

Philosophy Beyond the Rational Mind?

Human Potential for Transformation in the Mirror of the Quest for Artificial Superintelligence

By Regina Sibylle Surber

March 2, 2021

The basic idea of the research field of artificial intelligence (AI) is to create software that can solve intellectual tasks without and in place of humans.¹ Many theories and methods of AI research are based on theories of rational thought.² Some are even modelled on the structure of the human brain.³ This means that much AI research is looking for ways that humanity can design technologies that can replicate what has been previously been considered a unique ability – rational thinking. For this reason, we can call the goal of AI research the *imitation of rational thinking*. The goal of many researchers is to completely imitate human rational thinking. This so-called 'strong AI' could learn and understand every intellectual task analogously to how humans do these things.⁴ If the artificially created intelligence even surpassed human intelligence, one would be justified in speaking of artificial superintelligence (ASI).⁵ Strong AI and ASI are hypothetical states, whose feasibility scientists still strongly argue about. Many dismiss especially ASI as mere science fiction.⁶

Whether ASI will ever exist is not the subject of this article. One cannot observe a world with ASI today and we should not fantasize about scifi scenarios. The following thoughts are based on a different observation: AI research *strives* for ASI. In other words, ASI is AI's ultimate research purpose. For, research seeks to artificially produce rational thought always better and better. This 'ever-better' creation of AI would only stop at the point where AI becomes so sophisticated that it would take on this creation itself – as 'Strong AI', or eventually as ASI.

Observing that many are striving for ASI seems fair: human curiosity and our research drive are nothing new. They did not stop before trying out the atomic bomb. Also, AI research already has a strong grip on war practices of investment-rich states, offering them, e.g., highly automated swarms of combat drones.⁷ AI research is also advancing at an enormously rapid pace and is being

¹ Today, AI can, e.g., generate near-fit movie suggestions (e.g., Netflix), distinguish medically important from less important body values, see, e.g., Amisha et al, 'Overview of Artificial Intelligence in Medicine,' Journal of Family Medicine and Primary Care 8, no. 7 (July 2019): 2328-31, https://doi.org/10.4103/jfmpc.jfmpc_440_19, create photorealistic images and videos, see, e.g., Generative Adversarial Networks: Antonia Creswell et al, 'Generative Adversarial Networks: An Overview,' IEEE Signal Processing Magazine 35, no. 1 (January 2018): 53-65, <https://doi.org/10.1109/MSP.2017.2765202>. or analyze and create human language, see, e.g., Natural Language Processing: K. R. Chowdhary, 'Natural Language Processing,' in Fundamentals of Artificial Intelligence, ed. K.R. Chowdhary (New Delhi: Springer India, 2020), 603-49, https://doi.org/10.1007/978-81-322-3972-7_19.

² Stuart J. Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall Series in Artificial Intelligence (Englewood Cliffs, N.J: Prentice Hall, 1995), 6-7.

³ The human brain serves as a source of inspiration for so-called 'deep learning'. See, for example, Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning, Adaptive Computation and Machine Learning* (Cambridge Massachusetts, London England: The MIT Press, 2016).

⁴ Russell and Norvig, *Artificial Intelligence*, 29.

⁵ Nick Bostrom, 'Ethical Issues in Advanced Artificial Intelligence,' in *Maschinenethik und Roboterethik*, von Wendell Wallach und Peter Asaro, ed. Wendell Wallach und Peter Asaro, 1. Aufl. (Routledge, 2020), 69-75, <https://doi.org/10.4324/9781003074991-7>. Der Moment, in dem die KI die menschliche Intelligenz übertrifft, wird oft als Technologische Singularität bezeichnet, 'TECHNOLOGICAL SINGULARITY by Vernor Vinge', abgerufen am 18. September 2020, <https://frc.ri.cmu.edu/~hpm/book98/com.ch1/vinge.singularity.html>.

⁶ Famous singularists are, for example, Raymond Kurzweil: *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*, A Penguin Book (London: Penguin Books, 1999)., Ray Kurzweil, *The Singularity Is near: When Humans Transcend Biology* (New York: Penguin books, 2006). or Bill Joy: 'Why the Future Doesn't Need Us | WIRED', accessed 18 September 2020, <https://www.wired.com/2000/04/joy-2/>. Skeptics are Jaron Lanier, chief strategist by Microsoft, Mitch Kapor, founder of Mozilla, Microsoft co-founder Paul Allen, or Jaan Tallin, co-founder of Skype. For a good discussion and juxtaposition of singularists and skeptics, see, for example, Kurt Andersen, 'Enthusiasts and Skeptics Debate Artificial Intelligence,' Vanity Fair, accessed 18 September 2020, <https://www.vanityfair.com/news/tech/2014/11/artificial-intelligence-singularity-theory>.

