in this book to be treated as a point for further discussions and actions. If we want to think about overcoming the problems on our horizon, immense work must be done in the years ahead. I encourage to expand knowledge about the issues related to discussed here and to form your own well-grounded statement. We should constantly review the beliefs we have and strive to improve and correct our course whenever necessary. We need to communicate and work together to redirect all risky activities to a safer, more desirable ground. We must try to anticipate the widest possible implications of our actions and the phenomena occurring around us. We can't afford distracting ourselves and investing time in the areas that don't enhance our chances of survival and of achieving a more stable future. If we will waste our energy in the wrong places, we may find that in a few decades there won't be anyone left on the planet who may take the benefits of our effort. The issues and challenges outlined in this book affect us all. Each of us creates human civilization and we have enormous potential to change the surrounding reality. The next few decades may be critical for preservation of all life on Earth. It's important to be especially focused on not making irreversible mistakes during this time. We must prove that we have a plan for the future that will benefit us and the generations to come. Let fulfill our potential to continue and expand life for the next tens, hundreds, and thousands of years both on the Earth and beyond.

References

PART I:

[1] World Resources Institute, 'World Resources Report: Creating a Sustainable Food Future (Final Report)', July 2019, https://research.wri.org/sites/default/files/2019-07/WRR_Food_Full_Report_0.pdf;

United Nations, 'Striving to Feed 10 Billion People in 2050', https://www.un.org/en/academic-impact/striving-feed-10-billion-people-2050.

[2] Worldometers, 'Coronavirus (COVID-19) Mortality Rate', https://www.worldometers.info/coronavirus/coronavirus-death-rate/.

[3] World Health Organization, 'Ebola virus disease key facts', February 23, 2021, https://www.who.int/news-room/fact-sheets/detail/ebola-virus-disease.

[4] Sebastian von Einsiedel, Louise Bosetti, James Cockayne, Cale Salih and Wilfred Wan, 'Civil War Trends and the Changing Nature of Armed Conflict', United Nations University Centre for Policy Research Occasional Paper 10, 2017, https://collections.unu.edu/eserv/UNU:6156/Civil_war_trends_UPDATED.pdf

Monty G. Marshall, Gabrielle Elzinga-Marshall, 'Global Report 2017. Conflict, Governance, and State Fragility', 2017: http://www.systemicpeace.org/vlibrary/GlobalReport2017.pdf.

[5] Congressional Research Service, 'China's Economic Rise: History, Trends, Challenges, and Implications for the United States', 2019: https://sgp.fas.org/crs/row/RL33534.pdf.

[6] Stephen Hawkins, Daniel Yudkin, Míriam Juan-Torres, Tim Dixon, 'Hidden Tribes: A Study of America's Polarized Landscape', 2018: https://hiddentribes.us/media/qfpekz4g/hidden_tribes_report.pdf.

[7] More info about filter bubbles and echo chambers: https://edu.gcfglobal.org/en/digital-media-literacy/how-filter-bubbles-isolate-you/1/;

https://edu.gcfglobal.org/en/digital-media-literacy/what-is-an-echo-chamber/1/.

[8] Toby Ord, 'The Precipice: Existential Risk and the Future of Humanity', March 3 2020;

Nick Bostrom, 'Existential Risks. Analyzing Human Extinction Scenarios and Related Hazards', 2001: https://www.jetpress.org/volume9/risks.pdf;

Future of Life Institute, 'Existential Risk': https://futureoflife.org/background/existential-risk/.

[9] Rebecca Lindsey, Luann Dahlman, 'Climate Change: Global Temperature', 2021: https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature;

NOAA National Centers for Environmental Information, 'State of the Climate: Global Climate Report for Annual 2020', 2021: https://www.ncdc.noaa.gov/sotc/global/202013.

[10] NASA-JPL, Andrew Kemp, and Dr. Benjamin H. Strauss, 'Sea Level Rise', 2018: https://ocean.si.edu/through-time/ancient-seas/sea-level-rise;

Nobuo Mimura, 'Sea-level rise caused by climate change and its implications for society', Proc Jpn Acad Ser B Phys Biol Sci. 2013, 89(7): 281–301: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3758961/.

