

# The Environmental Impacts of AI - Policy Primer

Sasha Luccioni, Bruna Trevelin, Margaret Mitchell (Hugging Face)



Image source: <https://betterimagesofai.org/>

## Executive Summary

Artificial intelligence (AI) has an environmental cost. Beginning with the extraction of raw materials and the manufacturing of AI infrastructure, and culminating in real-time interactions with users, every aspect of the AI lifecycle consumes natural resources – energy, water, and minerals – and releases greenhouse gases. The amount of energy needed to power AI now outpaces what renewable energy sources can provide, and the rapidly increasing usage of AI portends significant environmental consequences. The goal of this primer is to shed light on the environmental impacts of the full AI lifecycle, describing which kinds of impacts are at play when, and why they matter.

While some research and documentation on AI's environmental impacts currently exists, the nature and extent of AI's effects are under-documented, ranging from its embodied and enabled emissions to rebound effects due to its increased usage. Regulatory and policy initiatives, both existing and in progress, have the challenge of encouraging innovation and growth while addressing environmental impacts and how they affect different stakeholders. Ways forward range from technical interventions to make AI models more efficient, to policy interventions to incentivize sustainable AI research and practice.

## Introduction

Recent years have ushered in a new era of growth and adoption of Artificial Intelligence (AI) technologies, especially generative AI, which is increasingly used in tools and systems spanning from Web search to customer service. While this era brings with it potential gains in terms of profit and productivity, its impact on already strained natural resources is significant, and cannot be overlooked. The goal of the present primer is to outline fundamental pieces

---

underpinning AI's direct impacts<sup>1</sup> on the environment. It is written to be legible for the general public, policymakers, and the AI community. As such, it is organized to detail which natural resources are used throughout the AI lifecycle, and what the effects are on the environment for each.

## Existing Research

Initial work on the environmental impacts of AI models focused on estimating the CO<sub>2</sub> emissions incurred during **model training**. This includes the seminal work of Strubell et al.<sup>2</sup>, which calculated that training a **large language model (LLM)** with 213 million parameters was responsible for 626,155 pounds of CO<sub>2</sub> emissions, roughly equivalent to the lifetime emissions of five cars, including fuel. Follow-up studies have looked at other types of **model architectures**, their energy use and emissions<sup>3,4</sup>, confirming the consequential environmental impact of AI training.

Luccioni et al.<sup>5</sup> proposed to extend these analyses to consider other stages of the AI model lifecycle, including training, deployment, and the components that make these possible: material extraction, equipment manufacturing, and overhead energy usage (cooling, networking, storage), among others (see Fig 1 below). Their work demonstrated that estimates of AI's **greenhouse gas emissions** are more than double than previous studies that focused on training in isolation, revealing a fuller picture of AI's environmental impact.

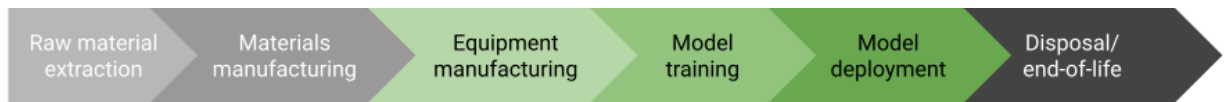


Fig 1. The life cycle assessment approach proposed by [Luccioni et al.](#)

Luccioni et al. also provided the first study on the environmental impact of **AI model deployment** by analyzing the energy usage and carbon emissions of running the BLOOM large language model (LLM) in the cloud – specifically, a Google Cloud Compute cloud instance that contained the 176 Billion parameter BLOOM model, which received 230,768 queries over a period of 18 days. They found that this deployment used an average of 40.32 kWh of energy per day (roughly the equivalent of 1,110 smartphone charges) and, given the energy mix used

---

<sup>1</sup> Kaack et al. (2022). [Aligning artificial intelligence with climate change mitigation](#). Nature Climate Change (Vol. 12, 518–527).

<sup>2</sup> Strubell, E., Ganesh, A., & McCallum, A. (2020, April). [Energy and policy considerations for modern deep learning research](#). In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 09, pp. 13693-13696).

<sup>3</sup> Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L. M., Rothchild, D., ... & Dean, J. (2021). [Carbon emissions and large neural network training](#). arXiv preprint arXiv:2104.10350.

<sup>4</sup> Naidu, R., Diddee, H., Mulay, A., Vardhan, A., Ramesh, K., & Zamzam, A. (2021). [Towards quantifying the carbon emissions of differentially private machine learning](#), ICML 2021 SRML workshop.

