THE TRAJECTORY TO THE "TECHNOLOGICAL SINGULARITY"

Casey Burkhardt
Department of Computing Sciences
Villanova University
Villanova, Pennsylvania 19085

Abstract

The idea of the technological singularity – the moment at which intelligence embedded in silicon surpasses human intelligence – is a matter of great interest and fascination. It is also an idea that rests on several problematic assumptions about the nature of human intelligence. This paper discusses the major proposals, originating mainly in the artificial intelligence community, concerning the nature of the technological singularity and the stages of progress toward the event itself. Attention is given to the problems raised by the concept of the singularity, the controversy that has surrounded the charting of milestones on the path to its realization, and the importance of responsible disclosure of technological progress and predictions.

1 Introduction

Learning about advances in technology is a commonplace activity in the daily lives of most people, whether or not they are directly involved in the field of computing. Is it possible that in the not-so-distant future one of these advances will yield a machine capable of matching the computational ability of the human brain? This paper investigates such a phenomenon, known as the technological singularity, by first defining the event, its implications, and the theoretical and computing challenges associated with its achievement. Detail is then given to several published predictions regarding timing for the singularity, and these views are reconciled into a more reasonable timeframe estimate. Some views opposing the possibility of such an occurrence are presented, and the paper concludes by addressing the social importance of responsible disclosure of claims and advancements in this area of research.

2 The Technological Singularity and Its Implications

The technological singularity is best defined as an event or point in time when a combination of computer hardware and artificial intelligence algorithms match or exceed the computational ability of the human brain. This may come about in a number of ways, including through the creation of "computers that are 'awake' and superhumanly intelligent, large computer networks (and their associated users) [that] may 'wake up' as a superhumanly intelligent entity, or computer/human interfaces [that] may become so intimate that users may reasonably be considered superhumanly intelligent." [Vinge, 1993] Essentially, the singularity will result in the creation of superintelligence, which is "an intellect that is much smarter than the best human brains in practically every field, including scientific creativity, general wisdom and social skills … [and] is conscious and has subjective experiences." [Bostrom, 1998]

To many, such an event is inconceivable, but should it occur, the implications of the technological singularity are even less clear. Because human thought will have been superseded by non-biological computation, humans will be unable to predict with any degree of certainty how technology will advance beyond the singularity. This is because most significant advances after the singularity will be made by superintelligence, not human intelligence, and will therefore be incomprehensible to any individual person. The concept of "recursive self-improvement" would also be in effect, and permit technology to improve upon its own level of intelligence at a perpetually accelerating pace.

3 Difficulties in Pinpointing the Singularity and Its Milestones

Knowing that the singularity occurs when technological computational ability exceeds the capabilities of the human brain, it is clear that the first difficulty in identifying the event is defining humanity's capabilities. This is not a trivial task, and one that may not presently have a precise answer. Moravec estimates intelligence in terms of raw computational power based on the relative size of the brain to the retina, but this approach is very one-sided and ignores the possibility that there is no single metric for human intelligence, or that collective human intelligence can be measured, let alone reproduced mechanically. It is important to acknowledge this limitation in pinpointing the singularity, but for the purposes of tracking the trajectory to the event, we will assume this is an attainable goal.

One of the most fascinating discussions surrounding the technological singularity is that of the predicted timing of its occurrence. It is important to understand that there will be no immediately measureable and instant event, but rather a more subtle increase in computational ability during an initial period of time. Many brilliant people fail to reach a consensus as to a time frame simply because each considers and weighs separate factors when making timing predictions. Advances in both hardware and software must be coordinated in a manner that allows artificial intelligence and technology to supersede human intellect. Because of its momentous scale, there is no concrete framework for tracking the progress of such a task.

