On the Sustainability of AI Growth

Dr. Avishek Ranjan
Department of Mechanical Engineering, IIT Bombay, India

ARTIFICIAL INTELLIGENCE (AI) DENOTES the intelligent behavior of computer programs and machines that depend on these, in contrast to the "natural" intelligence of biological species. It can be argued that since silicon, one of the elements used in computer chips, is a naturally-occurring resource, and programming languages are invented and programs are written by (biological) humans, there is nothing "artificial" in AI. The term AI, however, has become synonymous with technological superiority and development. It was first coined by John McCarthy, one of organizers of a workshop at Dartmouth College in 1956. Earlier, the British mathematician Alan Turing had already stated that "machine intelligence" is possible in 1950 article "Computing Machinery and Intelligence".[1] A rapid progress in AI ensued thereafter but came to a halt due to a sudden drop in research funding, in part triggered by a report from the British mathematician, Sir James Lighthill said that many of AI's most successful algorithms would not work for real world problems. Apart from mathematical logic and programming, human language and its processing were recognized as being closely-linked to intelligent behavior of computers. In that era, it was thought to be nearly impossible to teach computers the nuances of language such as context. What followed until early 1980s is sometimes termed as 'AI winter' with very little research progress. Thereafter, since the late 1980s, the research progress in AI accelerated with the advent of internet, availability of large quantity of data, lowering cost of computer hardware which became smaller and faster, and the invention of Graphical Processing Units (GPUs) for fast parallel processing. Though it has been researched on and talked about for several decades now, it only since 2010 that AI along with its

Dr. Avishek Ranjan of IIT Bombay, India enjoys teaching fluid dynamics, heat transfer and thermodynamics. He is widely published.

associated mathematical methods, often termed as "Deep Learning" or Machine Learning (ML), has seen an unprecedented growth in terms of visibility and adoption, so much so that it has now become a fashion to add ".ai" in the internet domain addresses, as an expression of modernity and technical strength. The recent progress in the last five years has been claimed a huge breakthrough, some have claimed that it is perhaps as big as the invention of modern computers or that of electricity or even that of fire![2] While these may seem exaggerated, it true that some of the awe-inspiring tasks that can now be performed by computers were once only a part of imagination and science fiction.

ChatGPT, an AI tool introduced by OpenAI, is both a symbol as well as a driver of recent growth. (Here, GPT stands for Generative Pre-trained Transformer and "Chat" is perhaps used due to the fact that one can interact with the AI-tool as if it was a human). Such was the buzz around it that ChatGPT acquired 100 million users worldwide in

AI will soon affect all our lives in some way or another

just two months. Similar tools have emerged from other major companies, Google, Amazon and Meta and there is a fierce competition among these to attract more customers and retain their advertising revenue. The most recent version of ChatGPT, called GPT-4o (where 'o' stands for 'omni') includes a "chatbot" which can communicate in a 'human-like' way, along with giggles, pauses, exclamations, and understand other nuances of a natural language, such as the context.[3] It can even help solve math problems, analyze plots, translate, write or troubleshoot computer program along with human-like voice assistance. Even though it is still in its infancy, this is indeed a phenomenal achievement that was once thought to be impossible. There is a lot of debate in the news on the advantages and caveats associated with the ever-growing capabilities of AI. On the one hand are the strong proponents who promise that, just as the computers made our lives more comfortable and advanced the society in several ways, the adoption of AI by humanity is bound to make us more technologically "advanced" and even solve some of our major problems. On the other hand, there are the skeptics who ask if we should indeed allow the machines to become so powerful that they may take control of the human society one day. There is probably some truth on both sides of the argument and, as a result, the actual outcome of AI adoption may be somewhere in the middle – neither utopian nor dystopian. There is very little discussion in the mainstream media, however, on the resources needed to support this

near-unstoppable growth of AI, and hardly anyone questions if we really have them. In particular, while there is a shortage of clean energy, where will be the energy needed for the tremendous amount of computing come from? It has been argued that by 2027, AI computer servers worldwide would consume similar electricity as Argentina consumes in a year, equivalent to about 0.5% of the world's electricity use (and these numbers may change in the future).[4,5]

