A Hard Energy Use Limit of Artificial Superintelligence

Klaus M. Stiefel[1] & Jay S. Coggan[1]

1. Neurolinx Research Institute, La Jolla, CA, USA

Abstract

We argue that the high energy use by present-day semiconductor computing technology will prevent the emergence of an artificial intelligence system which could reasonably be described as a "superintelligence".

This hard limit on artificial superintelligence (ASI) emerges from the energy requirements of an intelligent system more intelligent, and orders of magnitude less efficient in energy use than human brains. Furthermore, an ASI would have to supersede not only a single human brain, but a large community of humans, and hence expend multiple times the energy needed to replicate the power of a single human brain.

A hypothetical ASI would likely consume enormous amounts of energy, possibly orders of magnitude above what is available in industrialized society, making it impossible on energetic grounds alone. We estimate the energy use by ASI in excess of a human brain with an equation we term the "Erasi equation", for the *E*nergy *R*equirement for *A*rtificial *S*uper*I*ntelligence.

An additional challenge is the current developmental trajectory of AI research, the majority of which is not focused on the creation of superintelligent systems. An extremely sophisticated technology like the hypothesized ASI will typically not emerge by chance from scattered efforts.

Taken together, these arguments suggest that the emergence of an ASI is highly unlikely, if not impossible, in the foreseeable future based on current computer architectures, primarily due to energy constraints.

Keywords: Artificial Intelligence, Artificial General Intelligence, Superintelligence, Thermodynamics of Computation, Brain Energy Use

Introduction

The possible emergence of an artificial superintelligence (ASI) has been the subject of much academic discussion (Carlsmith, 2022) and science fiction literature (Lem, 1964). The idea of an entity which is significantly smarter than humans, comparable to the difference between humans and great apes, captures the human imagination. This paper outlines arguments that such a superintelligence is unlikely due to its projected energy requirements.

An important point in this context is the definition of an ASI. It is difficult to precisely define an entity which doesn't exist (yet), but its eventual architecture is neither known nor relevant for the present discussion, as the main argument relates to the estimated minimum energy use of such a system, which is independent of technical details.

The issue of whether the hypothetical ASI is directly in control of effectors (for instance the power grid of countries) or acts as an "advisor" for a government or private entity is not relevant either. The definition used encompasses any man-made computational system significantly more intelligent than humans, possibly with the ability to control the human population of Earth by means of manipulation, superior planning and foresight.

Results

We will outline arguments which show that the emergence of an ASI is highly unlikely in the foreseeable future. The main argument rests on the fact that the energetic cost of the computations performed would by far surpass the energy supply available to human civilization.

While we believe that ASI is technologically impossible to implement in present-day semiconductor technology and its high energy use, we do not believe that it is impossible in principle, as other authors do (Roli et al, 2021).

Energy Use in Biological and Engineered Computation

Whatever the architecture of an ASI turns out to be, it will be bound by the principles of thermodynamics of computation (Bennett, 1982). Reversible computation with no dissipation of energy has been proposed to work in principle (Frank, 2005) but is unlikely to be possible on the speeds necessary for conventional processors or even a superintelligent system, with great numbers of individual operations needing to be performed at great speeds.

A human brain contains about 10^{11} neurons and consumes about 12 W. A typical laptop processor uses 150 W. The fastest supercomputer at the time of this writing, Frontier, uses 21 10^{6} W to perform 1.685 ExaFLOPS (1.685 10^{18} floating point operations per second).

Assigning a computational speed to nervous systems commensurable to the widely used unit of computational power for digital computers, floating point operations per second (FLOPS), is at least not trivial, or at worst a mismeasurement or simply not comparable.

We hence give an order-of magnitude estimate of the computational efficiency of present-day semiconductor processors executing AI algorithms in comparison to biological brains. To do this we compare the energy use of a state-of-the-art, detailed simulation of parts of a mammalian brain to the energy use of an actual brain.

Our example comes from Switzerland's Blue Brain Project (BBP) of EPFL, which attempts to create a biologically realistic, data-driven reconstruction and simulation of an entire mouse brain. This intricate simulation includes details of molecules, cells and circuits that together participate in biological computation (Markram et al., 2015; Ramaswamy et al., 2018; Reimann et al., 2019; Zisis et al., 2021; Coggan et al., 2022).