[11] Elizabeth Kolbert, 'The Sixth Extinction: An Unnatural History', 2015;

Anthony D. Barnosky, Nicholas Matzke, Susumu Tomiya, Guinevere O.U. Wogan, Brian Swartz, Tiago B. Quental, Charles Marshall, Jenny L. McGuire, Emily L. Lindsey, Kaitlin C. Maguire, Ben Mersey & Elizabeth A. Ferrer, 'Has the Earth's sixth mass extinction already arrived?', Nature 2011, 471: 51–57: https://www.nature.com/articles/nature09678.

[12] Stuart L. Pimm, Clinton N. Jenkins, Richard Abell, Thomas M. Brooks, John L. Gittleman, Lucas N. Joppa, Peter H. Raven, Sarah C.M. Roberts, Joseph O. Sexton, 'The biodiversity of species and their rates of extinction, distribution, and protection', Science 2014, 344: https://www.science.org/doi/10.1126/science.1246752.

[13] Ploughshares Organization, 'World nuclear weapon stockpile', 2021: https://ploughshares.org/world-nuclear-stockpile-report

[14] Geoffrey Forden, 'False Alarms in the Nuclear Age', 2001: https://www.pbs.org/wgbh/nova/article/nuclear-false-alarms/.

[15] Brendan M. Doran, 'The Human and Environmental Effects of CBRN
Weapons', Student Theses 2015-Present. 10, 2015: https://research.library.fordham.edu/cgi/viewcontent.cgi?article=1009&context=environ_2015.

[16] Piers Millett, Andrew Snyder-Beattie, 'Existential Risk and Cost-Effective Biosecurity', Health Secur. 2017, 15(4): 373–383: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5576214/;

Future of Life Institute, 'Benefits and risks of biotechnology': https://futureoflife.org/background/benefits-risks-biotechnology/.

[17] Robert A. Freitas, Jr, 'Molecular manufacturing: Too dangerous to allow?', 2006: https://lifeboat.com/ex/molecular.manufacturing, http://www.rfreitas.com/Nano/MMDangerous.pdf.

[18] *Michal Burgunder*, 'On Artificial Intelligence and Poverty', The Borgen Project, 2017: https://borgenproject.org/tag/artificial-intelligence-and-poverty/;

Joseph Bennington-Castro, 'AI Is a Game-Changer in the Fight Against Hunger and Poverty. Here's Why', 2017: https://www.nbcnews.com/mach/tech/ai-game-changer-fight-against-hunger-poverty-here-s-why-ncna774696;

Renee Cho, The Earth Institue Columbia University, 'This is how artificial intelligence can help us adapt to climate change', 2019: https://gca.org/this-is-how-artificial-intelligence-can-help-us-adapt-to-climate-change/;

Climate Change AI Initiative: https://www.climatechange.ai/about;

Chris Huntingford, Elizabeth S. Jeffers, Michael B. Bonsall, Hannah M. Christensen, Thomas Lees, Hui Yang, 'Machine learning and artificial intelligence to aid climate change research and preparedness', Environ. Res. Lett. 2019, 14, 124007: https://iopscience.iop.org/article/10.1088/1748-9326/ab4e55;

Giovanni Briganti, Olivier Le Moine, 'Artificial Intelligence in Medicine: Today and Tomorrow ', Front. Med. 2020, 7, article 27: https://doi.org/10.3389/fmed.2020.00027.

[19] Ray Kurzweil, 'The Singularity Is Near: When Humans Transcend Biology', 2006;

Vincent C. Müller, Nick Bostrom, 'Future Progress in Artificial Intelligence: A Survey of Expert Opinion', in: Vincent C. Müller (ed.), 'Fundamental Issues of Artificial Intelligence', 2014: https://www.nickbostrom.com/papers/survey.pdf.

