An Overview of The Technological Singularity

Marcelo Fernandes
Software Engineer
Wellington, New Zealand
Email: marceelofernandes@gmail.com

Abstract—This paper briefly explores the history of the term "Technological Singularity" (TS) and analyses various hypotheses about how and when the TS is most likely to occur. Some of the most prominent researchers' definitions are reviewed in terms of how realistic they are, given today's technological trends and the evidence available to support their claims. Through identifying examples of accelerating changes in technology, the likelihood of these trends resulting in the TS is discussed. The paper then critically evaluates the broad potential socio-economic impacts of current and future accelerating technological changes. The paper concludes that the reviewed literature does not present sufficient evidence to identify what the TS definitively is or when it is most likely to occur. It instead calls for the creation of a regulatory framework that works to mitigate the most threatening impacts of future technologies on our economy, human rights, and society as we know it.

August 24, 2020

I. Introduction

The expression "technological singularity" (TS) is an umbrella term used to describe potential technological transformations that would lead to breakthrough challenges for human life. The term was popularised by V. Vinge [1] in 1993, when he explained that "the singularity" would be the "point where our models must be discarded and a new reality would rule", and he further associated the term with human extinction. The concepts around the term have earlier appearances. Back in 1958, J. Neumann was already speculating about the "accelerating progress" of technology that would culminate in what he described as an end of human affairs [2], and in 1966 I. Good [3] built on the theory adding that an ultra-intelligent machine would be the last that humankind would ever build.

The term became broadly used and acquired a diversity of meanings, despite the links that existed between early definitions (i.e. the existence of a machine in relation to the end of human life). Some definitions in the literature emphasise artificial intelligence as the main trigger for the TS, whereas others coined the term referring to a more general technological invention. Those ideas may or may not come together in a single definition, which has led to confusion and, ultimately, misunderstanding. A more thorough taxonomical analysis found that there are nine main definitions of the TS [4]. Those nine definitions can be encapsulated in two concepts: *intelligence explosion*, and *accelerating change*.

- *Intelligence Explosion (IE)*: Upon the creation of an artificial general intelligence (AGI), i.e, an artificial intelligence that has the capacity to solve any problem that humans can, the AGI would then rapidly trigger many self-improvement interactions by rewriting its own software. This would rapidly culminate in an artificial super-intelligence (ASI), which would surpass humanity's knowledge to unknown limits. This concept is extrapolated from a speculative scenario created by I. Good (1965) [3].
- *Accelerating Change (AC)*: Derived from the conversation between J. Neumann and S. Ulam [2], and supported by the "futurist" R. Kurzweil [6], the principle of this concept is that technology follows an exponential pattern. As soon as new technology is released, subsequent technology will use previous ones to be created faster. R. Kurzweil predicts that the singularity caused by the rapid advance of technological ruptures will occur around 2045 [6].

II. Mediating between conflicting opinions

The TS hypotheses sound fascinating, but how realistic are they? With many different theories supporting the high probability of a TS, it seems almost inevitable. However, there are many who would disagree. A research paper conducted in 2018 [7] asked machine learning specialists about the likelihood of an intelligence explosion occurring. The results showed that 50% believed it was "unlikely" or "quite unlikely", 21% believed it was "about even", and 29% believed it was "likely" or "very likely" to occur. The positions expressed by these specialists show that the research community is at least somewhat divided about the likelihood of an IE occurring.

The AC hypothesis has also been reviewed critically. J. Huebner [8] made a thorough examination of the course of technology and innovation between the Dark Ages and the early 2000's. His examination concluded that the innovation rate had long started to decrease, with its peak in around 1900. This coincided with a decrease in the number of patents per thousand, showing that human creativity was not accelerating. The decrease had been attributed to an economic and human brain limit. The primary claim was that it does not matter whether there are large amounts of information if there are no means to process it, or if there is no economic resource and interest to do so. J. Schmidhuber also attributed a "feeling" of accelerating change to humans tending to give special

importance to recent events, rather than analysing a full picture of time [9]. In addition, the fact that an AGI has the ability to upgrade its own code does not necessarily imply it will have the directive to do so. If the AGI does update itself, it may or may not enhance its intelligence.

In contrast, other theories are in support of these hypotheses. Moore's law states that transistors in integrated circuits double in number biannually (approximately), and this law has come to prove itself over time [10,11]. More transistors means more computational power, the flow-on effect of which is increasingly fast computing units. Whether this trend will continue into the future is up for discussion. A physical limit to the speed and amount of transistors that can be placed in a circuit may be a difficult barrier to overcome, culminating in an abrupt halt to computing hardware progress [12,13]. Alternatively, an accelerated advance could happen due to increases in software performance [14-15], but those again, present their own limitations [16].

