R Street
Free markets. Real solutions.

R Street Shorts
No. 129
September 2023

Artificial Intelligence, Energy and the Economy

By Wayne T. Brough

Rather than create overly rigid rules that can quickly become obsolete, regulators should work to ensure that electricity generation can keep pace with growing demand in order to fully capture the potential benefits generated by AI-based technologies.

Introduction

The release of ChatGPT on Nov. 30, 2022, sparked a global conversation about the future of computing.[1] By January 2023, this large language model had 100 million users—a number that rose to 173 million by April 2023.[2] Created and released by the company OpenAI, ChatGPT gave the broader public its first direct access to the computational power of artificial intelligence (AI) and its ability to provide human-like text generation with highly accurate contextual understanding.

These advances in AI and machine learning from ChatGPT and other AI models like Google's Bard, Meta's LLaMA and other open-source models could fundamentally alter the future of computing and improve our quality of life in countless ways, such as enhancing scientific learning, managing complex systems, and improving the productivity and output of virtually all sectors of the economy.

Despite these potential benefits, important ethical questions and societal concerns have been raised about some applications of AI, including possible misuse or malevolent use as well as privacy and cybersecurity risks.[3] Critics also fear job losses, as many routine tasks could be automated with AI, although the overall potential impact of substituting AI-assisted capital for labor remains ambiguous because the use of these technologies and the resultant expanding marketplace is expected to generate new employment opportunities in new fields.[4] These potential adverse consequences have caused many to call for regulation.

Additionally, transitioning to an AI economy raises concerns about energy consumption and environmental impacts, as the data centers that house the hardware and other resources required for AI technologies use a considerable amount of power. Estimates suggest that the centers consume 2 to 3 percent of U.S. and global power.[5] Given this concern, AI energy consumption and its resultant carbon footprint require an appropriate policy response that broadly evaluates the holistic impacts of AI,

Transitioning to an AI economy raises concerns about energy consumption and environmental impacts.

The data centers that house the hardware and other resources required for AI technologies use a considerable amount of power. It is estimated they consume

2–3%

of U.S. and global power.

considering both the direct power used by AI applications as well as their ability to improve energy efficiency and lower carbon emissions in key sectors of the economy.

In this paper, we explore these issues by briefly explaining the accelerated computing behind AI models and the key benefits the technology offers. We then discuss two of the bigger concerns surrounding AI and machine learning—energy consumption and regulation—and offer suggestions that policymakers can consider to ensure that the technologies are able to flourish safely and responsibly.

Much of the consternation over the existential threats posed by AI's ability to mimic human rationality is misplaced, given that current AI models are generative.

Understanding AI, Accelerated Computing and Energy Needs

Much of the consternation over the existential threats posed by AI's ability to mimic human rationality is misplaced, given that current AI models are generative. Generative AI is not the same as artificial general intelligence where machines think and learn like humans; rather, generative AI models create predictive content based on the datasets on which they are trained.[6] ChatGPT, for example, relies on a generative AI model—trained on a large dataset (175 billion parameters originally and 1.76 trillion in its latest iteration)—to predict patterns in responses to user queries and produce text-based content.[7]

Accelerated computing is at the heart of generative AI, providing the computational power necessary to build and run these complex models. It combines the use of graphics processing units (GPUs) with central processing units (CPUs) and other specialized hardware like tensor processing units (TPUs) to allow large datasets to be processed much more quickly and efficiently than traditional CPUs alone. And while CPUs rely on sequential calculations, GPUs, TPUs and other hardware accelerators have been developed to efficiently perform the parallel processing required to run effective AI.