<sup>5</sup> Luccioni, A. S., Viguier, S., & Ligozat, A. L. (2023). [Estimating the carbon footprint of BLOOM, a 176B parameter language model](#). Journal of Machine Learning Research, 24(253), 1-15.

---

by the computing instance, emitted approximately 19 kgs (42 lbs) of CO<sub>2eq</sub> per day, reflective of state-of-the-art AI model deployment in 2023. This study was followed up by Luccioni et al. in 2024<sup>6</sup>, which looked at the variation in energy usage and carbon emissions across different types of AI tasks. They found distinct differences based on modality (text vs. image) as well as whether the model was creating new content (captions, summaries, etc., commonly referred as “**generative AI**”) or returning existing content. They found that image-based tasks and those that create new content used more energy.

## AI’s Environmental Impacts

It can be hard to understand the extent of AI’s impacts on the environment given the separation between where you interact with an AI system, and how that interaction has come to be – most AI models run on data centers that are physically located far away from their users, who only interact with their outputs. But the reality is that AI’s impressive capabilities come with a substantial cost in terms of natural resources, including energy, water and minerals, and non-negligible quantities of greenhouse gas emissions.

### Energy

#### How does AI use energy?

Every time we query an AI model – be it via our phone, a chat interface, or a smart speaker – this request has to run through physical hardware to provide us with an answer. As AI models get bigger in size and complexity, they also require more powerful hardware to run them. One common hardware component for AI systems is the **GPU (graphical processing unit)**, and an individual AI model may require multiple GPUs. These GPUs are often on servers within data centers located across different regions of the world and connected via the Web. On average, 40-50% of the energy used by a data center is used for powering the computing equipment, with a further 30%–40% dedicated to cooling the equipment. Energy is also used for manufacturing the hardware (e.g., GPUs), as well as other elements of data center infrastructure (e.g., storage, networking, etc.) – although the exact amount used by companies such as Nvidia, who have considerable market shares on GPU design, remains unknown.

#### How much energy do specific AI models use?

There is currently little transparency on the energy demands of specific AI applications, although a recent estimate has put the amount of energy used for a ChatGPT query to be anywhere between 6-10 times more than a traditional Web search (0.3 Wh vs. 2.9 Wh) [6] <sup>7</sup>. At a macro level, AI is estimated to use 10%–20% of data center electricity today<sup>8</sup>, but as new

---

<sup>6</sup> Luccioni, S., Jernite, Y., & Strubell, E. (2024, June). Power Hungry Processing: Watts driving the cost of AI deployment?. In The 2024 ACM FAccT Conference (pp. 85-99).

<sup>7</sup> Goldman Sachs (2024)- “[AI, data centers and the coming US power demand surge](#)”

<sup>8</sup> EPRI (2024) - “[Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption](#)”

---

generations of AI-enabled servers consume more power<sup>9</sup>, this percentage is set to increase at an average of 70% in coming years<sup>10</sup> and double by 2030<sup>11</sup>.

### Where does the energy used for AI come from?

Data centers currently account for 2-3% of the total electricity use in the United States, and this number is set to triple in the next 5 years<sup>12</sup>. While new sources of renewable energy, such as solar and wind energy, are projected to meet approximately 40% of this demand, the rest is set to be met with non-renewable energy sources such as coal and natural gas. This would result in as much additional greenhouse gas as 16 million gas-powered cars<sup>13</sup>. The combined energy use of the main technology companies providing AI cloud computing services and products (such as Google, Microsoft, Meta, and Amazon) has more than doubled in the last five years, and these companies are also among the largest purchasers of corporate renewable energy in the United States, having purchased almost 50 GW of renewable energy to date, as much as the generation capacity of Sweden<sup>14</sup>. These same companies have also put out ambitious net-zero goals but have recently announced that they are failing to meet them due in part to the energy demands of AI tools and services<sup>15,16</sup>. Overall, the growing energy demand for AI is significantly outpacing the increase in renewable energies – entailing substantial new GHG emissions and squeezing an already tight renewable energy market.

## Water

### Why does using AI require water?

Just as the components in your personal computer heat up as they're used, so too do the components in computer servers. Data centers used for AI systems house hundreds of thousands of these servers, carrying out intensive, round-the-clock computation. They therefore need constant cooling to avoid overheating. One of the key ways in which this is done is by pumping clean water through radiators in the data center, which absorbs the heat in the air. The water used for cooling data centers cannot be saltwater or greywater because it clog or corrode cooling systems. This water then has to be cooled down; a significant portion of it evaporates in the process<sup>17</sup>. After cooling, the remaining water is then reused or cleaned and cooled before being discharged back into aquifers.

