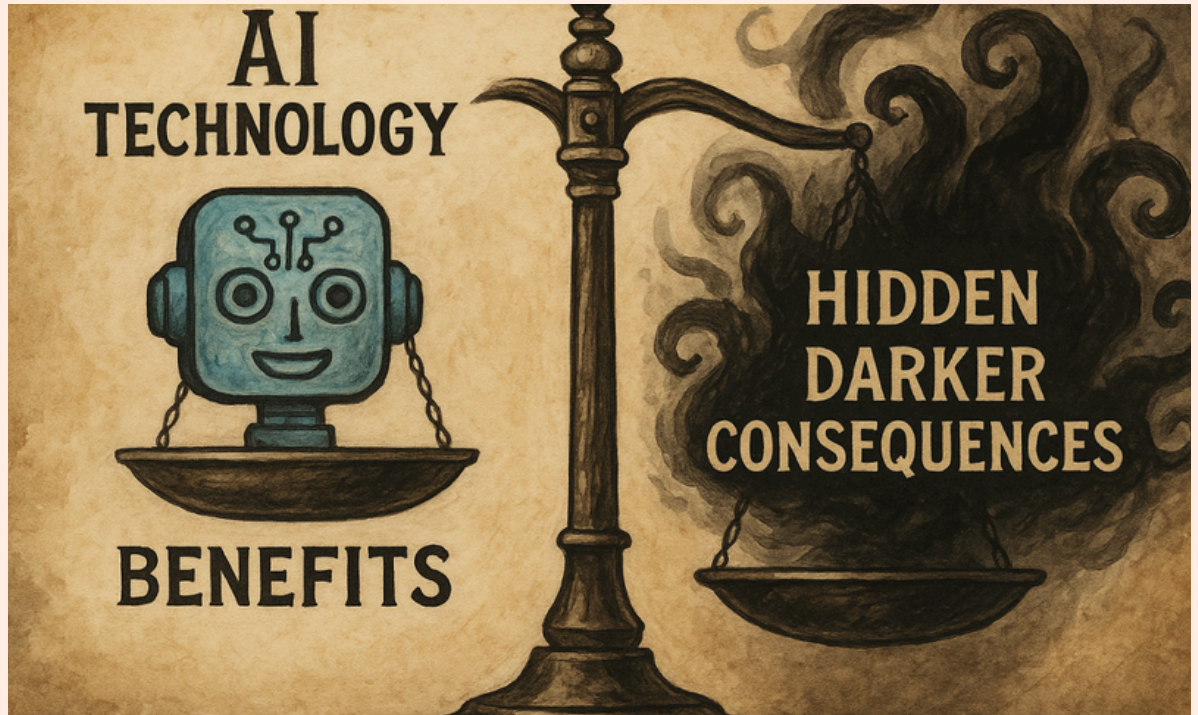


**Bansilal Ramnath Agarwal Charitable Trust's
Vishwakarma Institute of Technology ,Pune-411037
(An Autonomous Institute affiliated to Savitribai Phule University)
DEPARTMENT OF INFORMATION TECHNOLOGY**



Vishwakarma Institute of Technology, Pune

Welcome to the IT Bulletin – September 2026 Edition

The Hidden Carbon Footprint of an AI Query

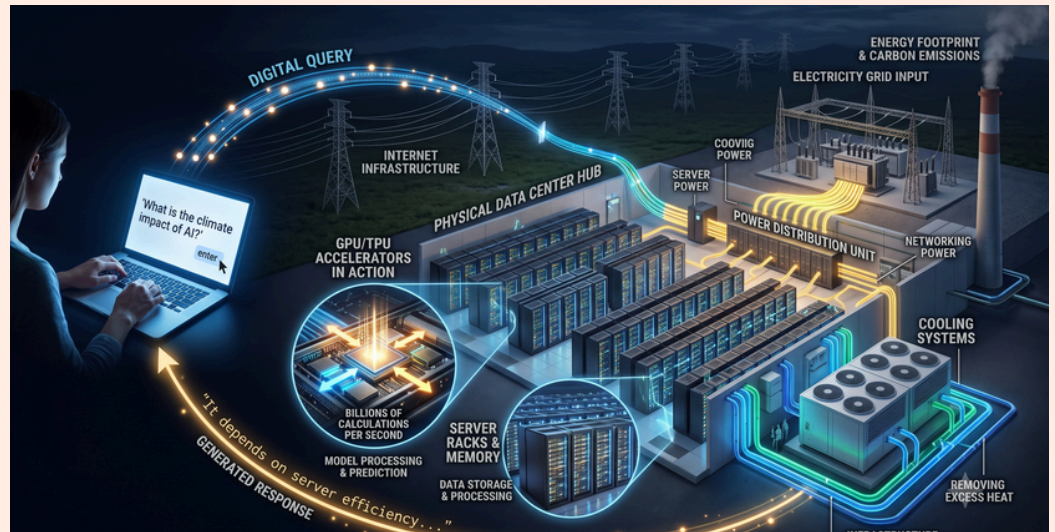
This issue dives into the unseen environmental toll of artificial intelligence, such as the carbon footprint of just one AI query. Given that AI technologies are becoming more integrated into education, software development, research, and life in general, each prompt demands computation, energy and data-center services.