The BBP uses a supercomputer roughly capable of 2 10^3 TFLOPS, with 400 TB of memory and 200 TB/s of memory bandwidth. The energy use for 720 processors involved in this simulation is around 400 kW. A simulation of 10 million neurons in a cortical circuit requires approximately 1460 TFLOPS and 270 kW to simulate 1 second of biological time, and took more than 8 hours of processing time, slower than nature by a factor of $3x10^5$.

Hence, when extrapolating to the entire mouse brain with 10^8 neurons, a simulation would require 2.7 MW. Extrapolating again to a human brain with 10^3 times as many neurons as a mouse brain, the energy requirement would be 2.7 GW (and 1.46 ExaFLOPS). This is orders of magnitude above the amount of energy a human biological brain is estimated to use, at 12 W. Based on the detailed simulations conducted by the BBP example, we estimate that biological computing is ***at least*** $9x10^8$ times more energy efficient than artificial computing architecture (Fig. 1).

We stress that this estimate is a lower bound. Although the simulations of the BBP are already highly detailed and the simulation is continuously increasing its biologically realistic complexity, the current energy estimates for simulations are a snapshot and do not yet take into consideration a significant amount of the computational complexity of brains. For example, many information-bearing processes of single cells are yet to be incorporated, such as allosteric proteins, which can assume several configurations based on binding states, biomolecular networks and numerous neuromodulatory, synaptic plasticity and adaptation factors. For these reasons, the estimated $9x10^8$ times energy efficiency differential for a large BBP mouse brain simulation still grossly underestimates the true value.