In this article, the questions "Is the growth of AI sustainable?" and "Is it really needed?" are discussed from various angles. The benefits of AI, that may perhaps justify the resources required, are discussed first. Thereafter, the energy and cooling requirements are discussed. I then move on to talk about the human vs machine efficiency that is closely linked to energy usage. Some thoughts on potential impact on human ingenuity, which is linked to efficiency, are portrayed before I summarize with final comments.

Necessity and Advantages of AI – recent progress

Over the course of history, the invention of better machines has helped save time, effort and money apart from making lives more comfortable. The potential in AI-based machines is usually demonstrated by its ability for fast object recognition, pattern matching, information and image processing that results in an associated "training" of the ML models. These models can then be used for applications such as self-driving

cars, or for diagnosing diseases quickly from a large number of available scans, for example. Potentially, there is possibility of time, money and human lives saved. While there are clear advantages, it is not easy to get accurate output from an AI model for an application other than the one it is trained on. Take for instance a self-driving car that is trained to run by itself in the US. It may find it extremely challenging to run in Europe (and nearly impossible in India!). Even within the US there have been several reports of fatal accidents due to self-driving cars, despite the fact that a lot of effort and money has poured in for research and development.

The development of Large Language Models (LLMs) has advanced the ways in which AI can be advantageous. For instance, if we travel to a country whose local language is completely unfamiliar to us, an AI-enabled phone camera can show the text in our native language, and an AI-enabled voice interpreter can help talk over a phone in the local language. Moreover, the advantages of ChatGPT (or its competitor tools such as Gemini by Google) are immense. These can generate a human-like text and/or voice response constructed from a database of existing knowledge on the internet. The language and paragraph-wise structuring of this response is often "too perfect" though the accuracy of content depends on the status of current knowledge. The ubiquitous "Google search" may be soon replaced (or augmented) by a response from a tool that can save a lot of our time and effort. Of course, sometimes its response may not concur exactly with the requirement(s), or it may take multiple iterations, but this technology will gradually become better.

Moreover, ChatGPT been trained to generate and correct (troubleshoot) computer code in many languages. This obviates the need to recruit and train a large number of software engineers, thus saving salary costs. Newer versions of ChatGPT, some available for free, have got several advanced features such as better understanding of nuances of natural language (including context, ambiguity and coherence), broader and more recent knowledge database in a large variety of domains (including technical, medical and legal), ability to respond based on image inputs, interactive assistance, better translation of multiple languages, enhanced ability to write creatively (e.g. poems, stories, scripts), suggestion of ideas based on inputs, better customization with the user, etc. An example of the usage is for disabled people in meeting their everyday challenges, such as a blind person trying to cross the road. Another example is as a personal assistant, or as a customized personal tutor for children who have no access to quality education, particularly in their language of choice. Khan Academy, the pioneer of digital education, has collaborated with OpenAI, the owner of ChatGPT, to come up with Khanmigo, a Chatbot designed to assist school students with subjects such as mathematics, science, humanities, and programming.[6] Rather than giving direct answers, the AI-tutor is designed to help step-by-step. For example, in order to calculate the area of a circle, rather than directly displaying the answer, the Chatbot will prompt the student to type the answer, or ask the doubts, offering hints and guidelines on the approach required. How many (human) teachers have been trained or have the ability or the patience to teach in the correct way? Whether students will really miss "the human element" of the interaction is a subject of debate, and is less likely to be a factor for the future generation of students who may be growing up with mobile phones. Let us now explore what are the energy and cooling requirements to enable a wider adoption of AI.