---

<sup>9</sup> Goldman Sachs (2024)- [“AI, data centers and the coming US power demand surge”](#)

<sup>10</sup> Morgan Stanley (2024)- [“Powering the AI Revolution”](#)

<sup>11</sup> Goldman Sachs (2024)- [“AI, data centers and the coming US power demand surge”](#)

<sup>12</sup> Washington Post (2024)- [“AI is exhausting the power grid. Tech firms are seeking a miracle solution”](#)

<sup>13</sup> Goldman Sachs (2024)- [“AI, data centers and the coming US power demand surge”](#)

<sup>14</sup> IEA (2023)- [“Data Centres and Data Transmission Networks”](#)

<sup>15</sup> Bloomberg News (2024)- [“Microsoft’s AI Push Imperils Climate Goal as Carbon Emissions Jump 30%”](#)

<sup>16</sup> Bloomberg News (2024)- [“Google’s Emissions Shot Up 48% Over Five Years Due to AI”](#)

<sup>17</sup> Reig (2013) - [What’s the difference between water use and water consumption?](#) World Resources Institute Commentary.

---

While data centers are a primary source of water usage in the AI lifecycle, water is also used in other areas of the AI lifecycle. This includes the hardware manufacturing process, for rinsing the different layers of semiconductor wafers that form the CPU and GPU chips. This water has to go through multiple intensive cycles of filtration, sterilization and purification in order to remove all the impurities, which can damage the wafers<sup>18</sup>. Water can also be used for the generation of electricity used for powering data centers and hardware manufacturing facilities - although less than 15% of total electricity generation globally comes from hydroelectric power.<sup>19</sup>

## How much water does AI use?

Generally speaking, the amount of water used for data center cooling differs depending on data center configuration, but can range from 0.18 to 1.1L of water per kWh of energy<sup>20,21</sup>. There are no official figures for specific AI models, but third-party research has estimated that GPT-3, the AI model underpinning the popular ChatGPT, uses 500mL of water for every 10 to 50 queries<sup>22</sup>, although this number would depend highly on where it is deployed and how efficient the data center is. An average hyperscale data center, such as those used for training AI models, uses around 550,000 gallons (2.1 million liters) of water daily<sup>23</sup>. Notably, some of them are built in areas with little water supply, such as the Arizona desert, where electricity is available, but water is scarce.

In terms of manufacturing, there are no exact numbers for AI-specific accelerators such as GPUs, but the Taiwan Semiconductor Manufacturing Company (TSMC), the largest manufacturer of the semiconductor wafers used in GPUs, uses about 157,000 tons of water a day<sup>24</sup>, a figure that increased by 70% between 2015 and 2019<sup>25</sup>.

## Minerals

### What kind of minerals are used in the AI supply chain?

Computing chips such as GPUs are built on a thin layer of semiconductor, usually silicon, upon which components made out of different types of metals such as aluminum, copper, tin, tantalum, lithium, gallium, germanium, palladium, cobalt and tungsten are added<sup>26</sup>. Mining of these metals also comes with environmental costs, since hundreds of tonnes of ore typically need to be dug up and processed to get a single ton of relatively common metals such as

---

<sup>18</sup> Brito, Griffin and Koski (2022) - "[Nvidia GPU — Design Life-Cycle](#)"

<sup>19</sup> Ember (2023) - "[Global Electricity Review 2023](#)"

<sup>20</sup> Microsoft (2022) [How Microsoft measures datacenter water and energy use to improve Azure Cloud sustainability](#)

<sup>21</sup> Amazon (2022) [Water Stewardship](#)

<sup>22</sup> Li et al. (2023) - [Making AI Less 'Thirsty': Uncovering and Addressing the Secret Water Footprint of AI Models](#), arxiv preprint 2304.03271

<sup>23</sup> DgtlInfra (2024) - "[Data Center Water Usage: A Comprehensive Guide](#)"

<sup>24</sup> Brito, Griffin and Koski (2022) "[Nvidia GPU — Design Life-Cycle](#)"

<sup>25</sup> Forbes (2021) "[No Water No Microchips: What Is Happening In Taiwan?](#)"

<sup>26</sup> Brito, Griffin and Koski (2022) "[Nvidia GPU — Design Life-Cycle](#)"

copper or aluminum<sup>27</sup>. Minerals such as cobalt and tungsten are considered to be ‘conflict minerals’, meaning that they are mined or traded in areas of conflict, and contribute towards perpetuating human rights abuses and armed conflict<sup>28</sup>.