AI models require powerful servers that use energy to process your question and other resources to cool and run a data center. While the carbon footprint of each query might be negligible, it adds up when there are billions of AI questions being asked around the world. The true impact also varies by model, complexity of the question, hardware, energy source, and data center.

By being aware of the environmental costs of AI, students and developers will also be mindful of the ways in which to use AI in a "responsible and sustainable way". Using more efficient models, limiting number of queries, training more effective models, and making better use of computing infrastructure, among others.

Find out how just chatting with AI can make a difference when it comes to energy, carbon, responsible computing, and the future of responsible technology - and why creating a more intelligent digital world can lead to a more sustainable one.

WHAT HAPPENS WHEN YOU SEND AN AI QUERY?



When you type a question into an AI chatbot and hit Enter, a lot happens behind the scenes in a fraction of a second. Your request travels over the internet to a data center, where it is processed by powerful servers running specialized chips called GPUs (Graphics Processing Units) or TPUs (Tensor Processing Units).

These chips perform billions of calculations to understand your input and generate a response. The servers require electricity not only for computation but also for memory, networking, storage, and cooling systems. Large data centers use sophisticated cooling infrastructure to prevent the hardware from overheating.

The AI model then processes your words, identifies patterns from its training, and predicts an appropriate sequence of words to form the response. Once generated, the response travels back through the internet and appears on your screen—often within seconds.

Although the process feels almost instantaneous, it relies on a large physical infrastructure of servers, networking equipment, cooling systems, data storage, and electricity. This is why every AI query, despite being digital, has a real-world energy and environmental footprint.

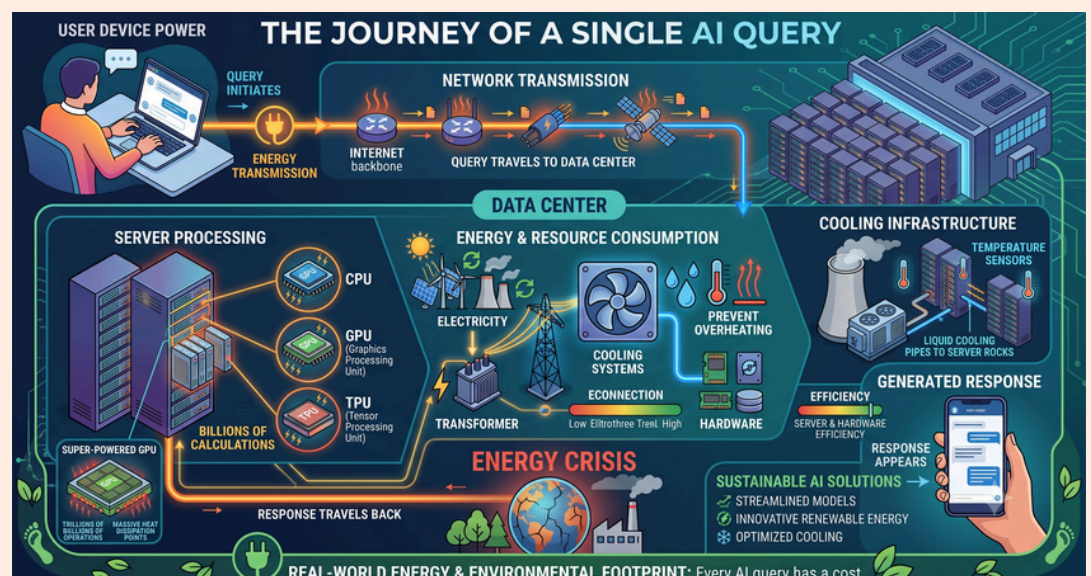
The Energy Behind Every AI Response

With every response AI sends back to you, energy goes from your devices to the data center, it does stuff with that energy and sends it back to you. If you were able to see the computers under that single line of text there, you'd see a super-powered GPU with its supporting hardware cranking out trillions of billions of operations.