Importantly, although AI-optimized hardware may lower per-task energy consumption because of its efficiency, the total energy consumption of AI workloads can still be significant because AI models are typically large and run continuously. This is especially true when creating models capable of tasks such as image recognition, natural language processing and predictive analytics. Consider, for example, the real-time data and analysis required to pilot autonomous vehicles, where data from onboard sensors is processed immediately and requires detailed navigation maps, object recognition and car-to-car communication. One estimate suggests that a single autonomous vehicle can generate up to 5,100 terabytes of data annually.[8]

Because of the size and energy requirements of AI models, concerns over energy consumption are factoring into to technological improvements, as energy efficiency is seen as a key factor in the sustainability and cost-effectiveness of AI technologies. Hardware manufacturers have economic incentives to develop products that reduce energy consumption to expand market share over their rivals. Likewise, data centers have incentives to reduce overall power usage with energy-efficient cooling systems and more efficient energy management.[9]

Potential Benefits of an AI Economy

The use of machine learning and AI modeling is increasingly important for the large-scale, real-time data processing and analytics that are underlying advances in many

fields. This increased processing power is useful for applications in which immediate responses are needed, such as the use of autonomous vehicles, where decisions must be made in fractions of a second based on large amounts of incoming sensor data. Similarly, traffic management systems must identify and resolve bottlenecks and other problems in real-time, which optimizes traffic management while reducing overall carbon emissions.

Economic Benefits

AI is also useful when data-intensive research is required, such as in the pharmaceutical sciences where new and personalized drug therapies could be developed at significantly lower costs and a more rapid pace. In fact, the pharmaceutical industry is already applying such technology, and several AI-designed drugs are now moving on to clinical trials.[10] In agriculture, AI can improve crop yields, reduce pesticide use and optimize the supply chain for agricultural products, and in the energy field, it can be utilized to optimize electricity grid management while also making buildings greener and smarter.[11] Table 1 highlights these and other potential benefits offered by AI and machine learning.

Table 1: Potential Economic and Social Benefits of Machine Learning and AI Modeling by Industry

Industry	Beneficial Uses of AI	Industry	Beneficial Uses of AI
Healthcare	• Improve diagnostics and patient care • Enable personalized treatment programs • Predict disease outbreaks • Optimize hospital operations • Accelerate drug discovery	**Retail**	• Personalize customer experiences • Optimize inventory management • Enhance logistics • Enable "smart" marketing
Agriculture	• Optimize crop yields • Predict disease and pest outbreaks • Automate farming tasks • Improve supply chain efficiency.	**Finance/ Insurance**	• Assess and manage insurance risks • Detect fraud • Optimize investment strategies • Personalize financial services
Transportation and Logistics	• Enable autonomous vehicles • Optimize routing and delivery schedules • Improve supply chain efficiency • Enhance safety	**Education**	• Personalize learning • Automate grading and feedback • Predict student outcomes • Enable new modes of online learning
Energy	• Optimize grid management • Increase the efficiency of renewable energy sources • Reduce energy usage in buildings and industry	**Environment**	• Enhance climate models • Optimize resource use • Monitor and evaluate environmental risks
Manufacturing	• Automate manufacturing tasks • Optimize production schedules • Improve quality control • Optimize supply chain management	**Public Safety and Disaster Response**	• Predict and respond to natural disasters • Enhance emergency services • Improve weather forecasting and storm tracking • Optimize resource allocation during crises

Greening AI

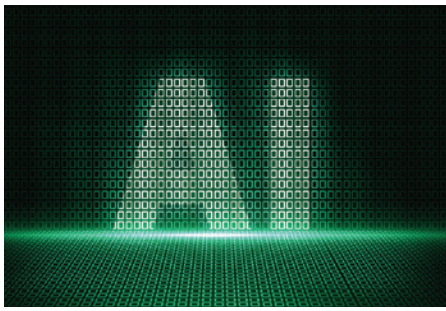
Despite the potential economic and social benefits of AI, one of the notable concerns with its deployment is its carbon footprint and potential impact on energy consumption. Fortunately, industry is adapting as it expands, adjusting and implementing practices to reduce the overall carbon footprint and energy consumption of AI technologies. One area of research focuses on improving the efficiency of the hardware required for AI modeling. As noted previously, data centers house specialized hardware designed specifically for accelerated computing, such as the GPUs and TPUs needed for machine learning and large datasets. As the demand for AI increases, the development of more efficient hardware solutions that consume less energy becomes a market advantage.