There are arguments on both sides, but the literature does not seem to provide a consolidated conclusion about if the TS will happen, when it will happen, and how it will happen [17]. It is important to remember that futurologists and scientists have been wrong in the past: "We can never understand the contents of stars" [18], "heavier-than-air machines will never be able to fly" [19], "There's no indication that nuclear energy will be ever obtained" [20]. These are just a few examples of universal generalizations that only took a single counter-example to become completely invalidated. If scientists can be wrong, shouldn't we, at least, prepare for the TS? In order to answer this question, an analysis of the TS' impact on humanity is required.

III. Social Impact

Concerns about jobs being taken by machines have gained prominence in recent years [21]. E. Brynjolfsson (MIT) claims that the acceleration of technology and automation is likely to result in economic disruption; noting there has never been a better time to be a high-skilled worker, or a worse time to be a low-skilled worker [22]. With more advanced computing power, low skilled employees are increasingly being replaced by machines. Considering that E. Brynjolfsson is not taking into consideration the existence of an AGI, these impacts are just the tip of the iceberg.

The likely impacts of an AGI are considerably more significant. There is no reason to believe that an AGI will necessarily behave in favor of humankind [23]; on the contrary, an AGI could seek self-preservation at human expense, potentially working to terminate human affairs as a way of protecting scarce resources [24]. If an IE does not cause human extinction, it may still bring harm to society by following three additional universal principles of self-improved systems that compete with current technology and systems: Creativity for adaptation, acquisition of resources, and efficiency enhancements [25].

While many potential scenarios are somewhat apocalyptic, there are more positive theories that discuss the potential for human enhancement through AC. As bio-engineering technology progresses, so does human life expectancy. If technology one day goes as far as allowing people to "replace" old components of their body, natural aging would no longer be a cause of death [26, 27]. Looking further, virtual immortality may be possible through digital mapping of the human brain that could be uploaded to a computer [28-31]. Such scenarios would significantly impact society as we know it but seem far from being actualised. More tangible goals can be achieved in some specific areas with a controlled AI, e.g: assistants for people with mild cognitive impairments, monitoring and looking after patients with COVID-19, reducing or completely eliminating road crashes with automated vehicles [32-34].

Having examined a broad range of advantages and disadvantages of various TS scenarios, it seems only reasonable to suggest that humanity needs to establish policies that mitigate the potential for catastrophic collaterals of a TS. AC has already led to forms of oppression, especially on those most powerless in society infringing on their established human rights [35]. Humanity should be vigilant and act to create regulation that maintains safe AI development.

IV. Conclusion part 1. What is the TS?

As this essay has discussed, the TS is a hypothetical point in time in which technology will become so advanced that society will be faced with a complete paradigm shift. Due to a lack of strong evidence it is difficult to be more specific about the ways the TS might play out. The evidence and predictions presented here do not provide sufficient information to make more detailed claims. Given this uncertainty and limited understanding, it would be unhelpful to give a restrictive definition of what the TS is, or a prediction of when the TS would happen, and how it would happen as this could limit our ability to provide an adequate response.

There is a useful taxonomy that can be extracted from this discussion. "Accelerating Change" is a well evidenced term when not directly associated with the TS; it is fairly acceptable to say that technology is accelerating in general, and that society is already being affected by this dynamic, however It would be speculative to say that this acceleration will necessarily lead to a singularity. The other terms (self improving technology, emergence of super-intelligence, complexity disaster, infinite progress, etc) [4] that are also portrayed in the literature are useful placeholders for a more philosophical discussion. Anyone who claims to have a more clear prediction of how the TS might happen fails to acknowledge the multitude of possible outcomes and the lack of substantive evidence to support any one of them.

V. Conclusion part 2. Reflection

Even though a future with AGI is apparently not so close, the present is already being affected by accelerating changes in technology. The imminent challenges now happening (e.g. job losses due to technological innovations and AI [23, 36]) are examples that illustrate the urgency built by the topic. Society

needs to initiate a debate about the creation of a social framework, robust enough to prevent future-threatening technology from being used against humanity. Who should be delegated with this responsibility? A possible immediate answer would be the government, however governments are barely legislating fast enough to keep up with current technologies that are not nearly as threatening as a possible TS [37,38]. American laws are still dealing with technology related acts and codes from the previous millennium [39,40]. Should the TS occur in 2045 as R. Kurzweiil predicts [6], societies and governments are lagging far behind.

Three key questions about the TS should be prioritised by importance accordingly:

1) How will society respond if the TS happens?
2) When will the TS happen?
3) What, precisely, is the TS?