Here the energy is dissipated at points. The hardware of AI systems consists of devices performing the computation, which run with data transferred via networking hardware and controlled via temperature sensors for the server room's temperature and cooling system. While the development of large AI models can involve huge computing power, the on-going demand is constant as millions of users make queries.

A lot of time, a single query is actually composed of a number of individual searches on models, what the size of the information you request is, what content is used and how efficient the server and hardware are. There is actually no fixed energy cost for each AI response.

An Energy Crisis The ever-more importance of having to consume less energy as AI grows bigger. The continued quest to find even more efficient processors and more streamlined AI, plus new forms of renewable energy, innovative cooling and data centre optimisation will all be vital in driving sustainability for AI.



DATA CENTERS: THE POWERHOUSE OF AI

I can imagine how behind every AI answer there is a network of data centers. These data centers are buildings that hold thousands of strong computers and servers. Each of those machines gives the power that trains AI models, answers user questions writes replies. Keeps huge amounts of information.

When AI works it usually needs GPUs and other specialized processors. GPUs can finish the calculations that AI models need at the same time. GPUs and other processors use a lot of electricity and make a lot of heat. I see that cooling is an important part of data center work.

The environmental impact of a data center depends on several things. It depends on how efficient the hardware's how good the cooling systems are and where the electricity comes from. I notice that a data center that uses power has a smaller carbon footprint than a data center that relies on fossil fuel.

As more people use AI data centers are changing to be energy-efficient and sustainable. New ideas like liquid cooling, better processors better use of servers and the use of energy are helping cut the resources that a data center needs to run AI . I see how these changes help keep the planet safe.

In terms: Whenever I talk to AI my request goes to a data center somewhere in the world. There powerful computers do the work, behind the scenes. I understand that these digital services even though they seem invisible have a physical and environmental impact.

UNDERSTANDING THE CARBON FOOTPRINT OF AI

Understanding the Carbon Footprint of AI

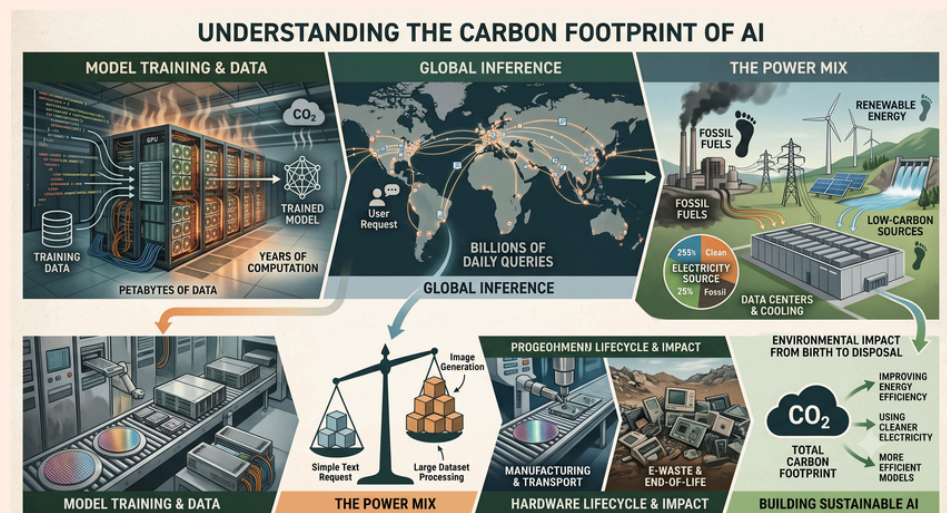
Every AI interaction has an environmental footprint because the computers running AI systems require electricity. The carbon footprint of AI refers to the greenhouse-gas emissions associated with the energy and resources used to develop, operate, and maintain AI systems.