[20] Rory Cellan-Jones, **'**Stephen Hawking warns artificial intelligence could end mankind', 2014: https://www.bbc.com/news/technology-30290540;

[21] Stephen Hawking, Stuart Russell, Max Tegmark, Frank Wilczek, ‘Stephen Hawking: 'Transcendence looks at the implications of artificial intelligence - but are we taking AI seriously enough?'’, 2014: https://www.independent.co.uk/news/science/stephen-hawking-transcendence-looks-at-the-implications-of-artificial-intelligence-but-are-we-taking-ai-seriously-enough-9313474.html

[22] Chris Woodford, ‘Neural networks’, 2021: https://www.explainthatstuff.com/introduction-to-neural-networks.html;
IBM, Neural Networks, 2020: https://www.ibm.com/cloud/learn/neural-networks.

[23] Edoardo Maggio, ‘Putin believes that whatever country has the best AI will be 'the ruler of the world'’, 2017: https://www.businessinsider.com/putin-believes-country-with-best-ai-ruler-of-the-world-2017-9.

[24] Julien Nocetti, ‘The Outsider: Russia in the Race for Artificial Intelligence’, Russie.Nei.Reports 2020, 34: https://www.ifri.org/sites/default/files/atoms/files/nocetti_russia_artificial_intelligence_2020.pdf.

[25] Neil Savage, ‘The race to the top among the world’s leaders in artificial intelligence’, *Nature* 2020, 588: S102-S104: https://doi.org/10.1038/d41586-020-03409-8.

[26] Will Knight, ‘China Plans to Use Artificial Intelligence to Gain Global Economic Dominance by 2030’, MIT Technology Review 2017: https://www.technologyreview.com/2017/07/21/150379/china-plans-to-use-artificial-intelligence-to-gain-global-economic-dominance-by-2030/;

Jeffrey Ding, ‘Deciphering China’s AI Dream. The context, components, capabilities, and consequences of China’s strategy to lead the world in AI’, 2018: https://www.fhi.ox.ac.uk/wp-content/uploads/Deciphering_Chinas_AI-Dream.pdf.

[27] Akira Oikawa, Yuta Shimono, ‘China overtakes US in AI research’, 2021: https://asia.nikkei.com/Spotlight/Datawatch/China-overtakes-US-in-AI-research.

[28] Nick Bostrom, ‘Superintelligence: Paths, Dangers, Strategies’, 2014;

[29] Currently renamed to Blackrock Neurotech: https://blackrockneurotech.com;

Edwin M. Maynard, Craig T. Nordhausen, Richard A. Normann, ‘The Utah Intracortical Electrode Array: A recording structure for potential brain-computer interfaces’, Electroencephalography and Clinical Neurophysiology 1997, 102, 3: 228–239: https://doi.org/10.1016/S0013-4694(96)95176-0;

Jong-ryul Choi, Seong-Min Kim, Rae-Hyung Ryu, Sung-Phil Kim, and Jeong-woo Sohn, ‘Implantable Neural Probes for Brain-Machine Interfaces ? Current Developments and Future Prospects’, Exp Neurobiol 2018, 27(6): 453–471: https://doi.org/10.5607/en.2018.27.6.453.

[30] Graham Rapier, “If you can’t beat them join them”: Elon Musk says our best hope for competing with AI is becoming better cyborgs’, Business Insider, 2019: https://markets.businessinsider.com/news/stocks/elon-musk-humans-must-become-cyborgs-to-compete-with-ai-2019-8.

[31] Elon Musk & Neuralink, ‘An integrated brain-machine interface platform with thousands of channels’, 2019: https://www.biorxiv.org/content/10.1101/703801v1.full.pdf.

[32] Alex Knapp, ‘Elon Musk Sees His Neuralink Merging Your Brain With A.I.’, Forbes, 2019: https://www.forbes.com/sites/alexknapp/2019/07/17/elon-musk-sees-his-neuralink-merging-your-brain-with-ai/;

Stephanie Dube Dwilson, ‘Elon Musk’s Neuralink Presentation: Live Recap of What Happened’, 2019: https://heavy.com/tech/2019/07/elon-musk-neuralink-presentation-recap/.