⁷ Turkey, in particular, recently made headlines with its recently acquired 'Kargu' type swarm drones: STM, 'STM - KARGU', STM, accessed 18 September 2020, <https://www.stm.com.tr/kargu-autonomous-tactical-multi-rotor-attack-uav>.

vigorously funded. Experts estimate global AI investment at 180 billion Swiss francs as of 2025.⁸ These observations raise an important question:

What light does the pursuit of ASI shed on human beings?

Our ability for rational thinking has given the homo sapiens a prominent status in evolution.⁹ This is why – at least up until now – on earth, his species is confronted almost only with itself.¹⁰ With ASI, however, humans would create a being that would be intellectually, possibly even physically and biologically, stronger than themselves, against which they would possibly 'lose'.¹¹ Hence, a striving for ASI would be a striving for a change in the human species' competitive conditions. These changes would be such that the characteristic 'ratio', which makes humans highly adaptive and their existence hard to jeopardize, may no longer guarantee their survival. Hence, the human would need to change in order to survive. Most importantly: this change would have to be independent of his capacity for rational thinking. The following question seems imperative:

Do there exist new development possibilities for the human that have nothing to do with rational thinking?

Figuratively speaking, we may put the question thus: if the pursuit of ASI were a mirror in which the human looked at his own reflection, could this human being recognize in himself new possibilities for development that are independent of rational thinking? Could the human being – biologically, and intellectually – remain as it is, and at the same time grow beyond his rational thinking? Could he acquire a kind of intellect that is qualitatively different from rational thought?

And, if so, how? Would this be an evolutionary path by which the human could develop a new biological trait? Unfortunately, this can hardly be answered in advance. For, new genetic characteristics always only ever develop by chance.^{12 13} Evolution is not a creative process. It has no goal.

⁸ 'Global AI Investment to Top £150 Billion by 2025', *Outside Insight* (blog), 31 July 2019, <https://outsideinsight.com/insights/global-ai-investment-150-billion-2025/>.

⁹ 'Evolution des Geistes: Wie der Mensch das Denken lernte', accessed 18 September 2020, <https://www.spektrum.de/magazin/wie-der-mensch-das-denken-lernte/828592>.

¹⁰ However, the battle between humans and microbes does not seem to be decided yet. At the same time, it also seems plausible to say that human ratiocination does not ill-equip humans against pathogens, having achieved some notable successes so far in the history of epidemics (e.g., plague vaccination, successful containment of cholera pandemics, smallpox vaccination and control programs).

¹¹ Mithilfe von Tissue Engineering oder 'Gewebezucht' wird biologisches Gewebe künstlich hergestellt, siehe z.B. Aldo R. Boccacini et al., 'A Composites Approach to Tissue Engineering', in 26th Annual Conference on Composites, Advanced Ceramics, Materials, and Structures: B: Ceramic Engineering and Science Proceedings (John Wiley & Sons, Ltd, 2008), 805-16, <https://doi.org/10.1002/9780470294758.ch90>.

¹² New traits are always a kind of by-product of 'mistakes' that occur when genetic material is passed on from generation to generation. When the father's sperm cell and the mother's egg cell fuse, the very first cell of the new human being is created. It contains the genetic material (genome). This contains - in encoded form - the entire building instructions for the later human body it will become. With each cell division, the entire genetic material is copied and passed on to the two new cells. Errors can occur during this copying. 'Evolution des Geistes'.