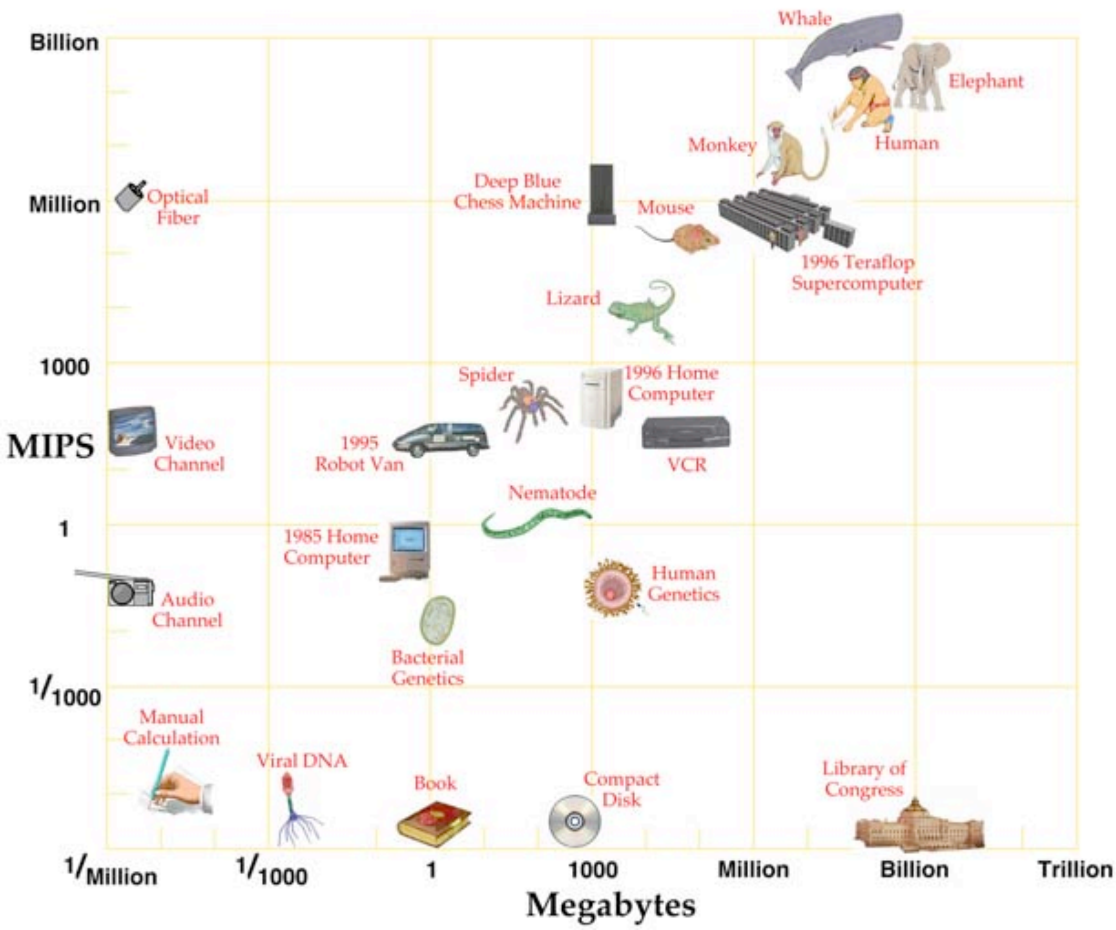
Ray Kurzweil provides some insight into a mechanism that can possibly be used to generate more accurate timing predictions for the technological singularity. He uses a methodology by which he estimates the date ranges of less significant yet technologically relevant events. These estimates can be seen as markers for plotting a trajectory to a larger, encompassing, event – in this case, the singularity. Although more useful and accurate than making a broad prediction from a smaller number of signals, this approach has its problems. For one, the marker events may not have a direct or equal correlation to underlying advancements leading to the singularity. Also, the time periods predicted for the marker events are not evenly distributed over a broad time scale, so developing a consistent trajectory is difficult, since sequential predicted events may be several years or more apart.

3.1 The Hardware Problem

One of the most challenging problems faced in reaching the singularity is the development of hardware capable of producing the raw computational power of the brain. It is believed that for a computer to imitate the complexities of the human mind, it would need to be capable of performing at a rate of 100 million MIPS (million instructions per second) or 10^{14} instructions per second. [Moravec, 1998] For the sake of comparison, Deep Blue, the computer that defeated chess grandmaster Garry Kasparov, operated at a rate of only 3 million MIPS. A single high-end multi-core consumer processor released in 2011 recently attained a throughput of 159,000 MIPS, still only $1/629^{th}$ the power estimated for matching the computational ability of the human brain. [Angelini, 2011] It may be some time before the required specifications can be reached and become publicly available.

