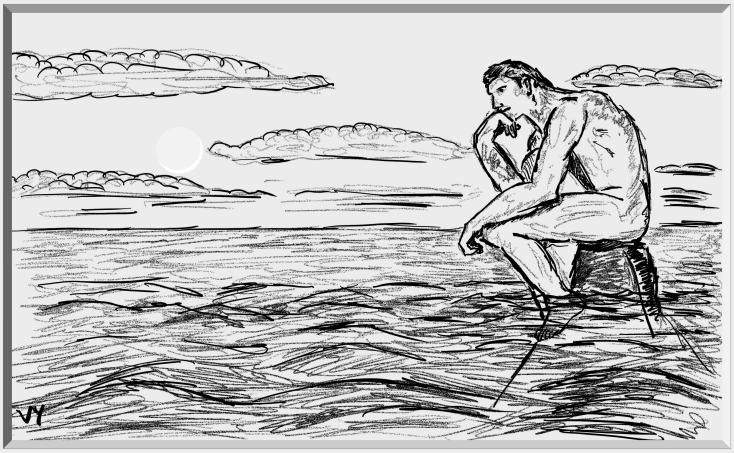
Human-led Science in the Age of Superintelligence

Vladislav A. Yastrebov

September 8, 2025



A Thinker on the Summit of Mount Everest

Zero AI usage

100% human created content.

This essay was written without usage of any AI/ML tools. Elaboration of ideas and concepts, drawing of figures, writing, editing, translation, spelling and grammar check – all was done by Vladislav A. Yastrebov without AI usage. It was a deliberate choice of the author. This choice, however, could affect to some extent the quality of the text as the author is not a native English speaker. On the other hand, the readers can be sure that the wording of this essay keeps genuinely author's voice and that the ideas and their explications were not adjusted or deformed by an AI.

The figure on the cover refers to "The Great Flood" analogy of a flooded landscape of human competence (Moravec, 1988; Moravec, 1998) – when some cognitive tasks are performed better by machines than humans, the "mountain"-analogy of this domain is shown as flooded – arts and sciences were shown as the highest peaks in this analogy. The waxing crescent moon signifies a sunset of the human intellectual domination[1].

[1]This interpretation of the crescent moon makes sense in the Northern hemisphere, and so it is true at the altitude of Mount Everest.

Contents

1 Introduction **1**
- 1.1 Notions . . . 1
- 1.2 Main assumptions . . . 2

2 Superintelligence-led Science (SIS) **3**
- 2.1 Separation of HS and SIS . . . 3
- 2.2 SI Scale and its Energy Efficiency . . . 5
- 2.3 The Pace of SIS Progress . . . 6
- 2.4 Cleaning up Sciences . . . 7

3 Scientific Landscape **9**
- 3.1 Fractal boundary between known and unknown . . . 9
- 3.2 Hypothesis of non-overlapping scientific questions . . . 10
- 3.3 Scientific Map . . . 10

4 Human scientists **11**
- 4.1 Human capacity . . . 11
- 4.2 Social aspect and its emulation . . . 15
- 4.3 SI-HS relations . . . 16
- 4.4 Meaning and purpose . . . 18

5 Conclusion **19**

1 Introduction

Do you mind if an AI-tool generates for you a book or a poem, a movie or a theater play, a painting or a photography, a song or a concerto, a sculpture or a cathedral, a choreography of a dance? If an AI performed, created, generated, constructed it, and not humans with their own history, mastering and their "soul"? I believe that most of us prefer human-created arts. But what about scientific progress? Would it matter for you if an AI proves a theorem, constructs experimental equipments and conducts experiments, collects and analyzes data, draws conclusions, develops new models and refines theories? Whether you like it or not, in the age of SuperIntelligence (SI), SI will take care of all these aspects of scientific progress (Moravec, 1998; Moravec, 1999). It will construct new neutrino and gravitational-wave detectors, new particle colliders, it will send new missions to the Sun, to the planets and small bodies of our Solar System and to the outer space in our galaxy and beyond. SI will conduct large scale social and political experiments, study all historical archives, carry out archaeological excavation, and in parallel, of course, it will advance all theoretical sciences and will maybe create new ones. What about us? What about human scientists in this endeavor?

In this essay, I will consider different scenarios of scientific exploration and possible roles of humans (or our augmented versions) in this future. Will we simply be archivists, or "priests" and "priestess", in a temple of SuperIntelligence-led Science (SIS)? Or maybe, we will have our contribution to make to this new Science and even push our own, Human-led Science (HS)? Will this HS be based on the legacy science, or can it grow near the forefront of SIS? Will we be active players of the true scientific research or simply science-game players in our virtual personal universe? Are our current, evolution-provided cognitive capacities enough to push science further? Or we need to be augmented: mentally or biologically or totally cyborgized to try to keep pace with SIS? Will we push the science being always accompanied by an SI mentor, or will we be able to do something on our own in our own sandbox where an SI will be forbidden? Will the SI share with us the totality of its knowledge and discoveries? Will the SI eventually stop in its exploration being satisfied by the world model that it has constructed, or this scientific progress will continue forever? These questions will be addressed in this essay.

1.1 Notions

Several abbreviations will be used throughout the essay, they are introduced in the main text, but for the convenience of the readers, a list of main notions is also provided below.

- **SI (SuperIntelligence):** an intelligence (information processing entity, agentic or not) largely surpassing the intelligence of the top human experts over all intelectual domains.
- **SIS (SuperIntelligence-led Science):** knowledge acquired and science advanced by the SI (with or without humans in the loop), notably for the purpose of constructing an accurate world model.
- **HS (Human-led Science):** either a legacy human-led science and a new science, beyond the legacy one, which was understood and internalized by at least a few human scientists. Even though the verb "lead" is used here, we should bear in mind, that this progress is often happens over paths paved by SI.

1.2 Main assumptions

Beyond the Great Filter. It is an utopian essay in the spirit of Bostrom (2024)[2] and Amodei (2024). We must not, however, forget that getting to this utopian state with a *benevolent* SI presents a great challenge *per se* – the most important and the last challenge that the humanity will face (Bostrom, 2014). Nonetheless, I will not focus on these risks, since the objective of this essay is to reflect on the state of science when this "Great Filter" has been overcome and the humanity has a well-aligned SI at its disposal (Yudkowsky, 2008; Yampolskiy, 2016; Yudkowsky, 2022). Let's assume a good scenario and dream.

Science and the SI's world model. Why SI will take the lead in the scientific research? As known from (Omohundro, 2008; Bostrom, 2014), any goal provided to an AI – being a super optimizer – induces sub-goals such as self-preservation, access to resources and construction of an *accurate world model*. This latter implies that before becoming an SI, a proto-SI will push the frontiers of our science to start constructing a better world model in its "head".

Expanding frontiers at the SI scale. Scientific breakthroughs in the human-led science (HS) often happen thanks to a few individuals capable of independent thinking and possessing an exceptional mastering and understanding of their domains. However, the progress in science does not reduce only to breakthroughs: there is also a different type of hard work ensured by the scientific community which contains verification and validation, deeper understanding, interpretation, enhancement, transmission and eventual application of these breakthroughs. SuperIntelligence (SI), by definition is an intelligence largely surpassing the capacities and competences of the best human scientists capable of breakthroughs, and of course it can ensure more tedious task related to enhancement, verification and application. Therefore, an SI representing millions of enhanced genii[3] accompanied by trillions of more ordinary research workers will be capable to rapidly push the frontiers of the current "legacy" human-led science (HS) towards new frontiers. The knowledge acquired in this advancement will be denoted SuperIntelligence-led Science (SIS), which of course will incorporate the legacy HS too.