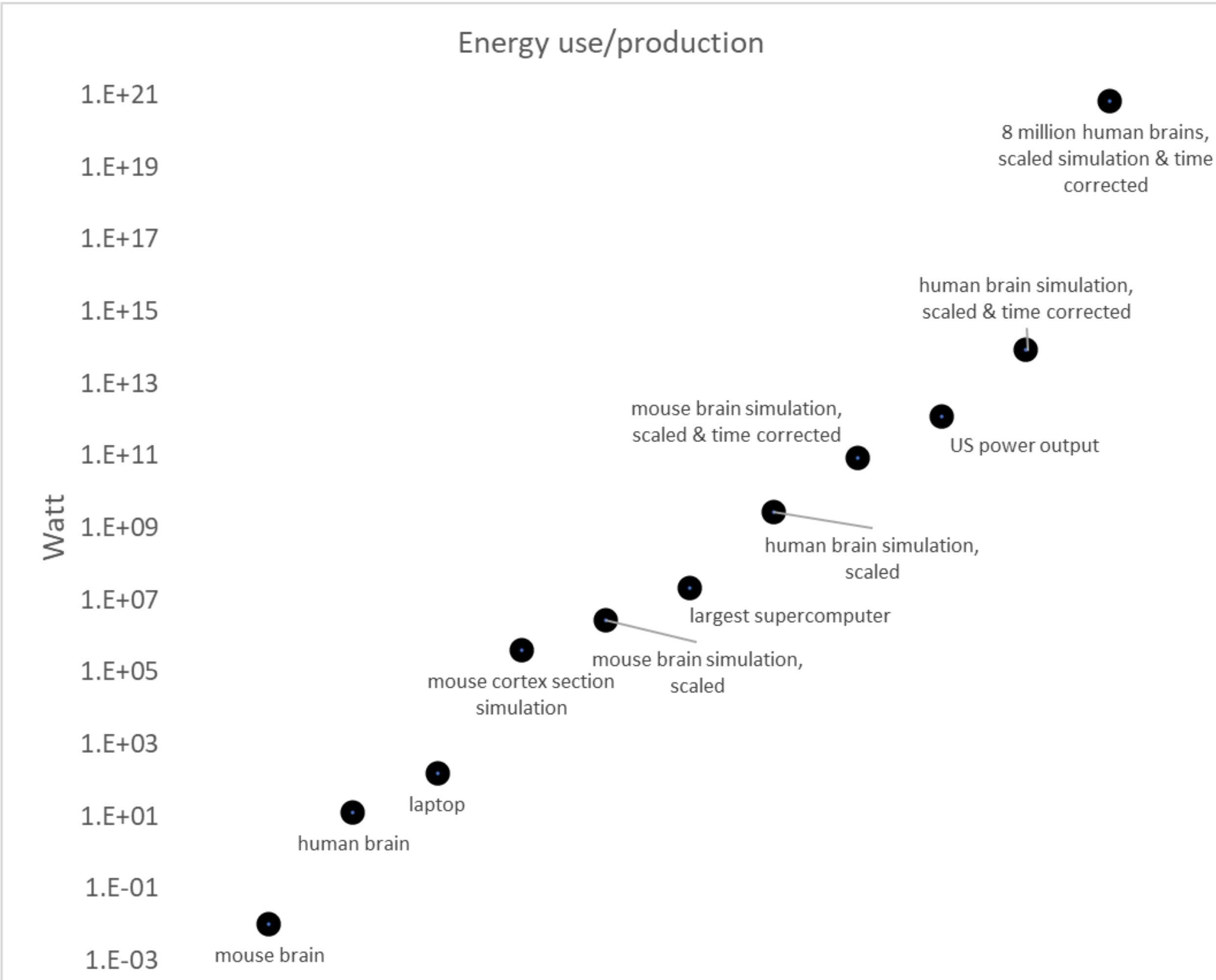

Figure 1: Energy use by the brain of a mouse, a human, a typical laptop processor, a leading supercomputer (Frontier), and the scaled energy uses (with and without corrections for processing time) for a complete mouse brain, a complete human brain and 8 million human brains.

Computing Time Considerations

This estimate above is based on 1 second of simulated biological time, but considering that it takes 3×10^5 times longer for the BBP supercomputer to simulate biological time, these simulations cannot be considered equivalent. Performing an action thirty thousand times slower is necessarily less energy demanding.

The most straightforward way to correct for this discrepancy is to multiply the relative energy efficiency of 9×10^8, derived above, by the 3×10^5, and we arrive at 2.7×10^{14} as the total relative efficiency of the human brain versus a silicone semiconductor processors running AI algorithms.

Simulation versus Emulation

The above approach is relevant especially since neuromorphic computing, computing based on architectures inspired by brain structure and function, is increasingly seen as a preferred strategy for implementing efficient computations (Indiveri et al., 2011; Wang et al, 2013; Shuman et al., 2022).

However, an important argument is that in order to replicate the performance of a human brain, one does not have to reproduce the exact structure and function of its biological intricacies. We agree with this notion, but argue that in any case the same amount of computation has to be carried out. Without doubt, a single neuron is capable of complex computations, and while they don't have to be simulated as electrical potentials traveling along axons and dendrites, the input/output relationships will have to be similarly complex. Highly simplified analog sigmoid transfer-function model "neurons" will certainly not suffice (Ananthanarayanan et al., 2009; Eliasmith & Trujillo, 2014).

And, even an estimated improvement of energy efficiency by a factor of 10^3 by an emulation (without precise biological detail) versus a simulation will only reduce, but not solve the fundamental energetic problems outlined above. It seems completely improbable, on energetic grounds, to surpass biological brains when using silicone semiconductor processors.

We speculate that only an approach that closely resembles biological computing strategies will be able to compete with biological intelligence. For example, an alternative set of large organic molecules, arranged in a multi-scale system, might be made to compute as efficiently as a brain. There is no necessity to use proteins and nucleic acids per se to build cells, but the principles of biology will have to be followed to be as energy efficient as biology. The pursuit of ASI might well benefit from biomimicry beyond today's neuromorphic strategies.

Human Group Intelligence

Humans are inherently social animals, it is therefore reasonable to compare the energy use of the brains of large human populations with that of a proposed ASI.

Even if we estimate that ~1% of the human population is mainly tasked with planning and coordination of human technological and social activities, and that they spend 10% of their lifetime actually engaged in these tasks (likely both under-estimates), then we have to assign the energy use of 8 million human brains (out of nearly 8 billion humans in 2022) to the human "group intelligence".

In reality, even the tasks performed in the construction of a footpath (involving spatial planning and the use of several tools to manipulate a variety of materials) require greater computational performance than any advanced AI system can do in 2022.

It is already remarkable that even given the astonishing computational efficiency of brains compared to computers, a large part of the planetary land area has already been modified to feed humans, and a large part of the caloric intake of humans is metabolically used by their brains.

Improvement in Understanding Reality

Another important point is by how much ASI will have to outperform humans. An often cited analogy is that ASI will be relative to humans, as we are relative to great apes. The brain of a chimpanzee is about a third the size of a human brain. Expecting one-third of the computational power and corresponding energy use for chimps is probably a reasonable minimum assumption.