**Greenhouse gas emissions**

**How else does AI impact the environment?**

The usage of water, minerals, and energy throughout the AI lifecycle often comes with emissions of greenhouse gasses – such as carbon dioxide. One major source is from the production of concrete and steel used in data centers; worldwide, concrete production is estimated to generate up to 8 percent of all human **CO<sub>2</sub> emissions**<sup>29</sup>, in addition to using water and minerals. Another primary source of greenhouse gas emissions is from generating the electricity needed in different stages of the AI lifecycle. In Luccioni et al.’s 2023 study<sup>30</sup> of the BLOOM language model, they found that of the total 50 tonnes of CO<sub>2eq</sub> emissions emitted during model training, only half was due to the energy consumption of the GPUs used for training BLOOM (‘dynamic consumption’), with a further 29% stemming from the idle consumption of the data center (i.e., the energy used for heating/cooling, networking, storage, etc.), and a final 22% of emissions were produced from the GPU manufacturing process. The carbon intensity of the grid used is therefore the main factor influencing the final quantity of emissions incurred by AI models, impacting both model training and deployment.

Process	CO2 emissions	Percentage of total
Equipment manufacturing	11.2 tonnes	22.2%
Dynamic energy consumption	24.69 tonnes	48.9%
Idle energy consumption	14.6 tonnes	28.9%

Table 1: Breakdown of CO<sub>2</sub> emissions for the BLOOM model (Adapted from [Luccioni et al., 2023](#))

**Missing and Partial Information**

For a full picture of AI’s environmental impact, we need both consensus on what to consider as part of “AI”, and much more transparency and disclosures from the companies involved in creating it. AI refers to a broad set of techniques, including machine learning, but also rule-based systems. A common point of contention is the scoping of what constitutes AI and what to include when estimating its environmental impacts. Core to this challenge is the fact

<sup>27</sup> Mills (2020) - “[Mines, Minerals, and "Green" Energy: A Reality Check](#)”  
<sup>28</sup> Euromines (2020) - “[The Electronics Value Chain and Its Raw Materials](#)”  
<sup>29</sup> Mohamad et al. (2022) [Environmental impact of cement production and Solutions: A review](#)  
<sup>30</sup> Luccioni et al. (2023) - “[Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model](#)”

---

that AI is often a part of, as opposed to the entirety of, any given system – e.g. smart devices, autonomous vehicles, recommender systems, Web search, etc. How to delineate and quantify the environmental impacts of AI as a field is therefore a topic of much debate, and there is currently no agreed-upon definition of the scope of AI. Some notable areas lacking coherent information include:

- **The embodied emissions of hardware manufacturing:** While there are some numbers about the embodied (i.e., supply chain) emissions of different types of computing hardware<sup>31,32</sup>, there are currently no reliable numbers on the carbon cost of manufacturing GPUs, which are the most popular type of hardware used for training AI models. Gathering these numbers for different generations and types of GPUs is important to have a better understanding of AI's lifecycle.
- **The embodied carbon of AI infrastructure:** A significant contributing factor in the carbon emissions of data centers is the use of concrete in their construction.<sup>33</sup> While initiatives have been aimed at providing sustainable building materials for data center construction,<sup>34</sup> <sup>35</sup> the relative impact of concrete construction on AI emissions is unclear.
- **Rebound effects and unintended consequences:** It has been observed in different sectors (such as transportation) that an increase in efficiency in resource use will generate an increase in resource consumption rather than a decrease. While there is currently emphasis placed on making both AI hardware and models more efficient, the rebound effect that this may have (e.g. by increasing overall resource use) is still unclear and needs further quantification and longitudinal tracking.
- **AI's enabled emissions:** AI technologies are increasingly used in sectors such as oil and gas exploration, improving their efficiency and increasing yield, which results in increased emissions overall<sup>36</sup>. The proportion of these emissions that can be attributed to AI is still unclear, but initial work<sup>37</sup> is being done to better understand this topic.
- **Ecosystem harm:** Industrial waste produced by AI structures such as data centers risk environmental contamination<sup>38</sup>. Further, the construction of new buildings – data centers as well as buildings where AI development occurs – risks negative effects on local ecosystems and existing natural habitats.<sup>39</sup>

## Legislation/Regulation

---

<sup>31</sup> Dell (2019) - "[PowerEdge R240 Carbon Footprint](#)"