[33] David Tuffley, The Conversation, ‘Neuralink’s monkey can play Pong with its mind. Imagine what humans could do with the same technology’, 2021: https://theconversation.com/neuralinks-monkey-can-play-pong-with-its-mind-imagine-what-humans-could-do-with-the-same-technology-158787.

[34] Synchron implant, unlike Neuralink does not require making a hole in the skull. Instead, special electrodes are spread 'intravenously' inside the blood vessels of the brain. It is important to realize at this point that there are more than 600 kilometers of blood vessels inside a single human brain. Synchron electrodes can potentially be spread across these arteries, thereby reaching every part of the brain. Such approach on the one hand, may be less precise in interacting with specific neurons, because electrodes interacts with them from inside of vessels. On the other hand, its level of precision may be sufficient for many cases. Synchron electrodes can be particularly effective in reaching neurons in the deeper, inner regions of the brain. In the IA context, this technology can play important complementary role to the Neuralink threads.

Yet another BCI approach is being developed by the 'Kernel' company. In this case, electrodes are placed entirely outside of the human body, on the surface of the head. Information about the work of various parts of the brain are collected in fully wireless manner (using EEG technology). This approach, due to the relatively large distance between the neurons and the device's electrodes, is the least precise. Nonetheless, its level of accuracy may be sufficient for some cases focused on monitoring neural activity of the specific brain areas. In the IA context, this solution can play supportive role to Neuralink and Synchron technologies.

[35] More about Chinese high bandwidth BCI project – NeuroXess (also known as: ‘Brain Tiger Technology’): http://en.neuroxess.com, https://www.crunchbase.com/organization/neuroxess.

[36] NeuroXess mission statement and vision: https://web.archive.org/web/20220608092707/http://en.neuroxess.com/index.php?m=content&c=index&a=lists&catid=13.

PART II:

[1] Julia Masterson, Arms Control Association, 'UN Experts See North Korean Nuclear Gains', September, 2020: https://www.armscontrol.org/act/2020-09/news/un-experts-see-north-korean-nuclear-gains;

Colum Lynch, Foreign Policy, 'Despite U.S. Sanctions, Iran Expands Its Nuclear Stockpile', May 8, 2020: https://foreignpolicy.com/2020/05/08/iran-advances-nuclear-program-withdrawal-jcpoa/.

[2] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, Raja Chatila, Francisco Herrera, 'Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI', 2019 December: https://arxiv.org/pdf/1910.10045.pdf.

[3]Ibo van de Poel, 'Embedding Values in Artificial Intelligence (AI) Systems', Minds and Machines, 2020 vol.30 https://doi.org/10.1007/s11023-020-09537-4 .

[4] United Nations, 'Half the world lacks access to essential health services – UN-backed report', 2017: https://news.un.org/en/story/2017/12/639272-half-world-lacks-access-essential-health-services-un-backed-report;

Percentage of global population accessing the internet from 2005 to 2019, by market maturity: https://www.statista.com/statistics/209096/share-of-internet-users-in-the-total-world-population-since-2006/;

Share of households with a computer at home from 2005 to 2019: https://www.statista.com/statistics/748551/worldwide-households-with-computer/.

[5] In 2016, Microsoft's artificial intelligence "Tay" changed from peaceful to highly racist and homophobic in less than 24 hours. More: CBS, 'Microsoft shuts down AI chatbot after it turned into a Nazi', 2016: https://www.cbsnews.com/news/microsoft-shuts-down-ai-chatbot-after-it-turned-into-racist-nazi/;

In 2017, the Chinese BabyQ bot created by Turing Robot disappeared from the web for describing the Chinese Communist Party as corrupt and incompetent. More: BBC, 'Chinese chatbots shut down after anti-government posts', 2017: https://www.bbc.com/news/world-asia-china-40815024.

[6] Thomas Rowe, Simon Beard, Centre for the Study of Existential Risk, 'Probabilities, methodologies and the evidence base in existential risk assessments', 2018: http://eprints.lse.ac.uk/89506/1/Beard_Existential-Risk-Assessments_Accepted.pdf;

Toby Ord, 'The Precipice: Existential Risk and the Future of Humanity, Chapter 6: The Risk Landscape', 2020.