For example, NVIDIA is developing a new generation of chips and hardware to more effectively optimize the use of GPUs and other hardware in ways that lower energy

consumption and improve the cost-effectiveness of AI and machine learning.[12] While NVIDIA is currently the market leader with respect to GPUs designed for machine learning and data centers, competitors such as AMD and Intel are developing their own GPUs, as are a host of startups.[13] Even Google and Amazon are entering the chip market.[14]

In addition to more efficient hardware, machine-learning algorithms can be optimized to reduce their computational requirements and thereby the energy required to run them. Techniques such as pruning, quantization and knowledge distillation can simplify models and reduce the amount of computation required, leading to energy savings.[15] Likewise, energy-efficient coding practices can reduce the energy needed to run AI applications.[16] For instance, coding that optimizes data structures, avoids unnecessary computations and runs on energy-efficient algorithms can contribute to reduced energy requirements for AI models. Additionally, smaller, compact models require less computation and therefore less energy consumption; identifying where they may be used in place of larger, more complex models while still yielding the appropriate level of accuracy and precision is yet another path to lower AI energy requirements.[17]

It is also possible to alter the location where computing is carried out to minimize energy use. The incorporation of edge computing, which allows local devices to perform some of the necessary computations rather than a data center, can reduce energy costs, particularly those associated with data transmission.[18] Edge computing can also reduce latency, which is especially important when using AI applications that require real-time responses. In effect, edge computing brings the benefits of AI to local uses, such as homes, factories, or smart instruments and appliances that make up the Internet of Things.



Data centers can lower energy consumption by adopting more efficient power supplies and cooling systems and improving server utilization. Importantly, AI itself can help data centers more effectively manage these factors and identify opportunities to reduce energy inputs.

Additionally, data centers can lower energy consumption by adopting more efficient power supplies and cooling systems and improving server utilization. Importantly, AI itself can help data centers more effectively manage these factors and identify opportunities to reduce energy inputs. In fact, Google, Meta and Microsoft all rely on AI to improve the functioning of their data centers.[19] Google, for example, used its DeepMind AI to reduce the energy used for cooling its data centers by up to 40 percent.[20] There are also opportunities to incorporate renewable energy more effectively into the operations of data centers. The Citadel Campus in Reno-Tahoe—the largest data center in the world at 7.2 million square feet—is powered by renewable energy.[21]

Finally, it is crucial to remember that beyond data centers, AI also has the potential to contribute significantly to reducing carbon emissions and lowering energy consumption throughout the economy. AI models can optimize energy use in buildings and industry, enhance electricity grid management and facilitate more efficient transportation systems, among other applications.

Thus, energy consumption by AI and machine learning is a concern worth addressing, with the understanding that AI is a new field emerging in a dynamic marketplace, and new technologies, greener coding and more efficient energy management are all already combining to reduce the current energy demand. These practices should be encouraged, and policymakers and regulators should be cognizant of both the evolving nature of the market and the inherent, market-driven incentives that reduce the overall power load of data centers.

Regulate or Innovate?

Given the rapidly evolving nature of AI technology and the concern regarding its high demand for energy, it is understandable that policymakers, technologists and other stakeholders are debating regulation. Yet the increase in the demand for energy must be evaluated within a proper framework, which requires comparing the energy use and outcomes achievable through the computing power of AI to the alternative. That is, we must consider how much energy would be required to generate the same outcomes and increases in productivity and economic growth in the absence of AI and machine learning, if those same gains are even feasible. Such an approach would highlight both the costs and benefits of the increased energy consumption of AI-based technologies.

Still, should a market failure be identified, it is essential that any regulatory framework considered be flexible enough to allow stakeholders and other parties to innovate and identify market-driven solutions to reduce energy demands. Overly prescriptive rules would likely become outdated quickly, locking in suboptimal technologies and unnecessarily impeding newer, more efficient approaches for managing energy consumption. Excessive regulation would also hamper the adoption of newer technologies or the deployment of renewable energy resources that could offer significant economic and social gains, including optimizing the power grid to more efficiently generate and deliver electricity.[22]

Regulating this technology would also be challenging because AI is not confined to a specific sector of the economy; it spans a wide variety of sectors such as healthcare, finance, transportation, energy and more—all of which have existing regulators with the authority to intervene in the marketplace. Before creating new regulatory bodies, it would be prudent to identify specific weaknesses, market failures or challenges that cannot be resolved within existing regulatory frameworks. To do so, policymakers will require a deep understanding of the technology, its applications and its implications. Close collaboration between policymakers, AI researchers, industry stakeholders and the public will yield a more robust policy framework for AI technologies.