Energy Consumption, Growth and Distribution

There is a lot of ongoing debate on the impact of AI on jobs that may become redundant. However, here I will refrain from dissecting the potential labor market impact and discuss
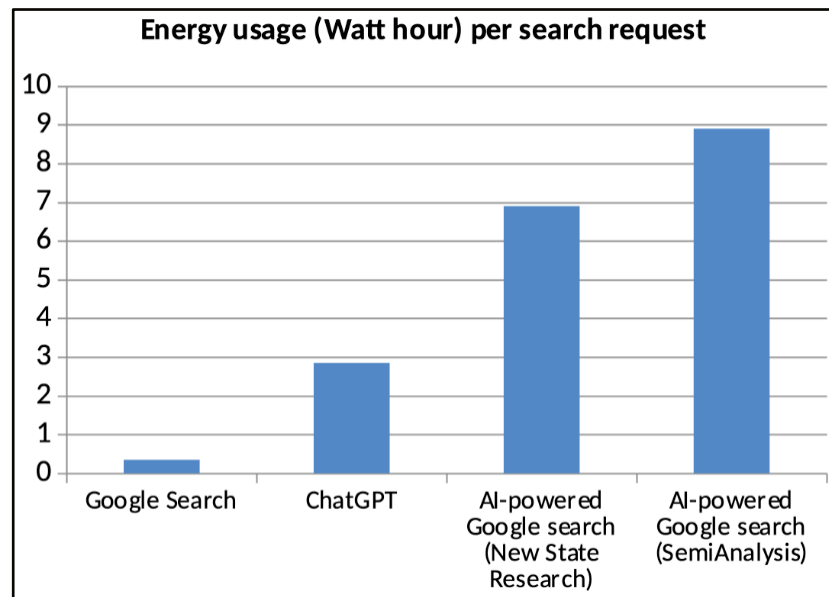

Energy usage (Watt hour) per search request

Fig 1 Estimated energy consumption for various search technologies[4]

We, as a collective society, should get to decide how much technology we must allow into our lives, so that we are its master and not the slave.

another relatively less-debated impact of AI, which has been pointed out, but probably not loud enough. According to de Vries,[4] by 2027, the total energy consumption by the AI sector could be between 85 to 134 Terawatt-hours (TWh) each year. (1 Watt-hour is equal to the energy consumed by a 1 Watt electrical device in 1 hour). This is equal to (or more than) the energy demand of many countries such as Argentina and Netherlands. A recent report by the International Energy Agency[7] offered similar estimates, suggesting that the annual energy usage at data centers, where all the internet data is stored, was around 460 TWh in 2022 and could increase to the range 620-1050 TWh by 2026. A comparison of the novel internet search methods by de Vries[4] shows that the energy consumption rises several manifolds for AI-integrated technologies as compared to the simple Google search (Fig. 1). It is possible that the hardware will also get more efficient with time, and there is indeed some evidence of this, however, the pace of this increase in efficiency is unlikely to match the phenomenal growth of AI.

Recently, attempts have been made[8] to study the carbon emissions due to ML models. However, this is specific to the computer server and its location and it is difficult in the absence of data. A wider adoption of AI-ML models requires a higher accuracy and a better training of these models. This requires more data and consequently a larger computation time to generate that data and as well as for training the model. This clearly translates to an ever-increasing energy usage. Some good practices that can be followed by the AI companies are: choosing energy-efficient hardware and computing servers or cloud storage from carbon-neutral companies, quantifying the emissions, reducing wastage of resources, identifying the problems where there is a real need of AI tools & avoiding frivolous usage, having local renewable energy sources such as solar panels along with storage, etc. While some companies such as Microsoft have initiated efforts in this direction others need to adopt. It has been argued[9] that perhaps AI can help in better (and faster) adoption and optimization of renewable energy. Even the Oil companies such as ExxonMobil and SinoPec have used AI-ML tools for improving productivity and efficiency.[10] Google has used its DeepMind