Moore's law indicates that significant returns in processor speed increases are expected due to the ability to double the number of integrated circuits on a silicon chip every 18 months. Assuming this holds true, it is plausible that we may have commercially available hardware capable of reaching 100 million MIPS by the year 2025. However, in recent years, Moore's law has come under a great deal of scrutiny, and it is widely believed that this trend of accelerating returns cannot continue in its current form. In 2005, Gordon Moore himself stated that, "in terms of size [of transistors] you can see that we're approaching the size of atoms which is a fundamental barrier … We have another 10 to 20 years before we reach a fundamental limit." [Dubash, 2005] Others, including Ray Kurzweil, believe that Moore's law will continue in another form, perhaps generating processor power increases through the use of "three-dimensional silicon chips, optical computing, crystalline computing, DNA computing, and quantum computing, [that will] keep the law of accelerating returns as applied to computation going for a long time." [Kurzweil, 2001]

Of course, sheer computational power is not the only hardware problem related to the technological singularity. There is also the issue of memory and storage -- a task at which the human brain excels. According to Moravec, this issue can be defined most clearly by determining a "time constant", or "roughly how long it takes the computer to run once through its memory." He claims that there is a one-to-one ratio between the number of calculations capable in a single second to the number of bytes of memory in an optimal system. This ratio extends to the human brain and nervous system, so "the 100-trillion-synapse brain would hold the equivalent [of] 100 million megabytes", or approximately 100 terabytes of data. [Moravec, 1998] In order to meet this capacity, fast-access storage mechanisms must be engineered 2 to 3 orders of magnitude greater than we have today. The following diagram from Hans Moravec's discussion exemplifies the linear relationship between the computations and storage requirements over a logarithmic scale.



[Moravec, 1998]

3.2 The Software Problem

Of course, the issue of hardware and physical computing potential is not the only problem in achieving the technological singularity. The field of artificial general intelligence must develop in a way that allows for the construction of computer software capable of superintelligence. Such advances are certainly nontrivial and require a multitude of research and academic discovery, which are likely to take many years. Most researchers in the artificial intelligence community believe that software is the larger half of the singularity puzzle, and as such, “it's … likely that devising the software will be a tricky process, involving lots of false starts and experimentation.” [Vinge, 1993]

Although a potentially correct path to developing these artificial intelligence strategies and software is currently unknown, there are some indications that taking a bottom-up approach, which is one that relies less on human input, to programming superintelligence may yield promising results. [Bostrom, 1998] In order for this methodology to work, there must be a reasonable model of the brain’s working architecture. Interestingly enough, some progress has been made in this effort by reverse engineering brain activity into a functional model. [Kurzweil, 2001 (124)] A complete software-based functional equivalent of the human brain, combined with a bottom-up learning algorithm and sensory input may be the most realistic strategy in the development of a rudimentary prototype of superintelligence.

The scale of such an undertaking would certainly be a significant limiting factor to the timely development of a software solution. It is safe to assume that a project of this nature would consist of many cross-disciplinary teams with limited understanding of the other’s roles and implementations. As such, it is plausible that the initial effort of organizing the project teams at scale would take a considerable amount of time. Finally, the challenge of close integration between hardware and software presents a significant problem. Traditionally, large software projects do not optimize well for a targeted platform, and there is often limited communication between those designing the hardware and the software that will run on it. For such an effort to produce a successful result, it would be of upmost importance for these teams to integrate closely and work together to develop a solution.

4 Several Optimistic Predictions

This section will present the perspectives and the definitive predictions of four subject matter experts regarding the timing of the technological singularity. Special attention will be given to their individual methodologies for determining a potential time period for the event, highlighting the strengths and weaknesses of each. The desired result of these case studies is to attain a better understanding of the predictive models each expert used, and to reconcile these perspectives into a more definitive and potentially more accurate timing prediction.