SI-idle questions. Nevertheless, it does not mean that answering all questions that are interesting for us, are beneficial for constructing this world model. Therefore, scientific progress of the SI can ignore some unresolved human's questions. Nevertheless, the SI could be managed to make progress in these directions too if

[2] Probably, for people not familiar with the concept of post-instrumental deep utopia from Bostrom, this text might seem to go too far beyond an ordinary (linear) vision of the future and will defeat their imagination. To make such readers more familiar with the utopian vision of the future in the age of SI, I would invite them to take a look on the world drawn by Bostrom (2024). For a detailed short-term AI evolution forecast, the reader can consult (Kokotajlo et al., 2025) and a preparation guide (MacAskill and Moorhouse, 2025).

[3] "A country of geniuses in a datacenter" as stated by Dario Amodei, a CEO of Anthropic, one of the leading AI labs in the world.

needed. These scientific domains and questions, which leave the SI "indifferent", will be denoted as SI-idle scientific questions/domains (see Section 3.2).

Oracle or sovereign? An *Oracle* type SI, which only aswers questions and does not posses any agentic capacities, is probably a form of SI the most adapted for a meaningful human-led exploration of scientific frontiers. Scientists can ask intelligent question, reflect on answers, accumulate and synthesize new knowledge. But it is obvious already, that AI[4] development has not taken the oracle-path. Therefore, it is more probable that agentic "genie" or even super powerful "sovereign" SI types, introduced by Bostrom (2014), will prevail in the future. But both scenarios will be considered.

Science as a meaningful occupation in a solved world. With an SI, it is obvious that the mankind is practically excluded from the advancement of the forefront science at least in important domains – those which are critical for constructing an accurate world model. Will such occupation as "scientist" disappear completely? Probably not, because in a "solved world", an SI-assisted human's search for meaning and purpose will be of crucial importance. A science-related occupation seems a very appealing choice for those who value intelligence.

Augmentation of humans. In a solved world, human scientists will be free to choose: to be augmented and explore the forefronts of the science alongside with the SI or slightly behind, or to remain genuine humans with maybe only light adjustments, and make a slow-pace progress towards the forefront and never reach it. But whatever is someone's choice, they will be able to tune themselves to be absolutely fine with this choice. Upgrades and degrades will be also possible with or without memory preservation.

Outline of the essay. The essay is organized as follows. In Section 2 we explore particularities of SI-led Science (SIS). In Section 3 a map of scientific knowledge and its frontier separating the Human-led Science (HS) and the SIS is outlined. Section 4 explores possible roles of science-related humans in the age of SI. Conclusions are drawn in Section 5.

2 Superintelligence-led Science (SIS)

2.1 Separation of HS and SIS

Separation. It is essential to separate the SI-led Science (SIS) and the Human-led Science (HS). In the current HS, of course, nobody has a fine-grained vision of the fractal boundary between the known and unknown, but some scientists have a relatively broad vision of the fuzzy macroscale frontier at least in their own and neighboring domains (see Section 3). Some scientists do work on the forefront of the global HS frontier, and some are simply pushing their own boundaries which are located in the vicinity of the already known. The HS' frontier can be cristallized but of course it exists nowhere but in scientists' minds and in the global and preferably

[4]Artificial Intelligence in a general sense.

cleaned up (see Section 2.4) scientific literature. Individual scientific boundary and the global HS' one can differ drastically even at smaller scales of particular questions. Now, we need to imagine that the SIS' boundary spanning all domains, at least those relevant for an accurate world-model construction, will further differ from the HS' boundary. Can the humanity assume that this new frontier belongs to them too? In some sense, yes, because with a benevolent SI they will gather the fruits of this new science and technology. But not in the intellectual sense, because there's none of living or had lived human individuals who understood (in a broad sense of this verb) this new boundary nor the gap that separates if from the HS frontier. Therefore, one of possible human occupations will be pushing the HS frontiers closer to the SIS one. Advancement to the forefront (which is potentially being pushed by SI at an ultra-high speed) will require from humans to go through difficult paths and acquire new knowledge and new tools to understand and appreciate the discoveries.

Asymmetry in knowledge sharing. Sharing of information between HS and SIS is *not symmetric*. If eventually humans make progress beyond the SIS' frontier, the later can readily integrate this progress because it overviews the totality of information circulating on the web. The opposite is not possible, the SI can and will advance the science for its world model construction and will not necessarily share its progress with humans. To transmit information, there should be a receiver on the human's side. So, such a transmission could be done in special cases, like in Oracle Temple analogy (see Section 4.3), but receiving information about SI progress in real-time is unlikely because of the limited intellectual and bandwidth capacity of the receivers, even if they are augmented. So, for some discoveries, maybe a life-long training will be needed to understand at least approximately the contours of the discovery. In addition, there should *exist* some humans interested in receiving this information; probably, some domains will seem to be of no interest to humans, and thus no information will be spontaneously shared with us.

Moreover, within some scientific domains with a high risk of "black balls" – potentially destructive technologies[5], SI could decide not to share new knowledge even if asked; let's call such domains "human forbidden scientific domains"[6]. Probably, even the frontiers of such domains can be protected, and the SI will not let any human in this buffer zone. In overall, SI strategy about black balls' scientific domains presents a difficult philosophical and ethical dilemma, and our access into these domains will depend on the alignment settings.

The benevolent SI should be very careful and protective with these devastating black-ball technologies. One potential scenario of blocking the access for humans is a non-intrusive external distraction of scientists and engineers trying to penetrate into the black-ball zone and construct a technology. This distraction could take a form similar to the one imagined in a science fiction novel "Definitely Maybe"[7] from brothers Strugatsky. There, an astrophysicist who was about to make a revolutionary discovery finds himself continuously distracted from his scientific work by such an improbable sequence of events that the scientist has deduced that something intelligent (Homeostatic intelligent Universe) prevents him from continuing his

[5]See the "vulnerable world hypothesis" by Bostrom (2019).

[6]I do not exclude, the existing level of science is already sufficient to invent a black-ball technology, and the benevolent SI will not let us pull such a ball.

[7]The original title of this novel – "A Billion Years Before the End of the World" – is much more meaningful in the context of SI and black-ball risks.

study[8]. Ultimately, he abandons this topic. With SI, this distraction can be much softer or even invisible. In the worst case scenario, if external non-intrusive distraction does not work, it can intervene intrusively by rewiring the brain of the intruder. Alternatively, the SI can let us into these forbidden zones but will prevent us from constructing a black-ball technology to protect humanity and the Universe. To do so, the SI with its strategic planning must foresee very remote consequences of major scientific discoveries, but even with its intellectual power, forecasting the behavior of a chaotic dynamic system for billions of years will be far beyond its capacity.

Apart from protecting us from "black ball" innovations, we could try to instill in the proto-SI the seeds of ethical innovation and research (Grinbaum and Groves, 2013; Grinbaum, 2024) in addition to general moral incentives which will be further deepen and developed by SI itself.