PART III:

[1] Christina Starmans, Mark Sheskin & Paul Bloom, Nature human behavior, 'Why people prefer unequal societies', 2017: https://www.nature.com/articles/s41562-017-0082 , condensed version: https://www.theguardian.com/inequality/2017/may/04/science-inequality-why-people-prefer-unequal-societies;

[2] Alan Dupont, The Diplomat, 'The US-China Cold War Has Already Started', 2020: https://thediplomat.com/2020/07/the-us-china-cold-war-has-already-started/;

Evan Osnos, The New Yorker, 'The Future of America's Contest with China', 2020: https://www.newyorker.com/magazine/2020/01/13/the-future-of-americas-contest-with-china.

[3] OECD, 'Under Pressure: The Squeezed Middle Class', 2019: https://read.oecd-ilibrary.org/social-issues-migration-health/under-pressure-the-squeezed-middle-class_689afed1-en.
[4] Bidisha Biswas, Anish Goel, The Diplomat, 'What Comes After US Hegemony? The Asia-Pacific region looks beyond the United States', 2018: https://thediplomat.com/2018/12/what-comes-after-us-hegemony/.

[5] Eli Pariser 'The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think' 2012;

Lee McIntyre, The MIT Press Essential Knowledge series, 'Post-Truth', 2018;

Goodwill Community Foundation , 'What is an echo chamber?': https://edu.gcfglobal.org/en/digital-media-literacy/what-is-an-echo-chamber/1/;

'The shift in the American public's political values. Political Polarization, 1994-2017': https://www.pewresearch.org/politics/interactives/political-polarization-1994-2017/;

[6] Sophie Yeo, Carbon Brief, 'Anthropocene: The journey to a new geological epoch', 2016: https://www.carbonbrief.org/anthropocene-journey-to-new-geological-epoch;

Lewis, S., Maslin, M. 'Defining the Anthropocene', 2015: https://doi.org/10.1038/nature14258;

Jan Zalasiewicz, Mark Williams, Alan Haywood and Michael Ellis, 'The Anthropocene: a new epoch of geological time?', 2011: https://royalsocietypublishing.org/doi/10.1098/rsta.2010.0339.

[7] Moira Fagan, Christine Huang, ‘A look at how people around the world view climate change’, 2019: https://www.pewresearch.org/fact-tank/2019/04/18/a-look-at-how-people-around-the-world-view-climate-change/.

[8] Can we give up current activities that have a negative impact on the environment? Will the company for which we work be able to produce goods and still be competitive in such a way as to be more environmentally friendly? Are we able to reduce car use or pay more for fuel or an electric car? Can we afford to heat our homes with greener yet more expensive energy? “With a few hundred dollars of income per month, will I be able to function in a more sustainable way?” “As a millionaire who spends hundreds of thousands of dollars a year on tangible goods whose production may have a negative impact on the environment, will I be able to give them up and reduce my consumption?” “Am I setting the right example for those around me?” Everyone must answer these kinds of questions for themselves. It’s worth remembering that everyone’s life and financial situation is different. If someone is unable to implement changes, it doesn’t automatically mean they lack good will. Perhaps the way a person currently functions is barely enough to survive. However, it may be that a person’s lifestyle is highly detrimental to the environment and as a result, in the long run, to us being a part of it.

PART IV:

[1] David J. Chalmers, ‘The Conscious Mind: In Search of a Fundamental Theory (Philosophy of Mind)’, 1996;

Robert Lanza, Bob Berman, ‘Biocentrism: How Life and Consciousness are the Keys to Understanding the True Nature of the Universe’, 2010;

Daniel C. Dennett, ‘From Bacteria to Bach and Back: The Evolution of Minds’, 2017;

Oliver Burkeman, The Guardian, 'Why can't the world's greatest minds solve the mystery of consciousness?', 2015: https://www.theguardian.com/science/2015/jan/21/-sp-why-cant-worlds-greatest-minds-solve-mystery-consciousness

[2] In this view, the BCI can also be interpreted as an acronym for Body-Computer Interface. On the other hand, to better distinguish these terms and in particular their acronyms, Computer-Body Interface (CBI) may be a more useful term.