technology to decrease the cooling power requirement and its cost by 40 percent.[11] An interesting technology that is being adopted in the industry is that of a "Digital or Virtual Twin".[12] For instance, in the digital twin of a power plant, all the actual production processes can be simulated on a computer in order to optimize the resources and minimize the energy intensive trial-and-error approach. As the primary goal of a business is to have a net profit, there need to be sufficient incentives for adopting renewable energy for the use of AI. Having said this, no matter how much are the efforts, efforts must also be made to encourage consumption for need rather than for greed, in the spirit of "Aparigraha" and "Santosha" as per the cardinal rules of Yama-Niyama from Ashtaanga Yoga.[13] We all are aware of the massive degree of income and wealth inequality where many people around the world may not even have electricity for basic use, or others are facing massive outages due to shortfall. However, prioritizing need over greed is easier said than done if the society is driven by profile maximization and human comfort as compared to the well-being of all. In a warming climate, we prefer to be inside air-conditioned (AC) rooms – however, this heats up the planet even more as the heat removed is vented out to the atmosphere. (This is mandated by the laws of thermodynamics). This leads to even greater warming, and greater electricity consumption, that may sometimes come from non-renewable sources such as coal, and the cycle goes on. Those who can afford can buy comfort, but what about those cannot or those who do not have a voice, for examples plants and animals? Should they be allowed to suffer? The principles of Neo-humanism by Shrii P R Sarkar,[14] which advocate the well-being of all are, therefore, even more relevant than before in this age of digital consumerism.

Need of Alternate Cooling Technologies

Computations such as those that happen in our laptops and mobiles produce heat that must be constantly removed by the use of a fan or other cooling methods. This is because of the Joule (also called Ohmic) heating that is proportional to the square of the current and electrical resistance of the

If humans learn to realize their true potential then the human may be far superior to AI. No matter how powerful or efficient AI is at generating text, writing codes, or creating art, it will never be conscious in the same way as the humans are.

"

material. As a consequence, doubling of power requirement means quadruple heat generation for the same device. At present, air or water cooling are the most popular cooling techniques for the removal of heat. Servers produce a lot of heat in very less time, which must be removed quickly, so this requires high heat transfer rate. A "thermal paste", which has large thermal conduction properties, is typically placed on a computer chip to enable this, and recently liquid metals such as Gallium have also been explored.[15] These elements must have high thermal conductivity, a property that conveys the capability of heat transfer by conduction. Metals also have high electrical conductivity so they must be used safely. Another promising technique that has been explored for data-centers is called Liquid Immersed Cooling.[16] Here a computer server is immersed in a dielectric, a thermally-conducting but not electrically-conducting liquid, which gets heated and flows to a secondary cooling circuit, transferring its heat. If this becomes economically-viable, it is an excellent option. Researchers are also exploring the use of mathematical optimization techniques combined with the laws of thermodynamics to minimize the cooling requirements and maximize the efficiency of data-centers.[17] At present water or water-based liquids are the best (and cheapest) coolants. However, given that there is a scarcity of drinking water in many parts of the world, such as in the Indian IT city, Bangalore, how appropriate is the usage of enormous amount of clean water for cooling the computers?

Human vs Machine Efficiency

When discussing the energy requirements, the question of efficiency (e.g. energy usage per calculation) naturally arises. It is without any doubt that mathematical calculations by even an old desktop computer are much, much faster than the speed what (most) humans can perform. It may seem that computers, therefore, are more efficient than humans but that is not always the case. According to an estimate, human brain uses roughly 20 Watts to work,[18] which is less than a laptop (with consumption