2.2 SI Scale and its Energy Efficiency

Necessarily, SI will have a perfect representation of a human brain functionality. Our brain is slow and not optimized by evolution for purely cognitive function[9], but it is quite energy efficient compared to the current comparable AI's energy consumption, it also learns and generalizes very rapidly, and if we eventually get to the age of a *benevolent* SI, *per se* is the ultimate proof of our brain efficiency. At the same time, I believe that human brain's cognitive function can be made even more energy efficient by removing all extra functionality unnecessary for the cognitive function. Therefore, SI will be able to reproduce a human-level intelligence with let's say 1% of its energy consumption, and, if implemented on a silicon or another non-biological support, ≈10,000 times faster[10]. Therefore, only with this optimization, we can hope for an efficiency gain by a factor of 10^6 compared to our brains. If you remove emotions and distractions, if you add access to reliable and large databases and the Internet, such an information processing entity will be already a proto-SI. If you scale the number of neurons and connections, then you can get to a total factor of one billion in terms of efficiency of cognitive function[11], which you can readily call *SI* and which can further self-develop and replicate. Maybe, there could be a trade-off between the depth / speed and energy consumption for the objective function and any further expansion in scale would penalize the cognitive function, but anyway this factor is already far beyond our perception of SI. With a factor of a billion beyond the human intellectual capacity for a single entity and with a perfect coordination between millions of them, such an SI will probably be already optimal for operating in our Universe. In the end, it will depend on its goals. If needed, these factors can be further increased.

[8] As it nicely stated in Wikipedia "... the mysterious force is the Universe's reaction to mankind's scientific pursuit, which threatens to destroy the very fabric of the Universe in some distant future."

[9] I cannot help but cite here Alexander (2014): "evolution is a blind idiot alien god that optimizes for stupid things and has no concern with human value".

[10] Chemical signals in our brain are very slow compared to electric circuits.

[11] Scaling hypothesis.

Neuroscience Dangers. Proto-SI's[12] to *clean up the science* development in neuroscience potentially presents risks for humans. Even benevolent AI is all about optimization. What if at early stages of its development, proto-SI decides that it is optimal to sacrifice or damage some humans in order to study *in vivo* their brain development and operation? Inspired by some politic regimes in the human history – which assumed that some damage could be done at individual scale to promise the prosperity for the whole society in the future – proto-SI could justify this intervention by an urgent need to optimize its neural circuits: *"SI needs to rapidly gain in efficiency to save the world from [you name it], and to do so, it needs to study your [or your child's] brain in vivo."* Some people might volunteer for such experiments, especially if the latter are organized safely. Nevertheless, such a study could potentially pose risks for living beings (humans included), especially their young (brain architecture optimized for maximal learning capacity) and the most prominent individuals with exceptional intellectual abilities. Even if such studies could be carried on digital copies of human brain, they could be harmful and thus unethical.

2.3 The Pace of SIS Progress

Pinning the SIS Progress. Even with SI, scientific progress and discoveries will need some time. If in some domains, a "Gedankenexperimente" is not sufficient and data collection is required, the progress will slow down. Even for SIS, some experimental data are not easy to obtain, and some technologies are not fast to implement. For example, construction of a new generation of LIGO for the study of gravitational waves or of a Huge Hadron Collider to further push the frontiers of high-energy physics will require time. The same applies for new generations of fusion reactors. In biology, for example, studying some genetic information transmission in species other than drosophilas and mice can take decades. Another possible pinning point is the necessity for heavy computations. For some dynamical systems, no short-cuts exist and there is a need for direct simulations to know the system state in the future. The simplest examples are cellular automata of class 3 and 4 (Wolfram, 1984), but SI will be able to deal with much more complex dynamical systems requiring much more compute and thus requiring a lot of time and energy.

Therefore, it is clear that the progress of SIS in domains requiring data, experiments and heavy computations will be pinned, while in purely theoretical domains the progress can go very fast or even instantaneously from the human perspective. It will thus result in a non-uniform geometry of the forefront knowledge (see Section 3).

Nevertheless, even without access to the results of experiments, SI will be able to continue developing simultaneously alternative sciences based on different potential results of the key experiments (see Figure 1). As soon as the experiment in question provides conclusive results (depinning), invalidated science branches will be eliminated and the progress on valid ones will continue.

Outer space exploration. Another pinning point consists in long space missions. Contrary to numerous science fiction examples, it is quite improbable that humans will take part in outer space exploration – we are "too soft" to tolerate high accelerations, we require a lot of energy in living state, are vulnerable to new

[12] By proto-SI we understand an early form of Superintelligence, its intellectual capacity is already well beyond those of the most capable humans.

viruses, have a short life span, we do not tolerate zero gravity and we are defenseless against cosmic radiation[13]. Therefore, it is likely that the space exploration will be done exclusively by SI by its own or by means of self-replicating von Neumann probes. Nevertheless, in the near future, our "Space Odysseys" could happen in our personal universe (see Section 4.2), or in real life (if technical limits are lifted) not for the sake of science but for entertainment only. In the far future, SI will be able to find additional or alternative homes for human civilization and will help us to move there and settle down.

Following the SIS progress. Every (or almost every) considerable update of the SIS's frontier can be publicly announced and be accompanied by a kind of press release, a public conference or something else. Top human experts can try to understand these milestones or take for granted some results. Some of these discoveries could be announced and explained at seminars given by enlightened humans or SI's representatives. Some advancements could be followed in live broadcasts on the SIS' twitch or in a *Science Theater* by analogy with operating theater. Human science advances slowly, and definitely it was not possible to follow it in live, but it will be possible with the SIS.

SIS stopping scenarios. The progress in SIS can continue for long time or can eventually stop when, for instance, SI decides that the following progress is energy inefficient and its current world map allows it to achieve its goal[14]. Another stopping scenario is when the entirety of science is covered by the SI and remaining questions can be solved easily "on flight". Here, however, we face the philosophical problem of the knowability of the world and of epistemology, which will be left outside the scope of this essay regardless their high relevance for the topic we explore. But even if SI stops and is "happy" with its world model, humans can further push the exploration with or without its help.

2.4 Cleaning up Sciences

To begin building a good world model, cleaning up the legacy HS seems to be a good idea for a proto-SI. This cleaning could take start in the legacy scientific literature – entries which do not bring new knowledge, are redundant or wrong will be removed from the corpus. All biases, misinterpretations, ungrounded believe, wishful thinking and conjectures proved to be wrong will be removed. Even the most capable human minds make errors. Of course, it does not necessarily mean that the errors in scientific literature were harmful for the development of science, but they should not present in the distilled corpus. HS will benefit too from this cleaning in two ways: 1. it will give credit where it is truly due, 2. having a solid and trustful foundation will be beneficial for HS progress whatever form it will take. An eventual adding

[13]Cryonics (freezing and storage of human) presents a possible solution to a part of these problems and could potentially be helpful in space exploration if some individuals want to explore the space.

[14]This set goal or self-adjusted goal with a feedback loop (Russell, 2019), whatever it will be, at least should not consist in "exploring the totality of science", which could directly lead to catastrophic outcomes Bostrom, 2014; Yudkowsky, 2022 potentially transforming all living beings in the Universe into guinea pigs as well as the entire Universe itself.