Question 3 is of lowest priority as its answer is insufficient when it comes to preventing and actioning a response to the TS should it occur. The answer to question 2 would be as important as the answer to question 1, except that as it has been shown, determining when the TS will happen is with today's understanding too difficult. Question 1 should be more strongly pursued as it is the most tangible and actionable of the three, and is therefore most important given that regardless of the answers to 2 and 3, humanity must prepare for the TS.

References

[1] V. Vernor, "The coming technological singularity: How to survive in the post-human era", in *Vision 21: Interdisciplinary Science and Engineering in the Era of Cyberspace*, 1993, pp. 11-22.

[2] U. Stanislaw, "Tribute to John Von Neumann, 1903-1957", in *Bulletin of the American Mathematical Society*, 1958, p. 5.

[3] I. Good, "Speculations Concerning the First Ultraintelligent Machine", *Advances in Computers*, pp. 31-88, 1966. Available: 10.1016/s0065-2458(08)60418-0 [Accessed 8 August 2020].

[4] A. Sandberg, "An Overview of Models of Technological Singularity", *The Transhumanist Reader*, pp. 376-394, 2013. Available: 10.1002/9781118555927.ch36 [Accessed 8 August 2020].

[5] M. Buchanan, "The law of accelerating returns", *Nature Physics*, vol. 4, no. 7, pp. 507-507, 2008. Available: 10.1038/nphys1010 [Accessed 9 August 2020].

[6] R. Kurzweil, *The Singularity is near*. London: Duckworth, 2006.

[7] K. Grace, J. Salvatier, A. Dafoe, B. Zhang and O. Evans, "Viewpoint: When Will AI Exceed Human Performance? Evidence from AI Experts", *Journal of Artificial Intelligence Research*, vol. 62, pp. 729-754, 2018. Available: 10.1613/jair.1.11222.

[8] J. Huebner, "A possible declining trend for worldwide innovation", *Technological Forecasting and Social Change*, vol. 72, no. 8, pp. 980-986, 2005. Available: 10.1016/j.techfore.2005.01.003 [Accessed 12 August 2020].

[9] J. Schmidhuber, "New Millennium AI and the Convergence of History", *Challenges for Computational Intelligence*, pp. 15-35, 2007. Available: 10.1007/978-3-540-71984-7_2 [Accessed 12 August 2020].

[10] E. Mollick, "Establishing Moore's Law", *IEEE Annals of the History of Computing*, vol. 28, no. 3, pp. 62-75, 2006. Available: 10.1109/mahc.2006.45 [Accessed 12 August 2020].

[11] C. Mack, "Fifty Years of Moore's Law", *IEEE Transactions on Semiconductor Manufacturing*, vol. 24, no. 2, pp. 202-207, 2011. Available: 10.1109/tsm.2010.2096437 [Accessed 20 August 2020].

[12] G. Cerofolini, "Realistic limits to computation I. Physical limits", *Applied Physics A*, vol. 86, no. 1, pp. 23-29, 2006. Available: 10.1007/s00339-006-3670-5.

[13] I. Tuomi, "The Lives and Death of Moore's Law", *First Monday*, vol. 7, no. 11, 2002. Available: 10.5210/fm.v7i11.1000 [Accessed 20 August 2020].

[14] M. Reformat, "Special section: Applications of computational intelligence and machine learning to software engineering", *Information Sciences*, vol. 259, pp. 393-395, 2014. Available: 10.1016/j.ins.2013.11.019.

[15] M. Wang, Y. Cui, X. Wang, S. Xiao and J. Jiang, "Machine Learning for Networking: Workflow, Advances and Opportunities", *IEEE Network*, vol. 32, no. 2, pp. 92-99, 2018. Available: 10.1109/mnet.2017.1700200 [Accessed 12 August 2020].

[16] "Tech Luminaries Address Singularity", *IEEE Spectrum*, 2008. [Online]. Available: https://spectrum.ieee.org/computing/hardware/tech-luminaries-address-singularity. [Accessed: 13 Aug 2020].

[17] D. Bostrom, "Existential Risks - Analyzing Human Extinction Scenarios and Related Hazards", *Journal of Evolution and Technology*, vol. 9, 2001. [Accessed 13 August 2020].

[18] I. Comte, *The Positive Philosophy*. Paris, France, 1842.

[19] L. Kelvin, "KELVIN ON SCIENCE. — British Lord Tells His Hopes For Wireless Telegraphy", New York Journal, 1902.

[20] A. Einstein, "Atom Energy Hope is Spiked By Einstein / Efforts at Loosing Vast Force is Called Fruitless,", *Pittsburgh Post-Gazette*, 1934.

[21] E. Brynjolfsson and A. McAfee, *The second machine age*. New York: Norton, 2016.