Taken together, a hypothetical ASI will have to outcompete the collective intelligence of at least eight of millions of humans, each with highly energy efficient brains, and it will likely have to outcompete them by a margin of at least three.

ASI Energy Demand

To outcompete human collective intelligence within the present technological boundaries by a large margin, an ASI would have to consume a considerable amount of energy. The equation describing this energy use is:

$$E_{ASI} = E_{brain}\ f\ G\ s$$

Energy use for ASI = Energy use brain X relative computational efficiency brain/AI X human group intelligence group size X AI superiority

E_{brain} is in Watts, all other parameters are unit-less. We name this equation the **Erasi Equation** (**E**nergy **R**equirement of **A**rtificial **S**uper**I**ntelligence).

The best assumptions which we derive here are that the relative efficiency is 9 10^8 times worse in computer hardware (a measure derived from detailed brain simulations, see above), and that we need to compare the performance of an ASI to the combined intellectual output of 8 10^6 humans. Additionally, the assumption is that an ASI would have to supersede human intelligence by a factor of 3, derived from the human-chimpanzee difference. In this case the following calculation represents our best guess for the cost of ASI:

$E_{ASI} = 12\ W \times 2.7\ 10^{14} \times 8\ 10^{6} \times 3 = 7.78\ 10^{22}\ W$

An alternative, much more optimistic assumption might be that ASI would have to supersede *only a single human brain* with an emulation which is *10^3 times more energy efficient* than a brain simulation. In this case the energy use would be:

$E_{ASI} = 12\ W \times 2.7\ 10^{11} \times 1 \times 3 = 10^{13}\ W$

In February 2022, the US had a power generation capacity of more than 1.2 10^6 MW (1.2x 10^{12} W). Hence the ASI would consume power ***between ten*** and ***ten billion times larger*** than the power generation of the USA, an obviously unrealistically high value, and a value which precludes the emergence of an ASI in the absence of radical engineering advances.

Just like in the case of the Drake equation (Wallenhorst, 1981), the Erasi equation describing the number of technological civilizations in the galaxy, the above equation describes the energy requirement for ASI ***given a set of assumptions***. Just as in the Drake equation, the assumptions are up to discussion, and values for revised assumptions can be plugged-in. We argue that with any reasonable set of assumptions, the energy use will be orders of magnitude higher than that of a large, highly industrialized nation.

Discussion

The intellectual and political discourse of the future of AI has recently focused on the potential dangers of an "AI takeover" by an artificial superintelligence. Here we argue that both the basic thermodynamics of computation make such a takeover highly unlikely.

AI has brought impressive results and multiple practical uses which have already change society. But despite these successes, our arguments demonstrate, in isolation and synergistically with each other, that it is highly unlikely, if not impossible, for an ASI to emerge which will turn humans into slaves. It is likewise premature to expect salvation from ASI-like architectures in the form of the hypothesized "singularity", a time when people could upload their virtual brains into an eternal cyber-world, thus achieving immortality.

While we believe that an ASI is unlikely on energetic grounds, we disagree with arguments like those in Roli et al. (2021) that only biological organisms can show agency and hence no non-biological entity can achieve a high level of cognitive functioning.

In essence, we believe that the intricate multi-level architecture of biological brains makes them so much more energy-efficient at computing that they can achieve computational powers far beyond what is possible with silicone semiconductor chips. We might only be able to build energy efficient AGI with organic molecules following the same rules as in biology. So basically, we will have to use some form of synthetic biology to emulate the energy efficiency of biology. The whole approach of using microchips is doomed to fail, we will need a revolutionary understanding of information processing and how to achieve it with organic molecules arranged in multiple levels in order to achieve ASI.

Additional Science Policy Arguments

Not only is the emergence of an ASI unlikely for energetic reasons, but it is also not the path which the majority of research into AI is taking presently. This is both true in for the commercial applications of AI as in academic research.

The majority of research in AI appears to be concerned with classification and sorting tasks, as well as with autonomous spatial navigation. By any standards these efforts are very successful, including success in classification tasks in very high dimensional data spaces. The very successful approach of deep learning is a specialized engineering solution for classifying such high-dimensional data (Sejnowski, 2018).