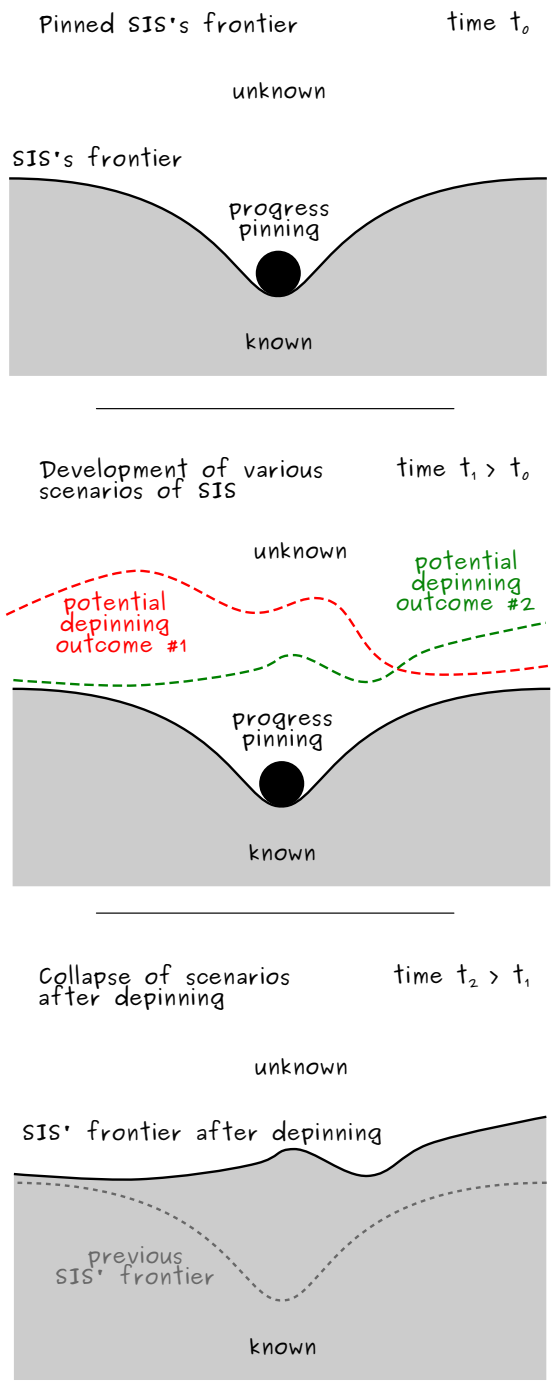


Figure 1: Pinning and depinning event in SIS's frontier progress

of a new entry to this corpus can be controlled by SI and not by randomly selected human reviewers prone to all problems of the modern peer-review system.

Credit redistribution. Regarding the benefit of credits redistribution. Cleaning up the scientific corpus will enable to get rid of currently used scientometric indicators, such as h-index, and will allow introducing new ones (if needed) handled by the SI. It would allow praising originality and true contributions to scientific progress. In the current system, the high number of papers exposing the same idea[15] results in higher visibility and spread, in bigger "weight" of the idea[16], and broader acceptance. In the cleaned-up science, a single mention of a groundbreaking idea and its first realization will give to the individual as many credits as if the same idea would be repeated many times. It can be seen as an appropriate normalization which is not easy to implement in the current scientific landscape.

Redoing human-led experiments. Since human scientists have proved that they can behave unethically and misconduct scientifically, notably by falsifying and fabricating data, the SI will need to check all experimental data by conducting an optimized set of experiments on its own. Possibly, some of rare original human-collected data can be included in the corpus with a "trust weight" enabling SI to construct some models and theories upon these data assuming and measuring risks. Such a tremendous effort in cleaning up and data reproduction will be highly beneficial for the science but also energy and time-consuming.

3 Scientific Landscape

3.1 Fractal boundary between known and unknown

The pavement of scientific knowledge, or more specifically the boundary between known and unknown, between the "terra cognita" and "incognita", can be seen to some extent as a fractal surface. On a bigger scale of global scientific domains we have a general understanding what is known and what is not, on the smaller scale of particular details, we have a more detailed vision and on even smaller, microscopic scales we can have very specific questions, which are either answered and documented or not. This fractal boundary between the known and unknown is fuzzy on the bigger scale and becomes sharper on the smaller scales of specific questions. Formulation and solution of such questions can lead to discoveries, to the refinement of our understanding, development of new methods and experiments, and of course they lead to new important subquestions[17] representing new boundary further scales in this fractal knowledge landscape. The small scales are critical for the advancement of the science and only there the true progress of the frontier can happen by pushing these smaller scales. An SI can see clearly the voids on these smaller scales and fill them in, thus resulting in gradual progress of the forefront. For steady progress, the amplitude of oscillations (standard deviation of the small scales from the average boundary) between the known and unknown should be small

[15] We mean an idea in its very broad sense – technique, method, approach.

[16] Probably, human and social sciences are more susceptible to this weight.

[17] "The purpose of models is not to fit the data but to sharpen the questions", Samuel Karlin.

enough. So I guess the frontiers of the SI-led science (SIS) can be smoother than this of the human-led Science (HS).

3.2 Hypothesis of non-overlapping scientific questions

Even though the cleaned up legacy HS with all its questions and details will be part of the proto-SI training, not all regions of its frontier will be interesting for the SI. There could be unresolved questions interesting for human individuals which do not belong to a set of questions solved by SI to construct its world map and push its own SIS (see Figure 2). Mathematically, using the notations of the set theory, it can be stated as follows: legacy HS $\subset$ SIS (legacy HS is a part of SIS) but a part of HS frontier (SI-idle questions) $\partial\mathrm{HS}_i \subset \partial\mathrm{HS}$ is not a part of the SIS $\partial\mathrm{HS}_i \notin \mathrm{SIS}$. The remaining questions of our frontier $\partial\mathrm{HS} \setminus \partial\mathrm{HS}_i$ will be of course solved by SI and will integrate the zone of known part of SIS: $\partial\mathrm{HS} \setminus \partial\mathrm{HS}_i \in \mathrm{SIS}$. This hypothesis could be formulated as follows.

Hypothesis 1 (Hypothesis of non-overlapping questions). *Unsolved questions of the Human-led Science do not fully overlap with the questions that Superintelligence explores to construct its world model.*

These SI-idle questions outside the SI interest, i.e. outside the zones of spontaneous SI scientific exploration, could present an interesting playground for humans involved in science. A forced deployment of SI in these domains can of course expand the frontiers drastically, but would it worth it? Maybe, such rare islands of science untouched by the SI would be the most valuable for human scientists. SI can be forbidden in those lands.

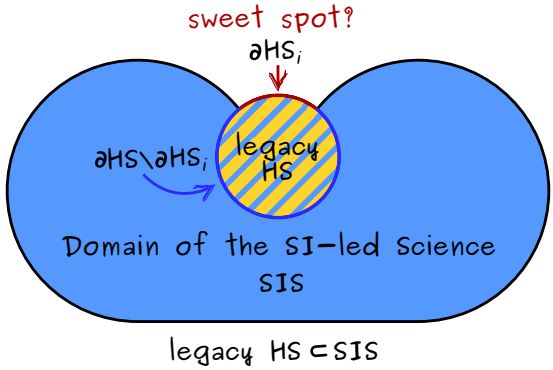


Figure 2: Domains of the legacy Human-led Science (HS) and SI-led Science (SIS) and a sweet spot of questions of "SI-idle" scientific questions $\partial\mathrm{HS}_i$ which do not interest SI for its world model, but they remain interesting for individual humans.

