Memory and FLOPS Hardware limits to Prevent AGI?

Joshua J. Cogliati

July 4, 2024

Abstract: Existing discussion of AGI safety have primarily involved preventing dangerous programs from running on computers. This article focuses on examing preventing AGI from running based on hardware Memory and FLOPS limits. We show that there exist computers that are sufficiently weak that an AGI cannot run, but are high enough that practical computers can run within the limits.

Stuart Russell proposed in an interview (Chia and Cianciolo, 2023) "we need to ensure that the hardware and the operating system won't run anything unless it knows that it's safe." For sufficiently powerful computers, this is very probably correct. However, this paper will show that an alternative way to ensure safety is to limit the memory and computational speed of hardware. For this paper, Artificial General Intelligence (AGI) is artificial intelligence that is capable of performing any scientific, technological, engineering or mathematical (STEM) task that a human could. Super-intelligence is harder to define but a working definition is that a super-intelligence AGI would be capable of out thinking an entire university or research laboratory for any STEM task necessary to gain independence.[1] Having hardware limits for AGI and superintelligent AGI would be useful because these would allow safer experimenting by keeping the computer hardware below the limit. In addition, this would allow computer hardware below the limit to avoid regulations need for safe computer usage. Note that with the AGI definition does not include any speed or timing considerations, only the super-intelligence definition includes speed.

It is worth noting that the Halting Problem and Rice's theorem are for Turing machines with an infinite tape; this paper is dealing with machines with finite space (memory + storage), so there are facts that are provable that would not be with a Turing machine.

1 AGI Limits

There do not seem to be many existing estimates for a limit below which an AGI is not possible. One estimate is that a human level AGI could be done on a 286 if the programmer is a superintelligent AGI or a "home computer from 1995" (which roughly corresponds to a 90 MHz Pentium) if the programmer is a human (Yudkowsky, 2022) but no method for how this estimate was calculated is provided.

This paper is concerned with an AGI that is capable of achieving independence. There are three basic ways that an AGI could use to achieve independence. The three are convincing humans to help, creating hardware in the environment, or expanding into other computer infrastructure. Expanding into other computer infrastructure is already something that has been done by computer virus for decades, and can gain other resources which can be used for one of the other methods to achieve independence. Convincing humans probably requires at least some level of fluency in language. Creating hardware in the environment requires both some knowledge of the environment and some ability to simulate it. Computer virus can be written in 10s to 100s of instructions, so preventing this is primarily a matter of preventing there from being available targets, and will not be

[1] For this definition, the university or research laboratory does not have electronic computing hardware, otherwise the floating point operations per second would be primarily from the computers there. This definition would be a university or research laboratory in roughly 1940. The reason the "gain independence" limitation is included is to prevent needing to simulate human brains, for which humans might have an inherent advantage.

discussed further. The method this paper uses to demonstrate that an AGI can't gain independence is to show that the available computing power does not allow fluent language and does not allow sufficient simulations.

An AGI restricted to 64 KiB of read and writable space (RAM and re-writable drives) would not be capable of gaining independence.

Fluent English language understanding and creation is likely highly to be impossible in 64 KiB. A typical human vocabulary of 20,000 words would not fit in this amount of space without compression[2] and adding definitions would increase the size even more so it would not be possible to fit a program that used unrestricted language fluently. A 64 KiB computer is unlikely to be able to simulate enough of the world to design hardware to gain independence. For example, a molecular simulation that used 4 bytes per float, 3 vectors of 3 floats, and a beginning and end state could not fit a simulation of a 10x10x10 cube in 64 KiB. So a 64 KiB computer can do neither fluent language nor complicated simulations.

There are several clarifications that need to be made about this 64 KiB limit. Networking together multiple 64 KiB computers would allow the memory space to be increased, which could allow AGI to be done on the combined computers. Even without an intentional network, there can be side channels that transmit and receive data, such as with radio waves. If time for computation is ignored (as it is for this paper's definition of independence gaining AGI) it does not matter if the storage is RAM, floppy drive, hard drive or flash drive, these all increase capabilities. Register or vector storage on the CPU needs to be counted as well. Write once, read many (WORM) media (such as paper tape, punch cards, CD-R or DVD-R) or media where there is manual work needed (such as original cassette drives that required the user to manually switch from reading to recording or UV erasable programmable read only memory (UV-EPROM)) are significantly different than RAM because of they can only be written once without intervention. Only writing once is a significant limitation for most uses in simulation or learning algorithms. In addition if they cannot be overwritten at the bit level[3] the data can be read back to see what computation was being run.