[22] N. Bostrom, "Ethical Issues in Advanced Artificial Intelligence", *Cognitive, Emotive and Ethical Aspects of Decision Making in Humans and in Artificial Intelligence*, vol. 2, pp. 12-17, 2003. [Accessed 13 August 2020].

[23] "The Risk of Automation for Jobs in OECD Countries", *OECD Social, Employment and Migration Working Papers*, 2016. Available: 10.1787/5jlz9h56dvq7-en [Accessed 20 August 2020].

[24] P. Wang, B. Goertzel and S. Franklin, *Artificial general intelligence, 2008*. Amsterdam: IOS Press, 2008.

[25] S. Omohundro, "The Nature of Self-Improving Artificial Intelligence", *Singularity Summit 2007*, 2007.

[26] T. Rando and H. Chang, "Aging, Rejuvenation, and Epigenetic Reprogramming: Resetting the Aging Clock", *Cell*, vol. 148, no. 1-2, pp. 46-57, 2012. Available: 10.1016/j.cell.2012.01.003 [Accessed 14 August 2020].

[27] A. Johnson, K. Akman, S. Calimport, D. Wuttke, A. Stolzing and J. de Magalhães, "The Role of DNA Methylation in Aging, Rejuvenation, and Age-Related Disease", *Rejuvenation Research*, vol. 15, no. 5, pp. 483-494, 2012. Available: 10.1089/rej.2012.1324 [Accessed 14 August 2020].

[28] S. BAMFORD, "A FRAMEWORK FOR APPROACHES TO TRANSFER OF A MIND'S SUBSTRATE", *International Journal of Machine Consciousness*, vol. 04, no. 01, pp. 23-34, 2012. Available: 10.1142/s1793843012400021 [Accessed 14 August 2020].

[29] S. BAMFORD, "A FRAMEWORK FOR APPROACHES TO TRANSFER OF A MIND'S SUBSTRATE", *International Journal of Machine Consciousness*, vol. 04, no. 01, pp. 23-34, 2012. Available: 10.1142/s1793843012400021 [Accessed 14 August 2020].

[30] K. SOTALA and H. VALPOLA, "COALESCING MINDS: BRAIN UPLOADING-RELATED GROUP MIND SCENARIOS", *International Journal of Machine Consciousness*, vol. 04, no. 01, pp. 293-312, 2012. Available: 10.1142/s1793843012400173 [Accessed 14 August 2020].

[31] G. Bell and J. Gray, "Digital immortality", *Communications of the ACM*, vol. 44, no. 3, pp. 28-30, 2001. Available: 10.1145/365181.365182 [Accessed 20 August 2020].

[32] A. Nesarikar, W. Haque, S. Vuppala and A. Nesarikar, "COVID-19 Remote Patient Monitoring: Social Impact of AI", 2020. [Accessed 15 August 2020].

[33] S. Demmel et al., "Global risk assessment in an autonomous driving context: Impact on both the car and the driver", *IFAC-PapersOnLine*, vol. 51, no. 34, pp. 390-395, 2019. Available: 10.1016/j.ifacol.2019.01.009 [Accessed 15 August 2020].

[34] J. Pagan, M. Mokhtari, B. Abdulrazak and M. Cabrera, "How AI Impacts Urban Living and Public Health", in *17th International Conference*, ICOST 2019, New York City, 2019.

[35] L. Andersen, "HUMAN RIGHTS IN THE AGE OF ARTIFICIAL INTELLIGENCE", *Access Now*, 2018.

[36] C. Frey and M. Osborne, "The future of employment: How susceptible are jobs to computerisation?", *Technological Forecast-

ing and Social Change, vol. 114, pp. 254-280, 2017. Available: 10.1016/j.techfore.2016.08.019 [Accessed 17 August 2020].

[37] B. Milano, "Government can't keep up with technology's growth", *Harvard Gazette*, 2019. [Online]. Available: https://news.harvard.edu/gazette/story/2019/02/government-cant-keep-up-with-technologys-growth/. [Accessed: 17- Aug- 2020].

[38] B. Schneier, "Policymaking must catch up with technology - before it's too late", *World Economic Forum*, 2019. [Online]. Available: https://www.weforum.org/agenda/2019/11/we-must-bridge-the-gap-between-technology-and-policy-our-future-depends-on-it/. [Accessed: 17-Aug- 2020].

[39] "THE DIGITAL MILLENNIUM COPYRIGHT ACT OF 1998". Washington DC, 1998. [Online]. Available: https://www.copyright.gov/legislation/dmca.pdf. [Accessed: 17- Aug-2020].

[40] "18 U.S. Code § 1030 - Fraud and related activity in connection with computers", *LII / Legal Information Institute*, 1984. [Online]. Available: https://www.law.cornell.edu/uscode/text/18/1030. [Accessed: 19- Aug-2020].