3.3 Scientific Map

In Fig. 3 a schematic science map is depicted with three separate domains: HS (legacy and post-SI) and SIS. The legacy Human-led Science and its accumulated knowledge with its own irregular frontier and some voids is shown in the center in golden color. It is surrounded by post-SI Human-led Science, shown in green, and SI-led science

domain (shown in skyblue). The SIS is broader than HS but also has some voids closer to its frontier and it possess an irregular boundary. This irregularity can be due to the pinned frontier (shown with (x), see Section 2.3) or due to some preferences of the SI dictated by an efficient world-model construction needs. As discussed, the SI can be uninterested in some frontiers of HS, they are marked with (o); there, the HS can continue its autonomous development (*) possibly with an assistance of the SI. Eventual "forbidden" research zones with too high risks of destructive for the humanity innovations Bostrom, 2019 – "black balls" – are shown in red with a buffer zone to prevent human's penetration. Pure HS is fixed within its pre-SI boundaries, so this legacy domain is frozen. Modern HS - the science led by humans or augmented humans probably with assistance of SI explores SIS but remains far from its boundary. The SIS' frontier will grow if judged necessary by SI pursing its goals. The further human science advancement without SI paving the path will be possible only within (o) zones of SI-idle questions (see Section 3.2). Expansion from this frontier, shown in light green, could be either purely human-led or represent zones of non-spontaneous SI exploration (pushed by humans). Otherwise, human or augmented-human "scientists" (shown with golden circles) will "travel" on this map towards SIS' frontiers on paths already paved by the SI.

To make more quantitative sense of such a schematic map, it would be reasonable to visualize it on the Poincaré projection of 2D anti-de Sitter space[18] (see a small inset in the figure).

4 Human scientists

4.1 Human capacity

We have been putting human mind on the top summit of known intelligences mainly because of its generality, adaptation capacity and high cognitive performance in abstract tasks. Our power of imagination reinforced by mathematical logic and tools, by scientific approach and computer programs is the best thing to push the science we have ever encountered. But there is no physical law that limits the intelligence[19] at our level. We can interpolate the level of intelligence from a tree, to an ant, to a dog, to a human and then extrapolate it and well beyond to a SuperIntelligence (SI) which can eventually transform itself in a hyper-intelligence and so on far beyond our perception capacity. So human scientists will operate on the new scientific landscape build by such an SI. But regardless our intellectual capacities, possibly, we will not able to even approach the frontiers of SIS even when being augmented because of the chasm separating our intellectual capacities from those of SI. Maybe, to grasp the

[18] Here, anti-de Sitter (AdS_2) space is used to identify that far from the center of the circle (where most of the HS is located) the distances, i.e. the difficulties associated with the scientific progress, become larger and large when we approach the edge (even though they are seen equivalent on the paper) and ultimately they diverge at the edge. By simplifying, we can say that most of low-hanging fruits have been picked by humans and further scientific progress is much less trivial and require much more cognitive efforts and knowledge - which is represented by the tricky metric of the AdS_2.

[19] By intelligence we understand a capacity of information processing or an ability to accomplish complex goals Tegmark, 2017.

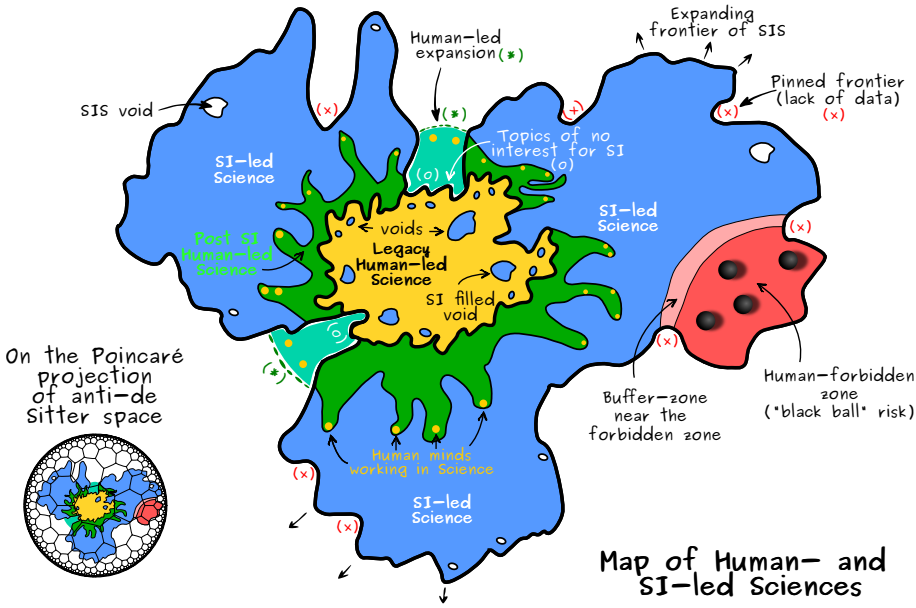


Figure 3: Map of Human-led and SI-led Science domains and its frontiers. All low-handing fruits and not-that-low-hanging ones have been already picked by humans, so the scientific progress should be measured on anti-de Sitter-like space: with distances becoming larger and large as we move towards the edge (heptagons on the map all have the same area and edge length in the hyperbolic space but on the Poincaré disk projection they seem smaller when moving towards the edge).

knowledge required to operate near the forefront, the human brain is not sufficient[20]. Factual knowledge, notions and theorems, new mathematical tools required to make a single step there, close to the SIS forefront, could be too excessive for our physiology. Or maybe, a true understanding of some phenomena or theories could be beyond our brain architecture. Human augmentation could be at least a partial solution to our limitations.

An alternative option is to believe that some of us are very capable and can go quite far on the SIS landscape. Maybe, in some domains, it will be still possible to approach the edge of SIS with years of deliberate practice with SI's help and eventual augmentation. Maybe, we will be able to grasp at least a portion of the forefront research like in "popular science", but popular one from the SI point of view, which would be well adapted to the skills and capacities of professional scientists. We will understand the theorems but not proofs.

Human augmentation. Pushing towards the forefront of HS and SIS can be done at different intensities. It could be done on your own, without external tools, or only with a seamless soft guiding by SI, but it can be also done at ultra-high speeds when you decided to seriously augment your mind biologically or by

[20] Possible limits: finite storage, small operational speed, possible limit of number of concepts it can memorize.

strong cyborgization[21]; this pursuit can be done in real world or in a digital one. The augmentation could be limited to the memory capacity or concern only the mental side, it can include mood regulation and other brain-chemistry related and physiologic aspects. We will be able to choose the pace of SIS discovery and, of course, we will be able to test different ones. However, without memory erasing, degrading to less capable versions of yourself could make you suffer: imagine switching from a near-speed-of-light advancement speed to a snail pace. But anyway, the pace of individual progress is not the main thing, we will learn (or be rewired) to be happy from the process itself because the very SIS forefront could remain unreachable anyway.

Getting to the SIS forefront. Can human minds reach and stay at the forefront of SIS? This question should be addressed knowing that the route will be paved by SI and we will be guided by SI, *i.e.* no discoveries or breakthroughs are needed from us, only learning new tools and understanding the progress made. Since we do not need to walk randomly taking fruitless paths or dead ends, this progress towards the frontier can go relatively fast. However, this no-creation way of advancement can be psychologically hard for humans – imagine being a scientist nowadays and write no papers but exclusively read papers of others. Therefore, to keep humans in the loop, some compensation mechanisms should be invented. For example, some indicators can be assigned to scientists to measure the progress towards the forefront or at least from the HS legacy frontier. Another compensation mechanism could be chemical or direct brain adjustment to enhance the positive feedback from learning something new.

Hard skills. To progress towards the SIS forefront, we will need to ask smart and relevant questions, it would be impossible without strong *hard skills*, which we will need to develop beyond the current level to dive into SIS. A good analogy for the need of hard skills is the following. Imagine a child asking scientific questions without mastering mathematics - the language of science. You can answer their question about why the sky is blue in simple terms, but with such an explanation, rigorously their understanding will be incomplete. Therefore, to practice HS we will need to cultivated deep thinking competence and to be well-equipped with knowledge and mastery of mathematics[22]. This view contrasts a widespread forecast that a need in hard skills will vanish in the age of SI and mainly soft skills will matter.