AI has produced extremely impressive results in limited domains which are very dissimilar from what humans have evolved to do. One example is the success in chess, where the reigning world champion was first defeated by software in 1997. It can be argued that in chess, AI has reached superhuman intelligence. However, the intellectual challenges in chess, a highly formalized game of logic, are very different from those encountered in navigating and manipulating the real world.

Artificial general intelligence (AGI), potentially leading to an ASI, is a niche within research in AI and is not receiving the attention which many other subfields of AI do. ASI will not likely emerge by chance, just as nuclear weapons, intercontinental ballistic missiles and particle colliders (to name three of many examples) did not emerge by chance from efforts in somewhat related disciplines, but were the results of massive, concentrated efforts of large numbers of scientists, engineers and support personal.

This argument about the soft limits in achieving ASI depends very much on the politics of science, which can change very quickly. This argument on its own does not preclude the development of ASI, but in the present day it acts in synergy with the argument about the energy consumption. Essentially the soft limit, caused by the socio-political situation in AI research, keeps the state of AI from even approaching the hard limit.

Acknowledgements

We thank Drs. Dan Brooks, John Jacobson, Johannes Jäger and Rudolf Hänel for discussion of the topic.

References

Ananthanarayanan, R., Esser, S. K., Simon, H. D., & Modha, D. S. (2009). The cat is out of the bag: cortical simulations with 109 neurons, 1013 synapses. In Proceedings of the conference on high performance computing networking, storage and analysis; 1-12.

Bennett, C. H. (1982). The thermodynamics of computation—a review. International Journal of Theoretical Physics, 21(12), 905-940.

Carlsmith, J. (2022). Is Power-Seeking AI an Existential Risk? Arxiv

Coggan JS, Keller D, Markram H, Schürmann F, Magistretti PJ. (2022) Representing stimulus information in an energy metabolism pathway J Theor Biol; 540:111090.

Eliasmith, C., & Trujillo, O. (2014). The use and abuse of large-scale brain models. Current opinion in neurobiology, 25, 1-6.

Frank, M. P. (2005). Introduction to reversible computing: motivation, progress, and challenges. In Proceedings of the 2nd Conference on Computing Frontiers; 385-390.

Indiveri, G., Linares-Barranco, B., Hamilton, T. J., Schaik, A. V., Etienne-Cummings, R., Delbruck, T., ... & Boahen, K. (2011). Neuromorphic silicon neuron circuits. Frontiers in neuroscience, 5, 73.

Lem, S. (1964) Summa Technologiae

Markram, H., Muller, E., Ramaswamy, S., Reimann, M. W., Abdellah, M., Sanchez, C. A., ... & Schürmann, F. (2015). Reconstruction and simulation of neocortical microcircuitry. Cell, 163(2), 456-492.

Ramaswamy S, Colangelo C, Markram H. (2018) Data-Driven Modeling of Cholinergic Modulation of Neural Microcircuits: Bridging Neurons, Synapses and Network Activity. Front Neural Circuits. eCollection 2018.

Reimann MW, Gevaert M, Shi Y, Lu H, Markram H, Muller E. (2019) A null model of the mouse whole-neocortex micro-connectome. Nat Commun.

Roli, A., Jaeger, J., & Kauffman, S. A. (2022). How organisms come to know the world: fundamental limits on artificial general intelligence. Frontiers in Ecology and Evolution, 1035.

Schuman, C.D., Kulkarni, S.R., Parsa, M. et al. Opportunities for neuromorphic computing algorithms and applications. Nat Comput Sci 2, 10–19 (2022).

Sejnowski, T. J. (2018). The deep learning revolution. MIT press.

Wallenhorst, S. G. (1981). The Drake equation reexamined. Quarterly Journal of the Royal Astronomical Society, 22, 380.

Wang, R., Cohen, G., Stiefel, K. M., Hamilton, T. J., Tapson, J., & van Schaik, A. (2013). An FPGA implementation of a polychronous spiking neural network with delay adaptation. Frontiers in neuroscience, 7, 14.

Zisis E, Keller D, Kanari L, Arnaudon A, Gevaert M, Delemontex T, Coste B, Foni A, Abdellah M, Calì C, Hess K, Magistretti PJ, Schürmann F, Markram H. Digital Reconstruction of the Neuro-Glia-Vascular Architecture. Cereb Cortex.: 5686-5703.