

What is the Singularity?

“... we are on the edge of change comparable to the rise of human life on Earth.”

– Vernor Vinge [1]

- The modern notion of [the Singularity](#) is usually attributed to computer scientist and SF writer Vernor Vinge.
- Vinge asserts that the creation of entities with greater than human intelligence will most likely occur within a couple of decades and lead to an exponential “intellectual runaway”. This event, [the technological Singularity](#), will render our current world models obsolete as extrapolation on them breaks down, and new models must be applied to keep up with the change. Other analogues:
 - At [the Singularity](#), a plot of the technological/intellectual progress curve jumps up nearly vertically, as if into a mathematical singularity (though in actuality, our curve may have to level off as physical limits are reached).
 - [The technological Singularity](#) can be seen as having a kind of event horizon, like a black hole (or a space-time singularity); it is impossible for us to predict what exactly will the world be like on the other side, lacking the intelligence of the beings that will mold it.

Transhuman intelligence

“Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever.”

– I. J. Good [2]

- **Transhuman intelligence** (or superintelligence) means a general intelligence that is above human capabilities in every respect (i.e. chess computers need not apply).
- As I. J. Good noticed already in the 60’s, this means that a such an intelligence would also surpass humans in designing even better intelligences; therefore, provided that we are able to — directly or indirectly — create even a slightly transhuman intelligence, it seems plausible that available machine intelligence would, at least for a while, increase exponentially.
- It follows that a smarter than human intelligence, if cooperative, is literally the *last* invention mankind need ever make.
- Superintelligences are often categorized into two groups: **weak** and **strong superintelligence** (*not* to be confused with weak and strong artificial intelligence).

Weak superintelligence

“Weak superintelligence is what you would get if you could run a human intellect at an accelerated clock speed, such as by uploading it to a fast computer.”

– Nick Bostrom [3]

- A **weak superintelligence** is an otherwise human level intelligence that just happens to work much faster than normal, whether it’s an overclocked human mind or a basically human-equivalent AI on a fast processor.
- This type of superintelligence can be understood by normal humans, given enough time. Specifically, most fictitious superintelligences tend to be of this type even if they’re purported to be strong, simply because the writer is of human intelligence.
- Though limited in itself, even this type of superintelligent entity would have tremendous impact in a society of biological humans.
- Also, a **weak superintelligence** may be an important stepping stone to **strong superintelligence**, especially if its cognitive structures are more understandable and malleable than those of a human brain.

Strong superintelligence

“Imagine running a dog mind at very high speed. Would a thousand years of doggy living add up to any human insight?”

– Vernor Vinge [1]

- A **strong superintelligence** is *qualitatively* superior to human level intelligence, as humans are arguably intellectually superior to other known animals.
- This type of superintelligence would need a cognitive architecture that allows it to “intuitively” grasp concepts that are merely comprehensible to us, as well as to efficiently think in abstract concepts not directly accessible to human cognition.
- Some differences of this kind are already visible in the spectrum of human intellectual ability, and there’s no reason to assume that the top of the current human range is the final word on this matter.
- It is difficult to predict exactly what a **strong superintelligence** would do, since that would require us to be that smart.
 - Vinge himself has, in his SF work, tended to deal with this by distancing the world’s superintelligences from where the story actually takes place.

Seed AI as a path to the Singularity

“The Singularity is beyond huge, but it can begin with something small. If one smarter-than-human intelligence exists, that mind will find it easier to create still smarter minds.”

– Eliezer Yudkowsky [4]

- An AI with enough general intelligence and knowledge to understand and improve upon itself may be considered a [Seed AI](#).
- It is notable that a [Seed AI](#) need not be originally transhuman, or necessarily even human level, as it may be built to specialize in self-understanding and self-improvement.
- Because of its access to its own source code and greater than human introspection capabilities, self-improvement would presumably be more straightforward for a [Seed AI](#) than a human.
- The problem still is that the original Seed AI needs to be generally intelligent to be of much use in its own development. If this can be achieved, however, this is probably the quickest path to the Singularity.

Other paths to the Singularity

“Computer networks and human-computer interfaces seem more mundane than AI, and yet they could lead to the Singularity.”

– Vernor Vinge [1]

- While AI is often the first thing to be associated with superhuman intelligence, it is not the only way to the Singularity. Other plausible approaches include:
 - **Biotechnological improvement** on the human brain (genetic engineering being an option, albeit a rather slow one).
 - **Computer/human interfaces** (and relevant software) may advance to the point that the user/device combination may be considered superintelligent.
 - **Intimate collaborative technologies** (probably mediated via computer/human interfaces) may enable multiple people (and computers) to effectively function as a superintelligent gestalt mind.
 - **Uploading** a human mind into a computer would yield, given enough computing power, a weak superintelligence, and the improved introspection and self-modification opportunities could enable a transition to the strong variety.

Transhuman ethics

“Success in Friendly AI can have positive consequences that are arbitrarily large, depending on how powerful a Friendly AI is. Failure in Friendly AI has negative consequences that are also arbitrarily large.”