Conclusion

Rapid advances in AI and machine learning are poised to drive economic growth by automating routine tasks, increasing productivity, enhancing business processes and reducing waste. Although managing the energy needs of AI models is an important issue, high energy costs offer natural incentives to develop technologies and products that increase efficiency and lower the energy costs required to run AI-based technologies. This continued evolution of AI technology requires a regulatory framework that fosters flexibility and promotes innovation. Rather than create overly rigid rules that can quickly become obsolete, regulators should work to ensure that electricity generation can keep pace with growing demand in order to fully capture the potential benefits generated by AI-based technologies.

Rapid advances in AI and machine learning are poised to drive economic growth. Although managing the energy needs of AI models is an important issue, high energy costs offer natural incentives to develop technologies and products that increase efficiency and lower the energy costs required to run AI-based technologies.

About the Author

Wayne T. Brough is policy director for R Street's Technology and Innovation team. He manages product flow on technology policy issues and conducts research in competition policy, innovation and intellectual property.

Endnotes

1. "Introducing ChatGPT," OpenAI, last accessed July 25, 2023. https://openai.com/blog/chatgpt.
2. "97+ ChatGPT Statistics & User Numbers in July 2023 (New Data)," NerdyNav, last accessed July 25, 2023. https://nerdynav.com/chatgpt-statistics.
3. Durga Ramakrishna (Krishna) Gadiraju, "Navigating the Unforeseen Risks of Generative AI Technology," IEEE Computer Society, July 6, 2023. https://www.computer.org/publications/tech-news/trends/risks-of-generative-ai.
4. Brendan Rearick, "Is AI Coming for Your Job? 65% of Workers Are Worried," *Money*, June 2, 2023. https://money.com/ai-job-loss-worker-concerns; Ajay Agrawal et al., "Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction," *Journal of Economic Perspectives* 33:2 (Spring 2019), pp. 31-50. https://www.aeaweb.org/articles?id=10.1257/jep.33.2.31.
5. Ajay Kumar and Tom Davenport, "How to Make Generative AI Greener," *Harvard Business Review*, July 20, 2023. https://hbr.org/2023/07/how-to-make-generative-ai-greener; Karen S. Freeman, "AI and Energy Consumption: Are We Headed for Trouble?," iMore, June 10, 2023. https://www.imore.com/apple/ai-and-energy-consumption-are-we-headed-for-trouble#:~:text=Energy%20from%20data%20centers%20consumes,average%20households%20for%20a%20year.
6. David Leslie, "Does the sun rise for ChatGPT? Scientific discovery in the age of generative AI," *AI and Ethics* (July 5, 2023). https://doi.org/10.1007/s43681-023-00315-3.
7. "GPT4 has more than a trillion parameters – report," The Decoder, March 25, 2023. https://the-decoder.com/gpt-4-has-a-trillion-parameters.
8. Rich Miller, "Rolling Zettabytes: Quantifying the Data Impact of Connected Cars," Data Center Frontier, Jan. 21, 2020. https://www.datacenterfrontier.com/connected-cars/article/11429212/rolling-zettabytes-quantifying-the-data-impact-of-connected-cars.
9. Office of Energy Efficiency & Renewable Energy, "Energy 101: Energy Efficient Data Centers," U.S. Department of Energy, Sept. 16, 2013. https://www.energy.gov/eere/articles/energy-101-energy-efficient-data-centers#:~:text=Data%20centers%20can%20become%20more,of%20energy%2Dintensive%20air%20conditioners.