The amount of carbon emissions is not the same for every AI query. It can depend on the size and type of AI model, complexity of the task, amount of data processed, computing hardware, data-center efficiency, and the source of electricity. A simple text request may require relatively little computation, while tasks such as generating images, processing large datasets, or running complex models can require considerably more.

AI's carbon footprint can be broadly considered across several stages:

Model Training: Large amounts of computation may be required to train an AI model.

Model Inference: Energy is consumed each time the trained model processes a user request.



Data Centers: Servers, networking equipment, storage, and cooling systems require additional energy.

Hardware Lifecycle: Manufacturing, transporting, maintaining, and eventually replacing computing equipment also involves environmental impacts.

The source of electricity is particularly important. If a data center uses electricity generated from renewable or low-carbon sources, the associated operational emissions can be lower than when electricity comes primarily from fossil fuels.

Why Does It Matter?

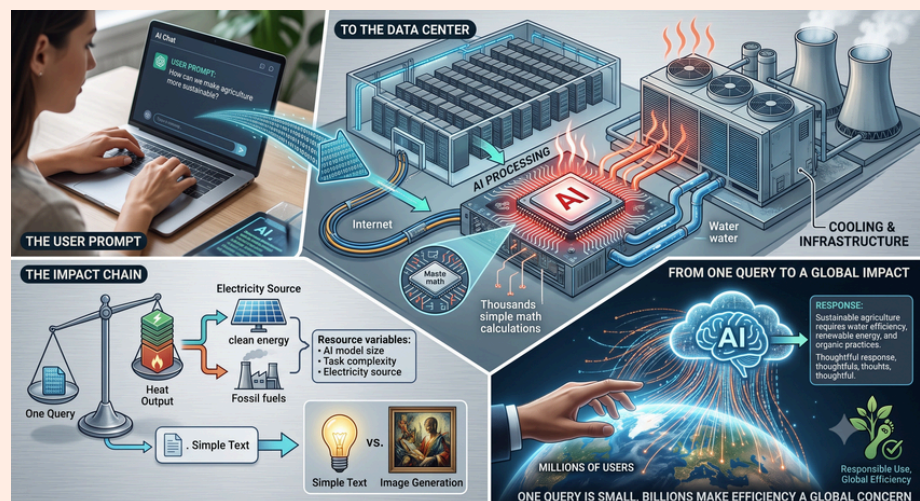
One AI query may have a relatively small carbon footprint, but AI services can process huge numbers of requests at global scale. As AI adoption grows, improving energy efficiency, using cleaner electricity, and developing more efficient models can help limit its environmental impact.

AI may be digital, but the computing power behind it has a physical footprint. Understanding that footprint is the first step toward building more sustainable AI.

WHY ONE QUERY CAN HAVE AN ENVIRONMENTAL COST

An AI query may seem like a simple digital action, but generating a response requires a chain of physical processes. When a user enters a prompt, it is sent to servers in a data center, where specialized processors perform calculations to understand the request and generate a response. These processors consume electricity and produce heat, requiring cooling systems to keep the equipment operating safely.

The environmental cost of a query depends on several factors, including the AI model being used, complexity and length of the request, computing hardware, data-center efficiency, and the electricity source. A short text query and a computationally intensive task such as image or video generation can have very different resource requirements.



From One Query to a Global Impact

User Prompt → Internet → Data Center → AI Processing → Cooling & Infrastructure → Response

Each step relies on physical infrastructure and energy. This is why digital interactions, although invisible to users, can still have an environmental footprint.

The goal is not to avoid AI, but to use it efficiently and responsibly while the technology industry works toward more energy-efficient models, hardware, and data centers.

One query may be small, but billions of queries can make efficiency a global concern.

TRAINING VS. USING AI: WHERE IS MORE ENERGY USED?

Training vs. Using AI: Where Is More Energy Used?

The energy footprint of AI comes from more than just answering user queries. Two major stages contribute to AI's computational demand: training and inference, or using a trained model.

Training an AI Model

Training is the process through which an AI model learns patterns from large datasets. It can involve running powerful GPUs or other specialized processors for extended periods. Large-scale models may require substantial computing resources during this stage.

Training also involves supporting infrastructure such as data storage, networking, and cooling. Although training happens less frequently than individual user interactions, it can involve a very large computational workload.

Using AI: Inference

Once a model has been trained, it can be used to answer prompts. This process is called inference.