4.1 Vernor Vinge

Former mathematics and computer science professor at San Diego State University, Vernor Vinge, is responsible for first creating and popularizing the term “technological singularity” in his 1993 essay, “The Coming Technological Singularity: How to Survive in the Post-Human Era.” The abstract of this work begins with a rather staggering prediction – “Within thirty years, we will have the technological means to create superhuman intelligence. Shortly after, the human era will be ended.” [Vinge, 1993] Vinge emphasizes the fact that superintelligent entities will be able to recursively self-improve themselves at an incomprehensible rate. Because of this, “from the human point of view this change will be a throwing away of all the previous rules, perhaps in the blink of an eye, an exponential runaway beyond any hope of control.” He asserts that at some point between the years of 2005 and 2030, technological advances will lead to the creation of superintelligence, causing human existence as we know it to cease. Vinge bases this estimate solely on the current trend in hardware performance growth, with a slight margin for error.

The prediction Vernor Vinge sets forth is accurate in regard to what it accounts for – the hardware problem, but fails to quantify time spent researching artificial general intelligence or developing computer software. As mentioned earlier, the software problem is a significant challenge standing in the way of the occurrence of the singularity, and if ignored in timing predictions, may lead to inaccuracies. Vinge takes note of this briefly, but only to recognize the problem's nontrivial nature, not to consider it explicitly in his prediction. In this case, I believe that the neglect of accounting for the software problem results in a significant underestimate in the amount of time required to discover superintelligence, somewhat discrediting the accuracy of his estimate.

4.2 Nick Bostrom

Nick Bostrom, the director of The Future of Humanity Institute at Oxford University, has authored a number of works regarding the impact and risks of the technological singularity. In his 1998 essay, "How Long Before Superintelligence?" Bostrom provides a very thorough prediction regarding the timing of the event. He estimates that within the first quarter of this century, hardware capable of matching the computational ability of human intelligence will be available. Furthermore, he makes reference to biological research that is attempting to understand the structure and workings of the brain, and emphasizes that he believes enough will be known to emulate it in software by the year 2012. His essay was updated with a total of four postscripts that further revise his predictions over time. In his latest addendum, he states that due to the rate of advancements in software, and revised hardware requirements, he would, "assign less than a 50% probability to superintelligence being developed by 2033." [Bostrom, 1998]

Bostrom provides great detail in his essay, not only about the need for raw computational ability, but also about the requirements of the artificial intelligence software that can potentially lead to the technological singularity. He provides an analysis of Moore's law, defending its potential validity through the next 10 to 15 years, and determines the computational ability of the human brain within an order of magnitude of previous estimates. Most notably, he introduces the bottom-up approach for constructing machine learning algorithms for use in superintelligence, citing several examples of algorithms that are likely and unlikely to yield results in the context of reaching the technological singularity. He also clearly identifies the need for an in depth understanding of the brain's structure and sensory input so as to be able to emulate related processes in software. Although Bostrom writes in-depth about the software problem, and the necessary biological research needed to emulate the human brain, he ends by underestimating the amount of time to completion for these research and software engineering tasks. This can be seen in the series of postscripts to his essay that, at first, seem optimistic, but gradually revise his estimate with increasingly conservative figures as progress on the software problem slows.

4.3 Hans Moravec

Robotics and artificial intelligence researcher and professor at Carnegie Mellon University, Hans Moravec, has provided a wealth of speculation regarding the predicted timing of the technological singularity and in particular, the hardware problem. His 1998 paper entitled "When will Computer Hardware Match the Human Brain?" establishes a detailed criterion for tracking and projecting the advances in computational hardware, using Moore's law as a basis. Throughout the work, Moravec identifies the specifications of Deep Blue as a point of comparison when illustrating computational requirements. He estimates that at some point in the 2020's, computer hardware capable of matching the performance of the human brain will be available at a consumer level. [Moravec, 1998]

Moravec goes to great lengths to show the computing hardware requirements of a system capable of producing the technological singularity. He places emphasis not only on overall computational power, but also on the memory requirements of such an application. This perspective is important because high-speed and high-capacity data storage is a factor of human intelligence that translates closely to any non-biological parallel. Like Vernor Vinge, Moravec does not include a discussion of the software requirements, nor does he account for research and construction of such software in his timing estimate of the singularity. However, he makes this clear by stressing that his prediction is relevant only to the consumer availability of capable hardware.