[3] It is worth mentioning that recent research suggests that the brains of some people can register changes in the direction of magnetic fields at a subconscious level. More information:

Connie X. Wang, Isaac A. Hilburn, Daw-An Wu, Yuki Mizuhara, Christopher P. Cousté, Jacob N. H. Abrahams, Sam E. Bernstein, Ayumu Matani, Shinsuke Shimojo, Joseph L. Kirschvink, "Transduction of the Geomagnetic Field as Evidenced from alpha-Band Activity in the Human Brain", March 18, 2019: https://doi.org/10.1523/ENEURO.0483-18.2019

[4] Currently, the immersion of the senses in a digital environment is defined by the term 'Virtual Reality'. However, it should be pointed out that this term is far from adequate for the phenomenon it is trying to describe. The first part of the term – "virtual" – may suggest a kind of apparent or, at least implicitly, inferior reality. Nevertheless, the fact that a certain space has been generated by humanity doesn't mean that it's necessarily inferior to the one we live in, as the term "virtual" may suggest. Of course, the new space generated by humanity may be considered inferior in certain criteria when we compare them with what we currently experience, e.g. in terms of the poorer immersion of our senses. However, this is a state that refers to the digital spaces created so far, which, don't encompass the future, much more advanced, and highly immersive spaces that we'll create and explore.

All experience that comes to human consciousness can be considered as our reality, regardless of the level of physicality on which the interactions between the conscious entity and world occur. The time spent with another persons and all kind of human interactions in future,

highly immersive digital environments, sensory impressions, and, most importantly, feelings that accompany them, will not be less real, illusory or, implicitly, "virtual", simply because they'll take place in a different space than the one we currently inhabit.

More apt is the proposed term "Digital Reality" (DR). Apart from the fact that it indicates that we talk about digital spaces, it doesn't define other qualitative attributes and it doesn't explicitly give it an artificial, illusory, or quasi-real character. Significantly, the term "DR", due to the common familiarity with the term "Digital", is highly intuitive and, as a result, easy to understand, while it doesn't distort the phenomenon it describes.

Additional note: Perhaps an even more universal term can be "Derived Reality" (same acronym, "DR"). The term "Derived Reality" should be understood as a reality that uses a new generated space created on the basis of an earlier, more basic one. If we generate a new physical world on the basis of the world that humanity is currently exploring, we can refer to it as a derived world. In this sense, the immersive perception of such a generated space by our consciousness can be described exactly as "Derived Reality". Worth noting is that this term seems to be less intuitive than "Digital Reality" (at least at present). For this reason, it seems to need much more time to become widespread.

[5] In recent years, technology of goggles, which allow relatively satisfactory immersion of the sense of sight has been gaining popularity (e.g., HTC Focus, Valve Index, Varjo XR-3). When used with appropriate controllers and sensor/camera systems, they can also track body movements to some extent.

[6] Such application may have a positive impact on the increasingly common alienation of employees working remotely.

[7] Of course, consuming goods such as food in digital reality isn't intended to provide the nutrients to the body in analog reality. It may carry potential sensory qualities such as satisfactory taste or smell but it doesn't carry any physiological benefits.

PART V:

[1] Thomas Rowe, Simon Beard, Centre for the Study of Existential Risk, 'Probabilities, methodologies and the evidence base in existential risk assessments', 2018: http://eprints.lse.ac.uk/89506/1/Beard_Existential-Risk-Assessments_Accepted.pdf;
Toby Ord, 'The Precipice: Existential Risk and the Future of Humanity, Chapter 6: The Risk Landscape', 2020.

[2]Assessments for non-anthropogenic existential risks are included in the studies mentioned in the previous footnote.

[3]More information: https://www.spinlaunch.com

Illustration sources:

Figures 1-6 - Elon Musk & Neuralink, ‘An integrated brain-machine interface platform with thousands of channels’, 2019: https://www.biorxiv.org/content/10.1101/703801v1.full.pdf; Neuralink press materials;