10. Carrie Arnold, "Inside the nascent industry of AI-designed drugs," *Nature Medicine* 29 (June 1, 2023), pp. 1292-1295. https://doi.org/10.1038/s41591-023-02361-0.
11. Akriti Taneja et al., "Artificial Intelligence: Implications for the Agri-Food Sector," *Agronomy* 13:5 (July 5, 2023), p. 1397. https://doi.org/10.3390/agronomy13051397; Sean Captain, "How AI Might Change the Way We Supply and Consume Energy," *The Wall Street Journal*, July 20, 2023. https://www.wsj.com/articles/artificial-intelligence-technology-energy-a3a1a8a7?mod=ig_energyreport.
12. George Lawton, "How Nvidia is driving greener computing," VentureBeat, Nov. 14, 2022. https://venturebeat.com/data-infrastructure/how-nvidia-is-driving-greener-computing.
13. Yiannis Zourmpanos, "AMD: Leading the Data Center Market," Seeking Alpha, Dec. 19, 2022. https://seekingalpha.com/article/4565157-advanced-micro-devices-stock-amd-leading-data-center-market; Cem Dilmegani, "Top 10 AI Chipmakers of 2023: In-depth Guide," AIMultiple, last accessed July 23, 2023. https://research.aimultiple.com/ai-chip-makers.
14. Danny Vena, "Alphabet Says Google's Latest AI Chip is 1.7 Times Faster Than Nvidia. There's a Catch.," The Motley Fool, April 6, 2023. https://www.fool.com/investing/2023/04/06/alphabet-says-googles-latest-ai-chip-is-17-times-f; Nicholas Rossolillo, "Amazon CEO Explains Machine Learning Chip Investments – Is Nvidia Stock In Trouble?" The Motley Fool, April 20, 2023. https://www.fool.com/investing/2023/04/20/amazon-ceo-on-ml-chips-is-nvidia-in-trouble.
15. See, e.g., "What is Pruning in Machine Learning?," ODSC, Aug. 5, 2020. https://opendatascience.com/what-is-pruning-in-machine-learning; "Benefits Of Using ML Quantization in AI Projects," rinf.tech, last accessed July 25, 2023. https://www.rinf.tech/5-reasons-why-machine-learning-quantization-is-important-for-ai-projects; Petru Potrimba, "What is Knowledge Distillation? A Deep Dive.," Roboflow, May 16, 2023. https://blog.roboflow.com/what-is-knowledge-distillation.
16. "Why Green Coding is a Powerful Catalyst for Sustainability Initiatives," IBM, last accessed July 25, 2023. https://www.ibm.com/cloud/blog/green-coding.
17. Gadi Singer, "Survival of the Fittest: Compact Generative AI Models Are the Future for Cost-Effective AI at Scale," Towards Data Science, last accessed July 25, 2023. https://towardsdatascience.com/survival-of-the-fittest-compact-generative-ai-models-are-the-future-for-cost-effective-ai-at-scale-6bbdc138f618.
18. Kashyap Vyas, "Edge AI: The Future of Artificial Intelligence and Edge Computing," IT Business Edge, Aug. 25, 2021. https://www.itbusinessedge.com/data-center/developments-edge-ai.
19. Kyle Wiggers, "Microsoft and Meta join Google in using AI to help run their data centers," TechCrunch, June 18, 2022. https://techcrunch.com/2022/06/18/microsoft-and-meta-join-google-in-using-ai-to-help-run-their-data-centers.
20. Richard Evans and Jim Gao, "DeepMind AI Reduces Google Data Centre Cooling Bill by 40%," Google DeepMind, July 20, 2016. https://www.deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-by-40#:~:text=Our%20machine%20learning%20system%20was,and%20other%20non%2Dcooling%20inefficiencies.
21. "Switch TAHOE RENO Now Open: Largest, Most Advanced Data Center Campus in the World," Switch, last accessed July 25, 2023. https://www.switch.com/switch-tahoe-reno-data-center-now-open.
22. Philip Rossetti, "The Environmental Case for Improving NEPA," R Street Institute, July 7, 2021. https://www.rstreet.org/commentary/the-environmental-case-for-improving-nepa.