4.4 Ray Kurzweil

Ray Kurzweil is perhaps one of the most well known advocates of the technological singularity. He is credited with dozens of publications on the subject in the form of books, essays, papers, and blog entries. Through these mediums, Kurzweil has made a staggering number of predictions regarding the singularity as well as the events and advancements leading up to it. Although most of these estimates do not consist of steps taken explicitly or directly toward the event, they define advancements that are side effects of technological milestones along the way. Ultimately, Kurzweil believes that the singularity will occur in 2045 because he claims that by that time, "the intelligence created per year … will be about one billion times more powerful than all human intelligence today, … [which] will indeed represent a profound and disruptive transformation in human capability." [Kurzweil 2006 (135-36)] A technological advancement of that scale would certainly have a significant impact on humanity, and Kurzweil believes that this impact will come in the form of the technological singularity.

Kurzweil takes a slightly different approach to approximating the timing of the singularity in that he uses global computational potential combined with economic estimations rather than the maximum throughput of a single processor. This strategy yields a more realistic estimate because it results in a time period indicating when the global impact of the technology is reached, and where it will ultimately have its greatest effect. Kurzweil's general approach to predictions is also somewhat different than the others profiled in this paper. He tends to comment on a broad range of topics, all in some way associated with technology, but with indirect ties to the singularity, rather than directly predicting milestones on the path to the realization of the event. This is unhelpful in the task of identifying direct progress toward the goal, but is certainly useful in reconciling a more accurate timing estimate based on overall prediction accuracy.

In October 2010, Kurzweil published a paper that evaluated 147 predictions that he had made for the year 2009. [Kurzweil, 2010] By accumulating statistics about these predictions, which were made nearly 15 years earlier, an accuracy model can be developed. Once we have a measureable sense of Kurzweil's prediction accuracy, it can be applied to his predictions about computational power and the timing of the technological singularity. This will further refine his estimates and may yield a more precise prediction.

5 Reconciling a Miscellany of Predictions

The best strategy for reconciling the predictions regarding the technological singularity is to form an accuracy model of a given predictor by establishing a track record of his past success in estimating outcomes related to technology. The most convenient example of such a predictor is Ray Kurzweil, who has both a large number of organized predictions, and a mechanism by which he evaluates them. In a recent publication, Kurzweil describes the aggregate outcome of 147 predictions made in the mid 1990's for the year 2009. He states that,

> "Of the 147 predictions for 2009 … 115 (78 percent) were entirely correct by the end of 2009, and an additional 12 (8 percent) are what I would call essentially correct, for a total of 127 predictions (86 percent) out of 147 that are correct or essentially correct." [Kurzweil, 2010]

These 147 claims were carefully reviewed, and it was determined that 80 percent was actually a more reasonable accounting of predictions that were either correct or essentially correct, so this is the value assumed as Kurzweil's prediction accuracy rating. The other 20 percent were found either to be unlikely to occur in the next several years, or at all.

With this information in hand, attention can be turned to the timing prediction for the technological singularity. In 2006, Kurzweil estimated this event would occur in 2045, 39 years in the future. [Kurzweil, 2006] Assuming Kurzweil's predictions hold at the projected 80 percent accuracy rate, and this rate is extensible to his estimation of the production of global non-biological computational intelligence, a more reliable estimate for the occurrence of technological singularity is the year 2053.

6 Some Discrepant Views of the Singularity

The possibility of an event like the technological singularity rests on the assumptions that all human intelligence is reducible to computing power and that humanity will learn enough about the function of the human mind to "build one" in silicon. Although there has been some progress in this regard, there is certainly a possibility that either the hardware problem, or the software problem, or both, will not be solvable. [Bostrom, 1998] This is a view with which many thinkers, including reputable computer scientists like Joseph Weizenbaum, have taken strenuous issue. In *Computer Power and Human Reason*, Weizenbaum asks, "What is it about the computer that has brought the view of man as machine to a new level of plausibility? ... Ultimately a line dividing human and machine intelligence must be drawn. If there is no such line, then advocates of computerized psychotherapy may be merely heralds of an age in which man has finally been recognized as nothing but a clock-work." [Weizenbaum, 1976]

Vernor Vinge, who originally coined the term "technological singularity", also expresses the possibility of it being an unrealistic phenomenon. "A plausible explanation for 'Singularity failure' is that we never figure out how to 'do the software' (or 'find the soul in the hardware', if you're more mystically inclined)." [Vinge, 2007] He describes this as "the age of failed dreams" and characterizes it by exposing symptoms including several failed large-scale software projects, a collapse of Moore's law due to decreased demand, not physical barriers, and an abandonment of the field of Strong Artificial Intelligence. Although this is a view that many in the computing field see as debilitating, it is important to keep in mind that part of not knowing what the future holds, is also not knowing its limitations.