~ 50 W). In the future, the laptops may come closer. However, other abilities in which humans still excel are imagination, ingenuity, ability to think out-of-the-box, discerning & judgment capacity, connecting the seemingly dissimilar ideas, etc. Indeed, in his article,[19] Noam Chomsky writes: "The human mind is not, like ChatGPT and its ilk, a lumbering statistical engine for pattern matching, gorging on hundreds of terabytes of data and extrapolating the most likely conversational response or most probable answer to a scientific question. On the contrary, the human mind is a surprisingly efficient and even elegant system that operates with small amounts of information; it seeks not to infer brute correlations among data points but to create explanations." Of course, some of these human abilities may take years to build, and even then may be present only in a tiny fraction of the population. That said, there are some abilities even in toddlers, such as learning of nuances in their mother tongue, which are acquired without a formal training, that are not easy to create in a computer with so few training resources. A couple of points in this regard that may be important in the foreseeable future are:

i. For complex AI tasks it is not just the energy consumption but also the power (that is energy per unit time) consumption that will matter. If the required power becomes lesser, then these machines will clearly be more efficient.

ii. A larger population does mean greater competition for the finite resources such as land, water, energy and food. The total cost of energy required in training a human mind (and this includes more than a decade of education and training) will very likely be more than the total cost involved for producing and creating a computer (perhaps this can also be quantified).

This is certainly controversial and may have unpredictable consequences. For example, those governments who are spending merely a pittance of the budget on education may become even more callous towards the citizens who depend on Govt. support. Energy-efficient humanoid robots will increasingly be used in the future, especially in countries where the population is declining. Perhaps

Immobile humans living in an artificial world controlled by computers – a still from the animated movie WALL-E (2008)

an increasing focus on AI may even help reduce the carbon footprint of humans on the planet, as has been claimed by some researchers.[20, 21] What may be difficult to replace is a creative and skilled/intelligent human. But, what percent of the world's population fall in this category?

A small digression seems apt here as we have attempted the difficult comparison of a human mind with a computer. The philosophy called "computationalism" argues that the relationship between the computer software and hardware is similar to that between the mind and body. Arguing that the human mind is the neuronal activity inside the (physical) brain, New York university philosopher David Chalmers identified the so-called "easy" and "hard" problems of consciousness.[22] The easy part is to know how the brain processes and control signals, makes a plan, etc through the complex electrical activity. The harder problem is explaining how why there is a consciousness or how it rises in the first place. However, this is not such a "problem" in the Indian philosophies of Vedanta and Tantra according to which consciousness is independent of the brain, and is projected from the supreme consciousness ("Shiva") to the unit ("Jiva") through the physical brain. Similar thoughts are echoed in Buddhist and Jain philosophies as well. As per these philosophies, if the humans indeed learn to realize their true potential (one of the objectives of "Yoga") then the human may be far superior. Continuing further on this line of thinking, it is not difficult to conclude that no matter how powerful or efficient AI is at generating text, writing codes, or creating art, it will never be conscious in the same way as the humans are.

Impact on Human Potential and Ingenuity

The impact of generative AI on human potential is a topic of intense debate. On the one hand, the use of tools such as interactive AI agents that can "talk" in a native language can be of massive help in filling the huge gap in facilities available in urban and rural parts. For instance, let us consider the status of education and the availability of trained educators in remote Indian villages. An estimate suggests the total number of students to be around 521 million and the number of teachers as 9.5 million, out of which a majority are in rural parts.[23] Over the last 6-7 years, "smart" mobile handsets (but this is still limited to one handset per family) and high-speed internet network connection have reached the Indian villages, along with the electrical power. This can be leveraged to help students learn with a personalized AI-based-tutor tailored to their needs, and in their own native language. Google Research, India has already built Multilingual Representations for Indian Languages (MuRIL), a ML-based model to help people build local language technologies supporting 16 Indian languages. This is being extended to 100 languages now[24]. The teaching standards in some schools are often so low (partly because teachers are poorly paid, and no one aspires to be a teacher!) that it is likely students may prefer AI-teacher. However, how this may impact the creative potential of young children remains to be seen. If the models are designed keeping this aspect in view, for example interactive tools that provoke interest and curiosity, they will be very useful, particularly at the level of school education.