¹³ Ideas of the so-called 'transhumanism' could perhaps be described as a kind of 'artificial evolution'. Transhumanism discusses ideas how humans could be artificially enhanced by means of new technologies and compete with potential ASI. So far, two ideas are roughly discussed: Either humans merge with technological implants and become some kind of hybrid being that could compete with ASI. See, e.g., 'Elon Musk Unveils Plan to Build Mind-Reading Implants: "The Monkey Is out of the Bag,"' the Guardian, 17 July 2019, <http://www.theguardian.com/technology/2019/jul/17/elon-musk-neuralink-brain-implants-mind-reading-artificial-intelligence>. Robert M. Geraci, *Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality*, Reprint Edition (New York; Oxford: Oxford University Press, 2012). Or research may succeed in manipulating human genetic material in such a way that human biology would be more difficult to eradicate. See, for example, Raya Bidshahri, 'The Power to Upgrade Our Own Biology Is in Sight-But Is Society Ready for Human Enhancement?', Singularity Hub (blog), 15 February 2018, <https://singularityhub.com/2018/02/15/the-power-to-upgrade-our-biology-and-the-ethics-of-human-enhancement/>. However, there are two problems with these ideas: on the one hand, such upgrades would likely come at a huge cost, which is why not all people could afford them. An imbalance between technically-biologically improved people and those who are not would be very difficult to justify ethically. On the other hand, especially ideas of genetic manipulation with the aim of a strongly prolonged human life bear

Here, though, we are asking about a possible development towards a specific goal – the transcendence of rational thinking. Hence, if this path of development exists and should be followed, the human being must somewhat become ‘active’. Therefore, this would not be an evolutionary development, but rather a path of an ‘intentional human transformation’.

The question of how the human could grow beyond his rational thinking must therefore be actively addressed. A big hurdle is the following: can the thinking human *think* at all about this question? On the one hand, the existence of an a-rational developmental path should be *acceptable* to rational thinking at least. Otherwise, this text itself would be nonsensical. Yet, on the other hand: is it possible to tread an a-rational way of development by recourse to rational thinking?

For this to be possible, human beings would have to be able to somewhat evade their rational thinking by thinking rationally. Is this feasible, or can the rational human always only think along the ‘wall of its rational thought processes’? Would the human perhaps not have to think rationally about its rational thinking, but – maybe first and foremost – *look* at his rational thinking in a different way? Yet, how would he get there? Would rational thinking have to make room for a kind of intuitive will of transformation? Is such a will already slumbering in the human being itself, and would he ‘merely’ have to open himself up to it? Or would the human first of all have to create this will? Also, how could one translate such a new intellectual element, if it were found, into a globally understandable language? Who would be the translator?

We are confronted with a cascade of questions. But, should we even ask them? We may just put the transformation problem on the back burner, or else dismiss it as a gimmick on the grounds that ASI may never exist anyway. At the same time, though, a confrontation with this question may have a great potential for humankind:

It seems plausible to say that today’s human self-understanding has created, and is still creating, a global togetherness that has produced, and still is, producing many crises.¹⁴ Of course, one would be doing him wrong if one were to reduce the human to his rational thinking. He also possesses a high emotional and physical intelligence.¹⁵ But there still seems to be a barrier to a holistic perception and a fruitful harvest of all his innate aspects. If not, in view of all these extraordinary characteristics, would the human not treat himself and others with more respect?

Perhaps the path of transformation that humans could take in light of their striving for ASI could be a possibility for a more holistic self-perception and perception of others? Maybe the aspired ASI offers the human a chance to grow into a more satisfied reasonableness, even into a cultivated insight?¹⁶

What would be the role of a successful philosophy in this context?

First of all, one would have to set up a clear argument for the fact that one should just simply consider the existence of such a human potential for transcending rational thought. At this point, one may only state that the right to claim that this potential does not exist is not stronger than the right to claim that it does.

the risk of an overpopulation of the earth. Without changes in the world economic and financial system, this would probably exert enormous existential pressure on large parts of the earth’s population.

¹⁴ And some were, of course, wonderfully managed.

¹⁵ See, e.g., John D. Mayer, Richard D. Roberts, and Sigal G. Barsade, ‘Human Abilities: Emotional Intelligence’, *Annual Review of Psychology* 59, no. 1 (January 2008): 507–36, <https://doi.org/10.1146/annurev.psych.59.103006.093646>.

¹⁶ Since this potential could probably be seized by every single human being, it could also transcend all national, social, religious or other borders and hierarchies. Possibly, everything that arises from this potential would be equally independent of these borders.

Then also, philosophy could accompany and support the human on this path by systematically raising the questions outlined above, by working on them, and by seeking answers. In this sense, successful philosophy would take a current epoch deeply seriously, and, at the same time, be a reflexive instrument on the way of a human 'becoming' to a possible new manifestation of the human 'being'.

The examination of the described mirror image could then perhaps also be informative for the question whether ASI will ever exist. Who knows, a new human self-understanding may possibly relativise the urge for the always-better, at least a little?