Every time a user asks a chatbot a question, generates an image, summarizes a document, or uses an AI-powered application, the model performs inference. Each request generally requires less computation than training the model, but inference happens repeatedly—potentially millions or billions of times.

So, Which Uses More Energy?

There is no single answer for every AI system. Training can require a very large amount of energy over a concentrated period, while inference creates an ongoing energy demand as people continuously use the model. The overall footprint depends on factors such as model size, hardware, number and complexity of queries, and data-center efficiency.

The Bigger Picture

Understanding both stages is important when considering AI's environmental impact. Improving training efficiency, developing smaller models, optimizing inference, and using energy-efficient hardware can all help reduce the resources required by AI.

Training teaches the AI how to work; inference allows it to work for us. Both stages have an energy cost.

THE ROLE OF COOLING AND COMPUTING INFRASTRUCTURE

The Role of Cooling and Computing Infrastructure

AI systems depend on powerful computing infrastructure to process large amounts of data and generate responses. At the heart of this infrastructure are GPUs, CPUs, memory, storage systems, and networking equipment. These components work together to perform the enormous number of calculations required by modern AI models.

How Are AI Data Centers Cooled?

Modern data centers use different cooling technologies, including:

Air Cooling: Fans and air-conditioning systems circulate cool air around servers.

Liquid Cooling: Liquid absorbs heat more efficiently and can be used for high-performance AI hardware.

Immersion Cooling: Some systems place computing components in specially designed cooling fluids.

Advanced Thermal Management: Sensors and automated systems continuously monitor temperature and adjust cooling requirements.

Cooling itself requires energy, meaning that the environmental footprint of AI is not limited to the electricity consumed directly by processors.

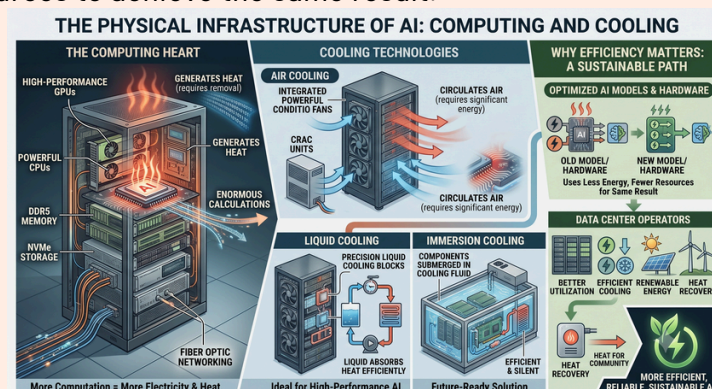
Why Computing Efficiency Matters

More efficient processors can perform the same amount of computation using less energy. Similarly, optimized AI models can reduce the computational resources required to produce a response.

Data-center operators can also improve sustainability through better server utilization, efficient cooling, renewable energy, heat recovery, and intelligent energy management.

As AI continues to grow, improvements in both computing hardware and cooling infrastructure will play an important role in making AI systems more efficient, reliable, and environmentally sustainable.

Efficient computing is not just about faster AI—it is also about using less energy and fewer resources to achieve the same result.



BILLIONS OF QUERIES: WHEN SMALL COSTS ADD UP

At an individual level, AI might seem to be using less energy than a human. The problem is that with millions or billions of requests, the cost of energy per query becomes significant, even if it is small.

Why Billions Matter

The amount of energy required by an AI to answer an individual query may be negligible, but this does not hold true if you consider the millions or even billions of queries that AI systems process. This means that computing, memory, networking, and cooling will require energy from the infrastructure of the data centre, and this minor cost of energy gets magnified through the sheer numbers of users. With the rising application of AI in search, education, software engineering, business, and entertainment, the number of queries keeps growing, thereby increasing the need for energy efficiency at the individual query level

The Scale Effect

Computing resources are activated every time an AI model is used to create text, analyze an image, write code, or reason. The energy needed can vary based on the hardware, prompt length, complexity of the task, and the model itself. The amount of energy consumed by various AI workloads can significantly differ, according to research estimates.

Small cost × huge number of queries = large total demand

For instance, a study conducted in the recent past found out that the median amount of energy used is about 0.34 Wh per query for some frontier-scale models with realistic deployment assumptions. In case of 1 billion queries, this would be equivalent to 0.34 GWh of energy use. Some difficult reasoning tasks can consume more energy per request.