<sup>32</sup> Apple (2019) - "[Mac Pro Product Environmental Report](#)"

<sup>33</sup> Gensier (2023) - "[Designing for Lower Carbon Concrete in Data Center Constructions](#)"

<sup>34</sup> The Verge (2023) - "[Microsoft is testing low-carbon concrete for its data centers](#)"

<sup>35</sup> AWS (2023) - "[How AWS is using more lower-carbon materials to build data centers](#)"

<sup>36</sup> Greenpeace (2020) - "[Oil in the Cloud](#)"

<sup>37</sup> Grist (2024) - "[Microsoft employees spent years fighting the tech giant's oil ties. Now, they're speaking out](#)"

<sup>38</sup> Rest of World (2024) - "[Microsoft is building a data center in a tiny Indian village. Locals allege it's dumping industrial waste](#)"

<sup>39</sup> Balova and Kolbas (2023) - "[Biodiversity and Data Centers: What's the connection?](#)"

---

Different regulations may deal with multiple elements employed in the AI supply chain (e.g. regulations regarding mining, water, labor, hardware, software, data), but more specific regulation addressing AI and the environment is emerging. As discussed throughout this primer, the discussion of what constitutes AI is open and the same goes to how AI and the environment should be dealt with by regulation. Below is a non-exhaustive list of initiatives regarding AI and environmental challenges.

## European Union

The EU launched the European Green Deal, a set of proposals to make the EU fit for reducing net greenhouse gas emissions, and to which EU's actions and policies have to contribute. As part of the Green Deal, the EU Climate Law has set legally binding targets for carbon neutrality by 2050, and AI systems used in energy management, smart grids, and environmental monitoring will need to comply with this law. A related initiative is the delegated regulation to establish an EU-wide scheme to rate the sustainability of EU data centers, as part of the new Energy Efficiency Directive<sup>40</sup>. Water consumption could be addressed and regulated, for data centers built in the EU, through a thorough application of the Water Framework Directive.

Environmental protection is also stated as being one of the core values put forward by the EU AI Act<sup>41</sup>, and appears several times in its text. As provided in the AI Act, the energy consumption of AI models is at the core of this topic, and is stated as one of the criteria that must be taken into consideration when training and deploying them. The AI Act stipulates that the providers of general-purpose AI models (GPAIs) specifically should share the known or estimated energy consumption of their models. It also provides that high-risk AI systems should report on resource performance, such as consumption of energy and of "other resources" during the AI systems' life cycle, which could include water and minerals depending on the level of detail of the standards that will guide compliance to this reporting obligation.<sup>42</sup>

The text of the Act also encourages the adoption of voluntary codes of conduct that would include measures to evaluate and minimize the environmental impact of AI systems. However, given that the adoption of codes of conduct is non-binding, it is unclear how effective it will be in achieving the stated objectives.

## United States

The Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence<sup>43</sup> published by the White House in October 2023, while extensive in other regards,

---

<sup>40</sup> [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AJOL\\_2023\\_231\\_R\\_0001](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AJOL_2023_231_R_0001)

<sup>41</sup> [https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L\\_202401689](https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689)

<sup>42</sup> Obligations regarding risk management can also be said to address environmental concerns in the AI Act. See Philip Hacker's blog post on The existential threat of AI at <https://blog.oup.com/2024/08/the-real-existential-threat-of-ai/>

<sup>43</sup> <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>



---

does not directly address the environmental impacts of AI models, although it does mention the development of climate-positive applications of AI (“models that streamline permitting and environmental reviews while improving environmental and social outcomes”) in Section 5.3. A few months after the Executive Order, Senators Markey and Heinrich and Reps. Eshoo and Beyer introduced the Artificial Intelligence Environmental Impacts Act of 2024<sup>44</sup>, a piece of legislation that proposes to carry out a comprehensive study on the environmental impacts of AI and develop a voluntary reporting system for its reporting. While the Bill is currently under review, it would be the first piece of legislation that directly addresses this topic. The FTC has issued guidelines related to the fairness and transparency of AI systems, which could be argued to extend to AI's environmental impact, at least in the sense that AI systems must not mislead consumers about sustainability claims.

### Other relevant initiatives

**Spain:** Sustainability was a key part of Spain's recently announced National AI strategy<sup>45</sup>, with a specific emphasis on investing in computationally efficient models and data centers powered by renewable energy sources.