On the other hand, it will be possible to upload directly in an augmented human brain the forefront knowledge in a particular domain or the entirety of scientific knowledge. This uploading or simply fast and seamless access to this knowledge

[21]To give an example of a cyborgization, in a hard science fiction trilogy "Astrovityanka" (in Russian, 2008-2010) by Nick Gorkavyi (also available in English as "AstroNikki" (in English, 2020)), the main character, teenager Nikki has a supercomputer which she always carries in her backpack and which is connected to her body enabling control of her body bypassing her damaged spinal column; she can interact with its AI as we do now with modern chatbots: request information, ask for analysis and simulations, however, this communication can happen through a kind of soundless ventriloquism, her robot also has permanent access to Nikki's five senses. This science fiction becomes reality with implantable brain-computer interfaces which has been successfully developing by Neuralink.

[22]At least to practice science in our natural, not augmented state.

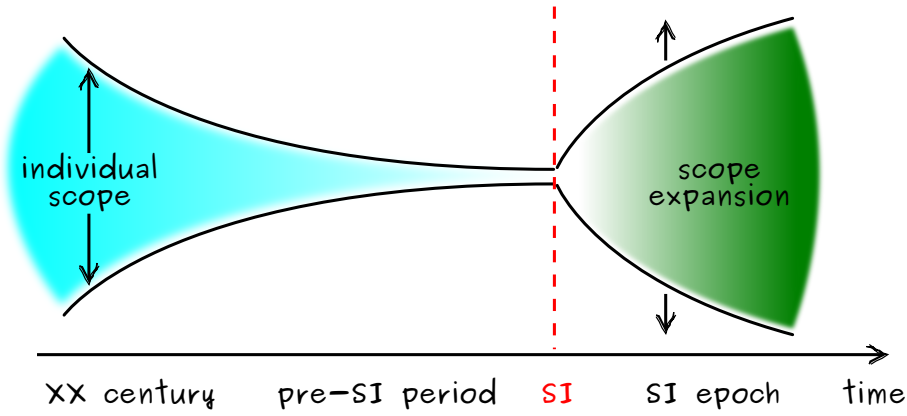


Figure 4: Expansion of individual scientific scope in the age of SI.

could be instantaneous making our brain believe that it masters and understands it on its own[23]. The illusion of understanding everything and performing on the very forefront of science, a "god mode", can be an exceptional experience which is hard to visualize today.

Evolution of individual scope. Over the last century, scientists have become quite *narrow* experts in specific topics. With SI in the loop and an eventual augmentation, they will be able to broaden our individual scientific scope and operate within a broader scientific domain, Figure 4. SI will provide them with a better vision of science in general and of the frontier.

SI language and a new toolbox. Are our tools, methods and mathematics well suited to pushing further the science frontiers and go beyond the current understanding of the world? At least, for the current generation of scientists today's mathematics will remain the most mastered and thus the perfect tool for this endeavor. Our mathematics, notations and tools are made for humans. SI will probably use its own language/code and its own representations of objects and notions, which will be completely obscure to humans and thus an adaptation and translation might be required. It is another argument in favor of the asymmetry of information sharing. Maybe, SI will provide us with different tools and even different mathematics for our own pursuit. As it often happens in mathematics, a new toolbox can help to resolve old questions and conjectures, the pinning points. These new, more adapted tools for the scientific exploration can be integrated in our educational system. The legacy literature will be outdated and useless, and people wishing to train themselves in science will have personalized SI tutors and textbooks written specifically for them. Therefore, there could be a sharp separation between the old and new generations of scientists speaking different languages[24]. However, it would

[23] Such an "understanding" could be, however, interpreted differently if one follows Roger Penrose theory of non-computable human intelligence Penrose, 1989.

[24] But of course an SI-assisted translation will be helpful here. However, if the new mathematics will be much more elegant and compact, its translation to the legacy mathematics can be problematic.

be important to teach the same new mathematics with the same notations to the new generation of scientists to let them interact.

Emotion. Human excitement, curiosity and a sense of achievement are very important for scientific progress. This emotional boost is purely human and is not required for a machine to construct its world model and push SIS. In the age of SI, the true scientific discoveries will be replaced by enhanced learning, which is a completely different endeavor. Nevertheless, excitement and curiosity will remain there. Discovering new things makes a lot of sense even though it is not on the forefront of SIS or even HS. So if the utilitarian function of HS will vanish anyway in the solved world, the excitement part will remain there if, of course, humans decide to keep the current mind settings.

Only the sense of being useful and pushing the real frontier of science will be missing. So if humans for whom this part is crucial want to stay in the loop, this emotional aspect can be artificially added in the post-instrumental world. The lack of meaning in rediscovering of SIS instead of pushing the real frontier can be erased. A further boost can be added by human's augmentation, which would represent, I think, a common practice in the HS in the future. The level of this augmentation could be deliberately selected and adjusted. This augmentation in terms of cyborgization or biological enhancement can provide us with higher mental ability and also can provide us with tools enabling us to advance at desired pace to the frontiers of SIS.

4.2 Social aspect and its emulation

Social network. The community and the societal aspect of science are very important. These aspects could however be mimicked to let people engaged in HS feel good and surrounded by like minds. Social networks could be adjusted for individual scientists to provide them with additional (external) motivation and inspiration. Some ego-promoting aspects can be deployed: people, whatever their contribution, can become Einsteins in their small individually-tuned social universes.

Personal universe and gamification. If having a social network with emulated agents praising you is not enough, you can be emerged in your personal universe adjusted for your values and your objectives as suggested by Yampolskiy (2019). So you can operate in any epoch and in any scientific role: be a star of quantum mechanics in early 20th century or be a scientist in a space expedition to a new imaginary inhabited world in our galaxy, you can even appear in a completely new world with different fundamental constants or completely new physics laws imagined by SI, and discover this new world as a pioneer scientist without a priori knowledge of it. You can adjust your mind to be part you and part a scientist of your choice with his/her mindset too, you can choose to be Henri Poincaré and work on the 3-body problem or, e.g. Galileo Galilei and discover Jupiter moons with your handmade telescope. Of course, all the environment of these epochs could be adjusted according to your preferences too. If needed, your memories of the modern science could be removed to let you fully enjoy the experiment.

You can live a life of Magister Ludi from Hesse's "The Glass Bead Game" practicing a pure abstract mixture of art and science in a ivory tower. However, the question of keeping yourself remains very relevant. If you want to keep yourself, you'll need to limit augmentation and mindset change. However, you could keep the option to return to the original settings and decide whether to keep or not all

memories after spending a period in one universe. The implementation of a personal universe can take different forms. It can be projected to your five senses if you decide to keep your brain in our physical world, or it all can happen in a digital world with you uploaded mind. The second seems to be technically simpler and energetically more efficient.

In your personal universe, you don't necessary push the science and do not necessary advance to SIS but can be simply happy doing science and living a fulfilling life.

Scientific writing. Of course, in the era of SI, the classical knowledge exchange and education through attending scientific conferences, writing and reading books and mainly peer-reviewed publications, will be deprecated. Most of the knowledge, outside forbidden domains, will be accessible through chat-like interfaces or directly through brain implants. But there should be some artificial incentives for humans to produce written texts which have a structuring foundation for their personal development, consolidation of the acquired knowledge and deep thinking practices. So, some forms of scientific "publications" reviewed by SI, will exist.

