

Ethical Perspectives on the Singularity

Victor LaBrie

What is the Singularity?

General definition

- According to Merriam-Webster:

singularity noun

 Save Word

sin·gu·lar·i·ty | \,sɪŋ-ɡyə-'ler-ə-tē , -'la-rə-\

plural **singularities**

Definition of *singularity*

1 : something that is singular: such as

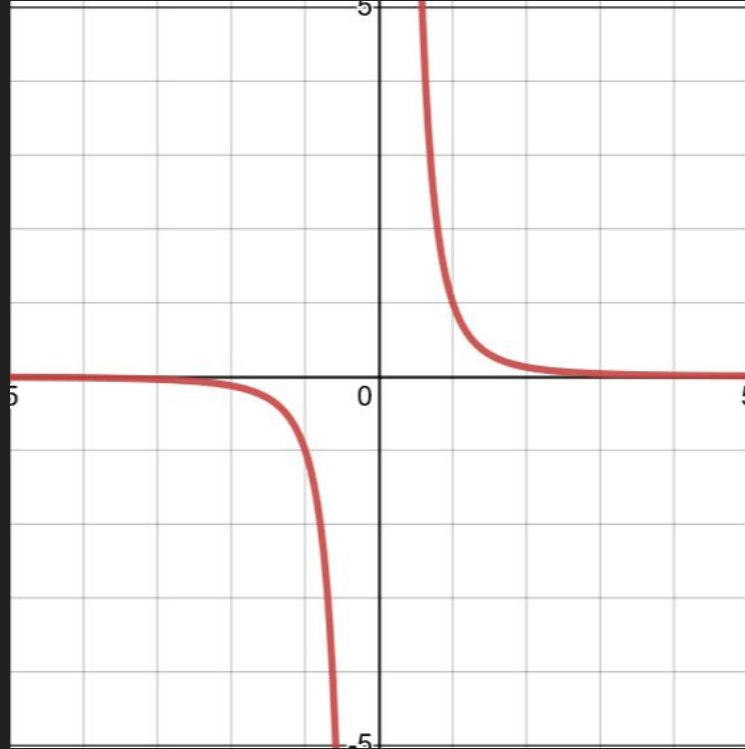
a : a separate unit

b : unusual or distinctive manner or behavior : PECULIARITY

2 : the quality or state of being singular

Singularities in Math

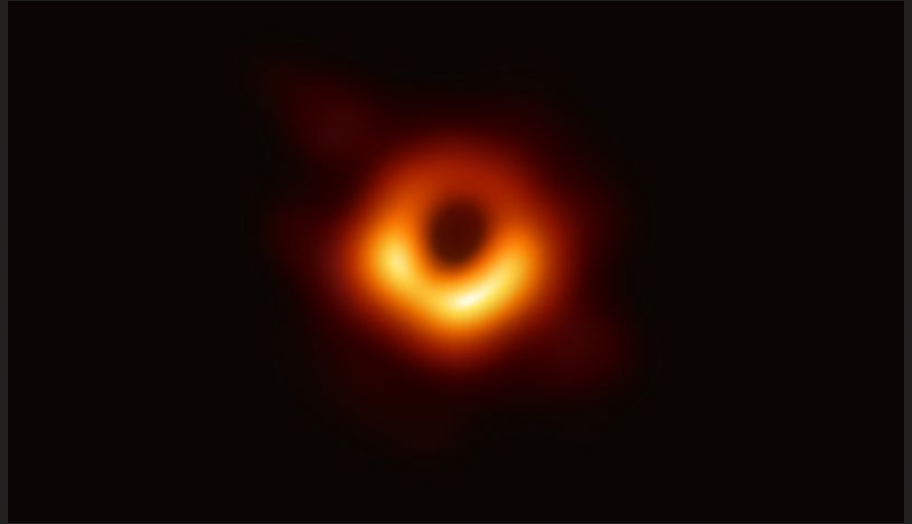
- Point at which a function “blows up”*
 - Undefined at this point, while everything around it is defined
- $y=x^{-3}$
- The singularity is (0,0)