– Eliezer Yudkowsky [5]

- A superintelligent being would presumably be capable of doing both great harm and great good, depending on its goals. This would be especially true to the first such entity, since there would be no equivalent opposition.
- It is plausible that the first such entity to gain power, possibly through developing strong [molecular nanotechnology](#), could then prevent such opposition from ever arising in its sphere of influence.
- It is therefore essential to equip any AI-based superintelligences with foolproof altruistic goal systems (see e.g. Yudkowsky: Creating Friendly AI [5]).
 - Any emergent (e.g. “brute-forced”) AI systems have an increased probability of being unfriendly, lacking proper benevolent goal system design.
- As most humans are to some extent egotistical creatures, creating a more or less human-based superintelligence is somewhat risky also.

Why take the chance?

“... if the technological Singularity can happen, it will.”

– Vernor Vinge [1]

- If it's possible, *somebody* will do it eventually, so it would be prudent to make sure it's somebody paying attention to the risks involved. Relying on the Singularity to be impossible or far away is too large a risk to take.
- A successful Singularity triggered by a Friendly AI [5] or a similarly altruistic being would have such a tremendous positive impact on the living conditions of sentient beings, current and future, that even a minuscule possibility of success makes effort towards it worthwhile.
- A positive Singularity would also be an effective shield against other types of **existential risks** — scenarios that “either annihilate Earth-originating intelligent life or permanently and drastically curtail its potential.” [6]
 - These include e.g. bio- and nanotechnological advancements, which can potentially be destructively harnessed even by small groups and individuals before we've had time to develop sufficient countermeasures on our own.

Time frame

“[Hard SF writers] saw that their most diligent extrapolations resulted in the unknowable ...soon. Once, galactic empires might have seemed a Post-Human domain. Now, sadly, even interplanetary ones are.”

– Vernor Vinge [1]

- Provided that the price of computational power continues to drop as it has up until now, human-level capacity should be attainable by 2024 at the latest, quite possibly much sooner [7]. It is possible, however, that the first human-level AI needs more than that to operate at a reasonable speed, unless an efficient design is found on the first try.
- The software availability is more difficult to predict; however, several projects are even currently aiming towards this. Human-parity hardware isn't necessary to start the basic design and development; some interesting results might be achieved with a proper design even on subhuman hardware.
- If truly intelligent software eludes human developers, the other, slower paths become more relevant. Safety, however, becomes harder to ascertain even theoretically.

Summary

“[The Singularity’s] coming is an inevitable consequence of the humans’ natural competitiveness and the possibilities inherent in technology.”

– Vernor Vinge [1]

- If the Singularity can happen, it will (barring an existential risk scenario); the potential benefits of transhuman intelligence will continue to attract people to tackle the problem.
- Projects to develop transhuman intelligence, or any intelligence capable of self-improvement, should proceed with extreme caution and care, but should nevertheless be undertaken.
- While the Singularity is an interesting subject of speculation in both SF literature and futurism, great care must be taken in handling the subject in any realistic manner. Pre-Singularity speculation, right up to the event itself, is comparatively easy; post-Singularity speculation should be taken at a proof-of-concept level only.

Bookshelf

- Vernor Vinge's *A Fire Upon the Deep*, *True Names* and *Marooned in Realtime* show the Singularity to the reader up close, without actually crossing into the unknown territory.
- Greg Egan's *Permutation City*, *Quarantine* and *Diaspora* help the reader visualize several aspects of mind uploading, alteration and improved introspection, among other things. Egan also has several short stories available on his web pages at <http://www.netSPACE.net.au/%7Egregegan/>.
- Iain M. Banks' *Consider Phlebas*, *Use of Weapons*, *Player of Games* and other Culture novels present a utopian picture of a society overseen but not exactly ruled by (apparently weakly) superintelligent Minds.

References

- [1] Vinge, Vernor, *What is the Singularity?*
<URL:<http://www.ugcs.caltech.edu/%7Ephoenix/vinge/vinge-sing.html>> (July 26th, 2003)
- [2] Good, I.J., *Speculations Concerning the First Ultrainelligent Machine*. Advances in Computers, vol 6, Franz L. Alt and Morris Rubinoff, eds, 31-88, 1965.
- [3] Bostrom, Nick, *The Transhumanist FAQ*.
<URL:<http://www.transhumanism.org/resources/faq.html>> (July 26th, 2003)
World Transhumanist Association
- [4] Yudkowsky, Eliezer, *What is the Singularity*.
<URL:<http://singinst.org/what-singularity.html>> (July 26th, 2003)
Singularity Institute for Artificial Intelligence
- [5] Yudkowsky, Eliezer, *Creating Friendly AI*.
<URL:<http://singinst.org/CFAI/>> (July 26th, 2003)
Singularity Institute for Artificial Intelligence
- [6] Bostrom, Nick, *Existential Risks — Analyzing Human Extinction Scenarios and Related Hazards*
<URL:<http://www.nickbostrom.com/existential/risks.html>> (July 26th, 2003)
Journal of Evolution and Technology, 2002, vol. 9
- [7] Bostrom, Nick, *How long before Superintelligence?*, 1998
<URL:<http://www.nickbostrom.com/superintelligence.html>> (July 26th, 2003)
International Journal of Future Studies, 1998, vol. 2