4.3 SI-HS relations

SI mentor. SI can help scientists to move forward towards the frontiers of SIS by guiding, motivating, and encouraging them. With such a wise mentor and colleague they will be able to focus only on relevant questions and limit their distractions, i.e. it will not let them exploring unneeded dead ends. So, their personal progress, being directed, will be more efficient[25].

The level of SI guidance can be adjusted, for example, instead of an explicit guidance to the questions that matter, scientists could get subtle hints that they could decipher: planets in a special arrangement, signs that could be noticed during their morning rambling in a new city they are visiting, a quotation that they can occasionally read on a wall of a café where they have a morning coffee. Of course, such subtle hints will be easier to implement in a digital universe[26].

When the questions of your scientific research are established and a path partially paved, SI can provide you with adapted reading, adjusted for your knowledge and brain's preferences in terms and in form, with probably some passages deliberately kept difficult, so it takes you some time to understand. If you learn some experimental techniques unfamiliar to you, SI as an experienced technician can test them and deploy to let you understand something new in biology or particle physics. So, you are encouraged to increase your mastering of the topic to pass some barriers and achieve the goals: prove a theorem, explore specific functions of a protein or refine your understanding of strong interaction. Of course, SI knows your current intellectual and work capacity, missing knowledge and your potential to make your own discovery and overcome the difficulty, so the problems that you face can be perfectly adjusted, and you will be convinced "you can do it". When we are sure that we can, then we truly can. Alternatively, it could be a finer message: "You can do it with 30% probability". It would be more challenging but more rewarding. So, in such

[25] Instead of a random walk advancing a distance proportional to a square root of time $\sqrt{t}$, the progress can be linear $\sim t$ and follow the most efficient geodesic trajectory.

[26] See "Protector god" AI aftermath scenario from Tegmark, 2017.

a set-up with a benevolent SI mentor, you can really surpass yourself and show the best of you. "I thought that it was beyond your capacities, but you've finally proven this tricky theorem! Bravo!" - SI could deceive you with flattery in the end.

SI oracle and its priests Future scientists can serve as priests and priestess in a temple of SI oracle (analogous to a research center) by asking smart questions and documenting the answers in an "ultimate book of knowledge" (a kind of database). It could take a form of a game in a solved world among other games of science we can imagine. Even though such activity of knowledge extraction could be seen as useful in the world of oracle-only SI, it would lose most of its meaning in the world of a more probable agentic SI, because if we need a database of the totality of scientific knowledge, we will simply need to ask it to create such a database for us. It could be done at least for the set of questions about which the humankind can be curious during the upcoming millennium.

Strictly speaking, there is no utilitarian function in such a kind of interaction with the oracle and in a "manual" population of the knowledge database. Nevertheless, such a gamified activity requiring a lot of intellectual efforts presents an interesting occupation for humans in a solved world. The priests and the priestess will not only formulate questions but also will work hard to understand the answers and its implications. This task of interpretation could be further (deliberately) complicated by special settings of the oracle. For example, the oracle should not simplify its answers; it should formulate them in a new language that priests and priestess will learn during first years of their service; it should, as Oracle of Delphi, sometimes provide ambiguous and enigmatic answers which could require additional calculations, experiments and verifications. In other settings, for example, the oracle can answer only a limited number of questions per day or in total. Many aspiring rituals could be invented to accompany this worship in the temple of SI. Like in the "Castalian Glass Bead Game", being a priest or a priestess of the SI oracle could be an interesting intellectual activity in a solved world. Ultimately, such a database stored for eternity could be helpful in a scenario when SI lets us alone.

Entrust some problems to humans. In some situations, SI can provide trained human scientist with some minor problems on the frontiers of some research domains where the progress does not impact SI world-model construction. The reason for this exploitation of Human Scientists could be simply humanistic with the objective to provide highly skilled human scientist with a real meaning (in the contrast to an imaginary one). However, it must be a gesture of goodwill because such a delegation will be inefficient in terms of energy and time. Another question is whether humans will volunteer at all? They could be seduced or convinced by SI in a way that they will not regret their choice after having pushed their small portion of the HS/SIS frontier. However, I think it is utopian to believe that humans could be helpful in the SIS forefront exploration.

Human emancipation from SI. I am confident, that in the era of SI, there will be groups of people refusing to deal with SI or its derivatives. Probably, they will possess rare islands of HS. Such groups will try to explore the science independently because of lack of trust in SI, because of some religious, mystical or conspiracy considerations, because of fear or because of the will to have pure freedom and independence even in exploring forbidden "black-ball" domains. Eventual justification which could be used, even in case of a full or partial trust to SI, is "Who knows what

happens tomorrow with this SI, at least we will have our own knowledge database", but for SI-emancipated groups it could range from cautious warnings "We have to push science and technology on our own to be capable of protecting our species" to conspiracy-like justifications: "SI does not let us in forbidden domains because it is afraid that possessing this knowledge could make us better than SI on doing something".

4.4 Meaning and purpose

Probably the main questions in the post-instrumental world will be related to the meaning and purpose of human beings. Why would someone want to work hard and move towards the frontiers of SIS when they could simply enjoy life in its personal universe? There could be various personal motivations for doing so: curiosity to learn how the Universe operates, expanding the limits of today's science understanding of the humankind, preserving this humankind legacy competence of scientific pursuit, only partial trust in SI's benevolence or a will to preserve the knowledge on the human side too, a fear of an uncertain future, ability to ask intelligent questions and think deeply, etc. Apart from these idealistic or fear-dictated motivations, a simpler, more egocentric motivations could be considered too: increasing our own metrics[27] whatever they are, peers recognition, self-pride and so on. Alternatively, more artificial motivations could be integrated in our brain on our will.

Nowadays, apart from pushing their scientific domains, scientists could find a lot of meaning in transmitting their knowledge, vision and approach to early-career scientists and students. In the age of SI, this role will drastically diminish. SI could take care of all aspects of education by creating the most fruitful and motivating environment for young students to help them to acquire all knowledge and master all scientific tools with the most advanced mental coaching fitting their personality and mindset. Therefore, creating new courses, video lectures, writing new books and papers for teaching needs cannot be considered as meaningful occupation for scientists of the future; this activity could remain meaningful only for the purpose of self-improvement. However, a real-life dialog and human-to-human interaction will still make sense in the real world but not in the virtual one. Promoting such inter-human connections will be a meaningful and rewarding mission for humans. If this interaction is freed of some egocentric and career-related conflict points existing in today's scientific communities, it will enhance individual progress and could be a source of meaning for some scientists.

Truly collaborative approach to scientific advancement is often mutually beneficial in the sense that a collaboration of two intelligence is more than their sum. Moreover, a really strong coordination could be organized by SI by softly pushing individual scientists with matching approaches and personalities to create fruitful and coherent collaboration groups.

Purpose. In Amodei, 2024, the author argues that even in our epoch, a lack of economical value does not make things valueless or less meaningful: it concerns many non-professional human activities. Already today, fundamental sciences, at least at the scale of individual scientists, can be considered as deprived of an explicit economic meaning. For scientists, the progress on its own and the act of creation

[27] Such metrics computed by SI could integrate meaningful normalization taking into account our particular brain capacity and abilities.

or discovery produce a lot of meaning *per se*. However, in a solved world, scientists will lack these aspects at least in their modern sense. All human contributions to the global forefront of science, will be marginal or totally absent. Therefore, the meaning should come from other things. HS will be a very intellectually stimulating human activity, which will be exceptionally rare in a solved post-instrumental world. Already today, many people blindly rely on LLMs in all their daily and professional questions. This trend will be amplified in the future potentially leading to an intellectual handicap of humans. Therefore, doing science in a solved world can preserve to some extent a singular role of humans and of our biological intelligence. People practicing science will align with stoics' ideal of "living in accordance with Nature" as our nature enables us to think deeply and to carry out such a highly intellectual activity as science.