It seems likely that 64 KiB of RISC-V RV64GCV machine language code would be more than sufficient to include a transformer model training and running program, and a simple simulation of Feynman's classical physics formulation (Feynman et al., 1963, Vol. 2 Table 18-4). Alternatively the program probably could fit the standard model and general relativity instead. It seems likely that a small program could easily include enough to get to a near AGI and a basic understanding of the universe in 64 KiB of code if run on a large and fast enough computer. So 64 KiB would not be enough to run an AGI, but might be enough to store an AGI.

64 KiB may be significantly lower than needed to prevent an AGI. The SHRDLU program was only capable of discussing blocks and had a vocabulary of approximately 500 words[4], and it used approximately 450 KiB (100 to 140 K of 36 bit words from the README in Winograd (1972)). From this, it seems possible that it might be able to be demonstrated that fluent English is not possible in 1 MiB of memory. For simulation 1 MiB can do more than 64 KiB, however this would be limited to less then about 15,000 elements in a molecular simulation.

2 Superintelligence Limits?

The amount of computational power to simulate the approximately 100 billion neurons (and roughly 10,000 synapses per neuron) in a human brain is estimated to be approximately 1 exa FLOP (10^{18} FLOPS) (Chen et al., 2019). This provides an upper limit for both AGI and superintelligence. Since a human is a general intelligence, then 1 exa FLOP of performance with enough memory for the all the synapses (approximately 1 petabyte) would be sufficient. Similarly, a superintelligence could be created by simulating 10,000 humans, so multiply the AGI limits by 10,000. This however is likely to be a overes-

[2] `https://www.mit.edu/~ecprice/wordlist.10000` for example is 75880 bytes. As well word vectors usually have vector length of at least 100 (Pennington et al., 2014), so 64 KiB would not even fit a 1000 basic words with the vectors.

[3] For example, on a paper tape using ASCII, a delete (0b1111111) can overwrite other characters.

[4] estimated by from counting the DEFS in the file dictio in the source code

timate of the computing power needed because of the different characteristics of computers versus human brains. Signals in human neurons travel at about 60 m/s (Stetson et al., 1992) and signal transitions take about 1 millisecond (Kandel et al., 2000, pg. 21). Signals in computers travel at near light speed (2.0e8 m/s) and signal transitions happen on the order of 1e9 times per second. This gives significant advantages for algorithms that do not parallelize well. Estimating the computing power needed to be a superintelligence from the other direction, a human can at most do less than 100 floating point operations per second, so 10,000 humans combined have 1 MFLOP for sufficiently parallelizable algorithms and less than 100 FLOPS for non-parallizable algorithms. Considering that most scientific, technological, engineering and mathematical tasks use calculations, to be conservative, the superintelligence limit should be closer to 100 FLOPS (1e2) than 10 zetta FLOPS (1e22). Proving that searching is needed might be one way to prove that there is a higher limit than 100 FLOPS.

The brain of a fruit fly has about 100 thousand neurons and about 50 million chemical synapses (Dorkenwald et al., 2023). Scaling by the number of synapses would give a simulation computational requirement of 50 giga FLOPS. A Intel 5160 processor capable of giga FLOPS of computation was used to defeat chess grandmasters (ChessBase, 2006) which gives some idea of how well humans can search. A 1976 Cray I computer had 166 MFLOPS and 32 MiB of RAM (Patterson and Hennessy, 1998, pg. 43), to give perspective on how long MFLOP sized computer have existed. Note that none of these examples provides an amount of computing power that can be used to demonstrated that the lower limit for superintelligence is greater than 100 FLOPS. Using those computations as a anchoring point, it does seem possible that future work could demonstrate a limit for computing power that cannot become a superintelligence that is a factor of a million higher or more.

3 Conclusions

An independence gaining AGI can be prevented by restricting all computers to less than 64 KiB of R/W storage without networking. Computer simulations and other uses of computers are very useful for solving other problems that humanity has so alternatively, computers below the AGI limit can be used without restrictions, and only run safe software on computers above this limit. 64 KiB of R/W storage is a useful amount computer power and systems like the Commodore 64[5], the Nintendo Entertainment System and Arduino UNO all had 64 KiB or less of R/W storage and these had sales figures in the millions (Amos, 2021; Team, 2021). This limit is however substantially below almost all modern computing systems, with the notable exceptions of low end embedded systems and retro computing.

Determining the threshold computational speed limit for a superintelligent AGI is harder and this paper was not able to demonstrate a lower limit value above 100 FLOPS. If a higher limit cannot be demonstrated, then the way to prevent superintelligent AGI is to limit memory below the regular AGI limit.

4 Speculation and Future Work

Raising the limits from 64 KiB and 100 FLOPS seems possible, and would definitely be useful future research. 1 MiB and 100 MFLOPS probably could be demonstrated, and would much more useful. Research on if networking can be allowed would be useful. Research how much Read only, Write only and Write once/read many can be allowed would be useful.