* Wolfram Alpha

Singularities in Physics

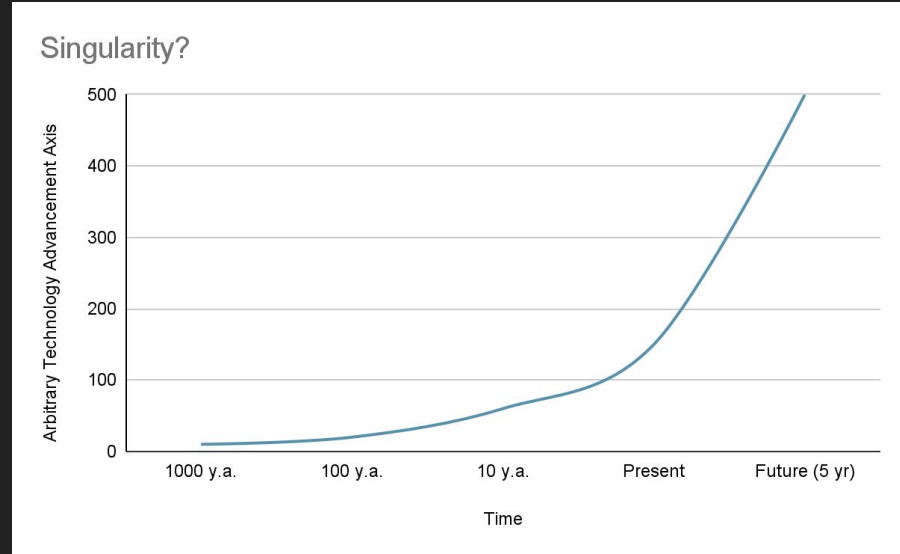
- Most widely known use of the word
- Gravitational singularities (black holes)
- Event horizon
- Physics breaks down



From NASA's website.

The Technological Singularity

- Future point beyond which reasonable predictions can't be made
- Coined by John von Neumann²
- Popularized later
- Vernor Vinge in 1993
- Ray Kurzweil really popularized it later
- Vinge argued that they were “on the edge of change comparable to the rise of human life on Earth.”⁴





<https://youtu.be/gpKNAHz0zH8>

What's going to cause it?

- According to Vinge⁴:
 - Conscious supercomputer
 - Conscious computer network
 - Brain computer interfaces
 - Advance in biotech/medicine
- “Intelligence Explosion”¹



<https://youtu.be/rsCul1sp4hQ>

Intelligence Explosion

- Super intelligence created
- Improves itself
- Repeat
- Greater intelligence = faster technological advancement rate (probably)
- This will likely start happening as soon as it becomes cheaper to have AI code rather than having coders code¹
 - GitHub Copilot

Will this actually happen?

- We don't know
- When AGI is achieved, singularity will happen
- Predictions for that range from 10 years to never⁵

3.) The earliest that machines will be able to simulate learning and every other aspect of human intelligence:		
Within 10 years	6	5%
Between 11 and 25 years	3	2%
Between 26 and 50 years	14	11%
More than 50 years	50	41%
Never	50	41%
Totals	123	100%

What happens after?

- There are many ideas of what could happen after
- Ray Kurzweil thinks that intelligence and consciousness will permeate the cosmos (somehow)³
- Could be terrible (I have no mouth and I must scream)
- Truthfully though we don't know

Let's get to the ethics

What questions should we be asking?

- Possible scenarios?
- Those are separate topics that other people will be exploring
- Not central to the idea
- The main idea is that we don't know what will happen
 - Event Horizon

What should we think about then?

Overarching Questions

- Is it good?
 - Unknown
- How should we act in the face of the unknown?
 - Allow or don't
- If we don't allow this to happen, what actions can we take?

How should we react?

Utilitarian Perspective

Act Utilitarianism

- Defining utility
- Moral luck

Rule Utilitarianism

- Analogy: Should you push a button that will either kill you or give you \$1 billion?
- Value of risk
- If the risk is high enough, does the reward matter?
- Singular

Kantian Perspective

Good will

- Varies from person to person

Categorical Imperative

- Universalization
 - Same situation as with Rule utilitarianism
- Treating people as ends unto themselves
 - Deprives choice

Overall it seems we should not allow this.
What should we do?

A new social contract?

- Amend the social contract
- “Don’t take actions that could lead to AI takeover”
- Should other tech not be pursued?
- “Better us than them”

Sources

- 1) Tegmark, Max. *Life 3.0*. Random House US, 2018.
- 2) Shanahan, Murray. *The Technological Singularity*. The MIT Press, 2015.
- 3) Kurzweil, Ray. *The Singularity Is near: When Humans Transcend Biology*. Duckworth, 2018.
- 4) Vinge, Vernor. "Technological Singularity." *Technological Singularity by Vernor Vinge*, <https://frc.ri.cmu.edu/~hpm/book98/com.ch1/vinge.singularity.html>. Originally published 1993.
- 5) Müller, Vincent C., and Nick Bostrom. "Future Progress in Artificial Intelligence: A Survey of ..." *Nickbostrom.com*, Future of Humanity Institute, <https://nickbostrom.com/papers/survey.pdf>.