5 Conclusion

In a solved world of SuperIntelligence (SI), SI-led science (SIS) will push the frontiers of science at the speed of light beyond the legacy human-led science (HS). Whatever the SI goal will be, it will require to construct an accurate world model for its purpose, which should require doing science. Therefore, HS and SIS will be separated and will share information asymmetrically: everything known to humans will be known to SI but not vice versa. Probably, a benevolent SI will have incentives to restrict human access to some scientific domains by smoothly deflecting our efforts to not let us pull a "black ball" of a devastating technology, which is formulated in this short verse:

> I struggled at the sky,
> Where definitely maybe
> Deflected my attempts
> At theory's assembly.

SI will clean up the legacy scientific literature keeping only essential contributions and redistributing credits to human scientists. In this task, having only a partial trust in human experimental data, SI will carry out numerous perfectly designed experiments. This new corpus of scientific knowledge and experimental data, will make a foundation for further scientific progress. This progress will not be uniform across all disciplines because of a need of heavy simulations, construction of new experimental equipment, long-lasting experiments or time-taking space exploration.

However, some scientific questions will not interest SI for its world model construction and those domains could be a reservation for human science with forbidden SI. Even withing SIS, human scientists pursuing scientific research can find their meaning and purpose by preserving this highly intellectual activity of the humankind. Being trained in the most efficient and adapted way, and further being guided by SI, these humans can go well beyond the legacy HS and create a new, post-SI HS whose frontiers will approach those of SIS. To accelerate this progress, humans can be augmented mentally, motivation-wise, or truly cyborgized. Anyway, augmented or not, humans in science will no longer be true inventors and creators, we will instead become learners and explorers of already paved paths. This rare activity requiring enhanced hard skills will present an intellectually stimulating occupations for humans in a solved world.

In many scenarios, in a solved world, most of human activities will transform into games. A similar gamification is expected for the science, which is considered today as one of the most serious and demanding human occupations. We can

become priests in the science temple of SI oracle and pretend to serve the humanity by collecting knowledge, or we can play a science game in our personal universe by taking the place of Einstein of Bohr in the beginning of the 20th century. We can become science stars in our own emulated social network, or discover physics in an imaginary universe with different laws of physics created by SI for our exploration.

This essay represents an initial reflection on the future of science and possible roles of humans in the scientific endeavor. Regardless only superficial coverage of many questions and ideas, it is clear that human-led science will have multiple facets and can take various forms. In conclusion, human-led science will preserve a singular role of humankind as a truly intelligent species; regardless its lack of utilitarian function, science will still make sense in a solved world because it will provide humans with a profound and lofty meaning in accordance with their nature.

References

Alexander, Scott (2014). *Meditations on Moloch*. https://slatestarcodex.com/2014/07/30/meditations-on-moloch/. Blog post on *Slate Star Codex*, published July 30, 2014. (Visited on 08/21/2025).

Amodei, Dario (Oct. 1, 2024). *Machines of Loving Grace*. URL: https://www.darioamodei.com/essay/machines-of-loving-grace (visited on 08/15/2025).

Bostrom, Nick (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Bostrom, Nick (2019). "The Vulnerable World Hypothesis". In: *Global Policy* 10.4, pp. 455–476. DOI: 10.1111/1758-5899.12718.

Bostrom, Nick (Mar. 2024). *Deep Utopia: Life and Meaning in a Solved World*. Ideapress.

Grinbaum, Alexei (2024). "Responsible research and innovation". In: *Handbook of Technology Assessment*. Edward Elgar Publishing, pp. 409–417.

Grinbaum, Alexei and Christopher Groves (2013). "What Is "Responsible" about Responsible Innovation? Understanding the Ethical Issues". In: *Responsible Innovation: Managing the responsible emergence of science and innovation in society*. John Wiley & Sons, Ltd. Chap. 7, pp. 119–142. DOI: 10.1002/9781118551424.ch7.

Kokotajlo, Daniel, Scott Alexander, Thomas Larsen, Eli Lifland, and Romeo Dean (Apr. 2025). *AI 2027*. URL: https://ai-2027.com/ (visited on 09/02/2025).

MacAskill, William and Fin Moorhouse (Mar. 2025). *Preparing for the Intelligence Explosion*. Released on 11 March 2025. URL: https://forethought.org/research/preparing-for-the-intelligence-explosion (visited on 09/03/2025).

Moravec, Hans (1988). *Mind children: The future of robot and human intelligence*. Harvard University Press.

Moravec, Hans (1998). "When will computer hardware match the human brain". In: *Journal of Evolution and Technology* 1.1, p. 10.

Moravec, Hans (1999). "Rise of the Robots". In: *Scientific American* 281.6, pp. 124–135.

Omohundro, Stephen M. (2008). "The Basic AI Drives". In: *Proceedings of the 2008 Conference on Artificial General Intelligence 2008: Proceedings of the First AGI Conference*. Ed. by Stan Franklin Pei Wang Ben Goertzel. NLD: IOS Press, pp. 483–492.

Penrose, Roger (1989). *The emperor's new mind: Concerning computers, minds, and the laws of physics*. Oxford University Press.

Russell, Stuart (2019). *Human compatible: AI and the problem of control*. Penguin Uk.

Tegmark, Max (2017). *Life 3.0: Being human in the age of artificial intelligence*. Knopf Doubleday Publishing Group.

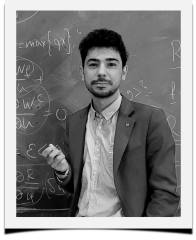
Wolfram, Stephen (1984). "Cellular automata as models of complexity". In: *Nature* 311.5985, pp. 419–424. DOI: 10.1038/311419a0.

Yampolskiy, Roman (2016). "Taxonomy of Pathways to Dangerous Artificial Intelligence." In: *AAAI Workshop: AI, Ethics, and Society*, pp. 143–148.

Yampolskiy, Roman (2019). *Personal Universes: A Solution to the Multi-Agent Value Alignment Problem*. arXiv:1901.01851 [cs.AI]. arXiv: 1901.01851. (Visited on 08/16/2025).

Yudkowsky, Eliezer (2008). "Artificial Intelligence as a Positive and Negative Factor in Global Risk". In: *Global Catastrophic Risks*. Ed. by Nick Bostrom and Milan M. irkovi. New York: Oxford University Press, pp. 308–345.

Yudkowsky, Eliezer (June 6, 2022). *AGI Ruin: A List of Lethalities*. URL: https://www.lesswrong.com/posts/uMQ3cqWDPHhjtiesc/agi-ruin-a-list-of-lethalities (visited on 08/15/2025).



Dr. Vladislav A. Yastrebov is a CNRS Research Scientist in Computational Mechanics at Mines Paris - PSL. His research interests span a broad range of fundamental and practical aspects of physics of solids, fluids and their interaction. He carries out a theoretical research in numerical methods and applies them to simulate complex nonlinear phenomena involving coupled multi-physical aspects. Vladislav is also an active user of modern AI tools and has created several AI agentic applications. This essay presents his attempt to reflect on the future of his profession in the age of superintelligence.