A 10 MFLOPS 512 KiB computer with 512 KiB removable re-writable storage, a 1200 bits/sec network connection, 2 MiB UV-EPROM and a 40 MiB WORM drive could be used for many things we currently use computers for including GUI word processing, spreadsheets, email, bulletin board systems, C and MicroPython programming. Remove the network connection and this is a vastly safer environment to run AI programs that we do not fully understand.

Using high powered computers for AI research is in some sense like using a 25 kVolt AC for experiments before fully understanding electricity. It would be much safer to experiment with 3 Volt DC. We need to have a better idea what computational amounts

[5]Note that a Commodore 64 with a disk drive did have more than 64 KiB of R/W storage, but Commodore 64 could be used with a manually operated cassette tape drive.

are low enough to be safe and which can lead to accidental AGI creation.

Lastly, there is usefulness in bans that are far above the provable limits, since the danger of accidentally creating an independence gaining AGI does increase as computational power goes up.

These are my own opinions and not those of my employer. This document may be distributed verbatim in any media.

References

Evan Amos. *The Game Console 2.0.* no starch press, 2021. ISBN 978-1-7185-0060-0.

Shanyu Chen, Zhipeng He, Xinyin Han, Xiaoyu He, Ruilin Li, Haidong Zhu, Dan Zhao, Chuangchuang Dai, Yu Zhang, Zhonghua Lu, Xuebin Chi, and Beifang Niu. How big data and high-performance computing drive brain science, 2019. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6943776/.

ChessBase. The last man vs machine match?, 2006. https://en.chessbase.com/post/the-last-man-vs-machine-match-.

Jessica Chia and Bethany Cianciolo. Opinion: We've reached a turning point with ai, expert says, 2023. https://www.cnn.com/2023/05/31/opinions/artificial-intelligence-stuart-russell/index.html.

Sven Dorkenwald, Arie Matsliah, Amy R Sterling, Philipp Schlegel, Szi-chieh Yu, Claire E. McKellar, Albert Lin, Marta Costa, Katharina Eichler, Yijie Yin, Will Silversmith, Casey Schneider-Mizell, Chris S. Jordan, Derrick Brittain, Akhilesh Halageri, Kai Kuehner, Oluwaseun Ogedengbe, Ryan Morey, Jay Gager, Krzysztof Kruk, Eric Perlman, Runzhe Yang, David Deutsch, Doug Bland, Marissa Sorek, Ran Lu, Thomas Macrina, Kisuk Lee, J. Alexander Bae, Shang Mu, Barak Nehoran, Eric Mitchell, Sergiy Popovych, Jingpeng Wu, Zhen Jia, Manuel Castro, Nico Kemnitz, Dodam Ih, Alexander Shakeel Bates, Nils Eckstein, Jan Funke, Forrest Collman, Davi D. Bock, Gregory S.X.E. Jefferis, H. Sebastian Seung, and Mala Murthy. Neuronal wiring diagram of an adult brain. *bioRxiv*, 2023. doi: 10.1101/2023.06.27.546656. URL https://www.biorxiv.org/content/early/2023/07/11/2023.06.27.546656.

Richard P. Feynman, Robert B. Leighton, and Matthew Sands. *The Feynman Lectures on Physics*. 1963.

Eric R. Kandel, James H. Schwartz, and Thomas M. Jessell, editors. *Principles of Neural Science 4th Ed.* 2000.

David A. Patterson and John L. Hennessy. *Computer Organization and Design 2nd Ed.* 1998.

Jeffrey Pennington, Richard Socher, and Christopher Manning. GloVe: Global vectors for word representation. In Alessandro Moschitti, Bo Pang, and Walter Daelemans, editors, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar, October 2014. Association for Computational Linguistics. doi: 10.3115/v1/D14-1162. URL https://aclanthology.org/D14-1162.

D S Stetson, J W Albers, B A Silverstein, and R A Wolfe. Effects of age, sex, and anthropometric factors on nerve conduction measures. *Muscle Nerve*, 1992.

Arduino Team. Introducing the arduino uno mini limited edition: Pre-orders now open, 2021. https://blog.arduino.cc/2021/11/24/introducing-the-arduino-uno-mini-limited-edition-pre-orders-now-open/.

Terry Winograd. Shrdlu source code, 1972. https://web.archive.org/web/20200817093131/http://hci.stanford.edu/winograd/shrdlu/code.tar.

Eliezer Yudkowsky. Ngo and yudkowsky on scientific reasoning and pivotal acts, 2022. https://intelligence.org/2022/03/01/ngo-and-yudkowsky-on-scientific-reasoning-and-pivotal-acts/.