7 Conclusion and Commentary on Social Responsibility

The technological singularity is of great interest to most of those who learn about it. To the mind of a layperson, it is at once a source of wonder and apprehension. To those adept in the areas of technology and artificial intelligence, it is almost irresistibly attractive. Concurrently, it assumes several points about the nature of human intelligence and the increasing improvement rate of computing machinery that are, to this day, uncertainties. By looking closely at several of the proposed predictions for the technological singularity, a more accurate estimate as to a plausible time period for its occurrence has been determined, and possible reasons for a failed singularity have been brought to light.

There is an understandable tension between enthusiastic predictions of the advance of artificial intelligence techniques and the sober recognition of real limitations in our current understanding of human intelligence. This highlights the importance of making ethical and responsible choices with regard to care in formulating further predictions based on advances in this area of computing. This is underscored by Weizenbaum's contention that, "The computer professional … has an enormously important responsibility to be modest in [their] claims." [Weizenbaum, 1972] It is essential to note that this need for social responsibility is applicable not only to futurists who wish to make timing predictions related to the technological singularity, but also to researchers who are working to advance progress in solving its fundamental problems. Failure to do so in this particular area of interest has the potential to generate unrealistic expectations not only within the field, but also through sensational treatment by the media, in the population as a whole.

Acknowledgements

I would like to thank Dr. William Fleischman for his guidance, encouragement, proofreading, and most importantly, for sharing his wisdom. I also thank Jane Morris of Villanova University's Center for Undergraduate Research and Fellowships as well as Villanova University's Department of Computing Sciences for their support.

References

Angelini, C. (2011), Benchmark Results: SiSoftware Sandra 2011, online at http://www.tomshardware.com/reviews/core-i7-990x-extreme-edition-gulftown,2874-6.html accessed 03.03.2011

Bostrom, N. (1998), How Long Before Superintelligence? online at http://www.nickbostrom.com/superintelligence.html accessed 02.27.2011

Dubash, M. (2005), Moore's Law is dead, says Gordon Moore, online at http://www.techworld.com/opsys/news/index.cfm?NewsID=3477 accessed 02.27.2011

Kurzweil, R. (2000), *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*, Penguin Group, New York.

Kurzweil, R. (2001), The Law of Accelerating Returns, online at http://www.kurzweilai.net/the-law-of-accelerating-returns accessed 02.27.2011

Kurzweil, R. (2006), *The Singularity is Near: When Humans Transcend Biology*. Penguin Group, New York.

Kurzweil, R. (2010), How My Predictions Are Faring, online at http://www.kurzweilai.net/predictions/download.php accessed 01.19.2011

Minsky, M. (1994), Will robots inherit the earth? Scientific American, 271(4), 108-111.

Minsky, M. (1982), Why People Think Computers Can't, online at http://web.media.mit.edu/~minsky/papers/ComputersCantThink.txt accessed 2011.02.21

Moravec, H. (1998), When Will Computer Hardware Match the Human Brain? online at http://www.transhumanist.com/volume1/moravec.htm accessed 01.19.2011

Vinge, V. (1993), The Coming Technological Singularity: How to Survive in the Post-Human Era, online at http://www-rohan.sdsu.edu/faculty/vinge/misc/singularity.html accessed 01.19.2011

Vinge, V. (2007), What if the Singularity does not happen? online at http://www.kurzweilai.net/what-if-the-singularity-does-not-happen accessed 02.27.2011

Weizenbaum, J. (1976), *Computer Power and Human Reason*, W.H Freeman and Company, San Francisco.

Weizenbaum, J. (1972), On the Impact of the Computer on Society: How does one insult a machine? Science, Vol. 176, 609-614.