Not Every Query Costs the Same

However, a basic question regarding text and reasoning cannot have the same energy needs. According to the IEA, there is a rapid decrease in the amount of energy consumed by AI applications due to advancements in hardware and software. More energy-hungry applications such as reasoning, video creation, and agentic AI are on the rise.



MAKING AI MORE ENERGY-EFFICIENT AND SUSTAINABLE



With growing size and usage of AI models, it has become increasingly important to optimize energy efficiency. Given that training and inference on large-scale AI models takes a lot of computation, even minor efficiencies can be very impactful at scale. Scientists are therefore looking into ways to increase performance and decrease the amount of electricity and computing power used in the process.

Smaller and more energy-efficient AI models represent one such way. Methods like model compression, quantization, pruning, and knowledge distillation allow achieving similar results with fewer computations and less energy spent on both training and inference on such models, especially when the model is used by millions of people.

Hardware efficiency is another very important factor to consider. Modern GPUs and AI accelerators are built specifically for performing large amounts of computations more efficiently than older hardware. Improvements in memory design, processing design, and workload scheduling can help cut down the amount of energy used for AI computations even more.

The infrastructure of data centers is an important part of the energy consumption puzzle too. AI servers need constant cooling and good power management in order to work. Using better cooling systems and making good use of server capacity and workload allocation can help save energy.

Sustainability of AI, finally, is dependent on responsible use. The selection of an adequate-sized model for a specific purpose, the reduction of unnecessary computing power, and efficiency in repeating tasks may reduce resource utilization without sacrificing any real performance. In the face of the continued expansion of AI into areas of education, health care, business, research, and daily application, sustainability will become increasingly important.

TOWARDS A GREENER FUTURE FOR ARTIFICIAL INTELLIGENCE



The goal is not to slow down the progress of Artificial Intelligence, but to make that progress more efficient and sustainable. By combining efficient models, smarter hardware, cleaner energy, optimized infrastructure, and responsible usage, AI can continue to evolve while moving towards a greener technological future.

Artificial intelligence (AI) is becoming ubiquitous in fields ranging from research and education to medicine and industries to our everyday life. The more AI systems are used, the higher their demand becomes on computing power, electrical power, water for cooling purposes, and even data centers. Therefore, in order to become sustainable, it is vital to combine the progress in technology with the sustainable use of natural resources.

The key part of this task is creating more energy-efficient AI models. Instead of endlessly increasing the size of AI models, scientists create small models that perform well and consume fewer resources.

Hardware improvements may aid in accomplishing this goal. More energy-efficient graphic processing units (GPUs), AI accelerators, and more powerful memory architecture will make it possible to conduct complex calculations using less energy. Proper scheduling of workload will help to ensure that the computational resources are not wasted and sit idle while still consuming energy.

Moreover, the infrastructure supporting the operation of AI should also become more sustainable. Cooling, power usage efficiency (PUE), server efficiency, and workload allocation may be used to lessen environmental damage caused by the operations of data centers. Usage of clean energy like solar, wind, or hydropower may be used to reduce the emissions from the energy consumed by AI.

Another important aspect of sustainability is the proper usage of AI technology. Different tasks require different amounts of computational resources, and running a big model for a small task will lead to waste of resources. Choosing the right model, limiting the number of requests, caching the results of the request and optimizing applications may help to lower the energy footprint of AI.

REFERENCES

- International Energy Agency (IEA) – Energy and AI
- A comprehensive report on AI, data centers, electricity consumption, energy efficiency, and emissions.
- IEA – Energy and AI
- International Energy Agency (IEA) – Energy Demand from AI
- Provides information about data-center infrastructure, servers, GPUs, cooling, and global electricity demand.
- IEA – Energy Demand from AI
- International Energy Agency (IEA) – AI and Climate Change
- Discusses the relationship between data-center electricity consumption and CO₂ emissions.
- IEA – AI and Climate Change
- Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. ACL.
- Academic research examining the energy and environmental costs associated with training deep-learning models.
- ACL Anthology – Energy and Policy Considerations for Deep Learning in NLP
- Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. Communications of the ACM.
- Discusses improving AI efficiency and reducing the environmental costs of increasingly computationally intensive AI systems.

STUDENT EDITORS



Avantika Chatterjee
TY IT-A



Dev Padhariya
TY IT-D