On the Sustainability of AI Growth

A concern, if I can use this word, is at a higher education level, where once again there are complex challenges such as non-availability, enrollment, affordability, quality, relevance, etc. Of course, the AI-based tools can be very useful if they are made easily accessible. However, the style of education, will have to shift from the purely "content-delivery" mode to a "content- with-skill" mode, an approach which is more hands-on and includes projects rather than based merely on rote learning. This can help attract more students towards undertaking higher education and help them enhance their creative potential. This is easier said than done and it is very difficult to implement if the class size is large and the teacher, who was trained in the "older" ways, is unwilling to put in the efforts. There is, however, a greater concern. If students (or in general, the human population, that is distinct due to their thinking capacity) leaves the "thinking part" of the task to the advanced tools made available by generative AI, this is likely to be detrimental to the development of mental faculty. If the goal is only generating profit, as is the case with a typical business, which may need a high cost of resources, then the development of human potential takes a back seat. The end result may be a situation where there is a large of pool of humans with basic literacy who are addicted to the AI tools, and may even agree to pay for the use, but they are unable to earn a living as it may be more profitable to employ computer algorithms instead. This will create a different type of inequality in the society, one in which on the one hand there will be those who will develop & control the AI tools, and the others who will be controlled (evidence exists that this is already present for tools such as social media). This inequality will extend to the availability and cost of energy as well – those who need electricity will have a lot of it, others may or may not get the very basic. This may have other social consequences such as increase in crime.

Final comments

In his article,[19] Noam Chomsky says "AI's deepest flaw is the absence of the most critical capacity of any intelligence: to say not only what is the case, what was the case and what will be the case — that's description and prediction — but also what is not the case and what could and could not be the case. Those are the ingredients of explanation, the mark of true intelligence." He obviously argues that AI cannot match the likes of Albert Einstein or Steve Jobs, and the creative geniuses of their kind. They, however, represent a small minority of humans, and this will very likely remain so in the future. The majority wants peace and comfort in their lives, and they are happy to be the end-users and customers, who think of technology as a magical wand. If AI tools help them in this regard, they will be widely adopted without a concern for the energy consumption, cooling-resource's demand or its sustainability. For some causes such as that of education in remote areas, perhaps this is justified. However, given its high environmental cost, AI should be used sparingly, and only if it is actually needed to make the lives better. GPU-based computing by Nvidia is known to be more energy-efficient than CPUs, and continuous efforts are on to improve upon this. Some efforts by the big tech companies such as Google and Microsoft to achieve a net-zero carbon footprint are steps in the right direction. Increasing the capacity of renewable energy sources with adequate energy storage is another need of the hour for sustainable growth. We can hope that these will continue in the future, and the advancement of AI technology does not come at a cost of excessively warming our planet, depleting the resources, leaving it uninhabitable, increasing the already high inequality, and leaving the humans lazy, stupid and dependent. This was shown in WALL-E, the 2008 animated movie directed by Andrew Stanton, written by Stanton and Jim Reardon (image above). The film, which some would say was much ahead of its time, portrays consumerism, corporatocracy, waste management, human environmental impact and concerns, obesity/ sedentary lifestyles, and global catastrophic risk.[25] The danger is not so much from the so-called malevolent-superpowerful-AI which "will rule over the humans" as it is from the (human) leaders who are spearheading the AI development. It is crucial that these leaders have their moral conscience in the right place, and that there is some global organization that closely monitors the progress. We, as a collective society, should get to decide how much technology we must allow into our lives, so that we are its master and not the slave.

The article, complete with all its references, is published on theneohumanist.com

Dr. Avishek Ranjan is a faculty member in the department of Mechanical Engineering at IIT Bombay, India. He completed PhD from University of Cambridge, UK, and enjoys teaching fluid dynamics, heat transfer and thermodynamics. Along with his research group members, he works on computational fluid dynamics and magneto-hydrodynamics, with applications in liquid metal batteries, aluminum reduction cells, geophysical flows in the atmosphere, oceans and the core of the Earth. He is widely published.