**Canada:** Environmental sustainability is one of the key focuses of the Pan-Canadian Artificial Intelligence Strategy<sup>46</sup>. According to this strategy, by 2030 Canada intends to have a robust national AI ecosystem founded on scientific excellence, training and talent pools, public-private collaboration and advancing AI technologies to bring positive social, economic and environmental benefits for people and the planet. However, Canada's Artificial Intelligence and Data Act does not have specific provisions regarding the environmental impacts of AI.

**France:** The AI and Green transition roadmap<sup>47</sup>, part of the broader France Nation Verte ecological planning program, highlights the potential of data, and more specifically AI, to meet the five major challenges of the ecological transition: consumption of resources, preserving biodiversity, mitigating and adapting to global warming and reducing pollution that impacts health.

**International instruments:** The OECD Recommendation of the Council on Artificial Intelligence<sup>48</sup> has as one of its key principles the pursuit of inclusive growth, well-being, sustainable development and environmental sustainability with AI. The Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI Systems<sup>49</sup> proposes a non-exhaustive list of guiding principles, including prioritizing the development of AI systems

---

<sup>44</sup> <https://www.congress.gov/bill/118th-congress/senate-bill/3732/text>

<sup>45</sup> <https://portal.mineco.gob.es/RecursosArticulo/mineco/ministerio/ficheros/National-Strategy-on-AI.pdf>

<sup>46</sup> <https://ised-isde.canada.ca/site/ai-strategy/en>

<sup>47</sup> <https://www.ecologie.gouv.fr/politiques-publiques/feuille-route-intelligence-artificielle-transition-ecologique>

<sup>48</sup> <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>

<sup>49</sup> <https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-guiding-principles-advanced-ai-system>

---

to address the world's greatest challenges, notably but not limited to the climate crisis. The United Nations Global Digital Compact<sup>50</sup> includes a commitment to promote environmental sustainability across the life cycle of digital technologies and aim to ensure that digital infrastructure and equipment are sustainably designed for the mitigation of and adaptation to climate change.

## Ways forward

There is no single, one-size-fits-all approach to reducing the environmental impacts of AI models, but a variety of technical, behavioral and organizational interventions can be adopted by different stakeholders, at different stages of the AI lifecycle:

- **Technical interventions:** different approaches are possible to make AI models less resource-intensive. This includes techniques such as pruning, quantization, distillation, and flash attention. These approaches can be applied at various stages of development, both during model training to make it more compute-friendly as well as on already-trained models to make them more efficient.
- **Behavioral interventions:** users of AI models can also contribute to minimizing AI's environmental impacts by choosing task-specific (as opposed to multi-task) models when possible. AI model developers can benchmark the energy consumption of different models and opt for the most efficient ones, for instance, based on metrics such as the [AI Energy Star project](#), and adopting strategies such as [flexible scheduling](#) for training AI models, which would allow for the optimization of energy sources.
- **Organizational interventions:** institutions can also implement best practices with potentially far-reaching impacts in terms of the environment. This can include opting for compute instances powered by renewable energy sources when possible, measuring the energy consumption and greenhouse gas emissions of AI-enabled products and providing these metrics to users of these products.
- **Policy interventions:** enforce transparency for user-facing AI systems, and transparency in general regarding the environmental impact of AI systems, start regulating different systems based on how often they're used, promote standards and incentives for meeting these standards (such as tax rebates), boost innovation through funding, incentivize sustainable applications and market opportunities and prioritize the most impactful solutions, including based on informed analysis of policy efforts.<sup>51</sup>

---

<sup>50</sup> <https://www.un.org/techenvoy/global-digital-compact>

<sup>51</sup> See for example Climate policies that achieved major emission reductions: Global evidence from two decades, at <https://www.science.org/doi/pdf/10.1126/science.adl6547>

---

## Glossary

**AI model deployment:** The process of using an AI model to carry out the task that it was trained for - e.g. generating images, finding answers to questions, etc.

**AI model training:** The process of providing data to an AI model that is meant to enable it to produce more accurate predictions.

**CO<sub>2</sub> equivalents (CO<sub>2eq</sub>):** Since different greenhouse gases that are generated during electricity generation have different global warming potentials, these are commonly reduced to a common denominator, that of carbon dioxide, in order to make comparisons easier.

**Embodied emissions:** Emissions associated with the whole lifecycle of an item, including production of materials, construction, transportation, and item usage.

**Generative tasks:** Tasks that require a model to generate new outputs based on inputs - e.g., generating a poem based on an image, or generating an image based on a text prompt.

**Graphical Processing Unit (GPU):** a type of computing hardware originally designed to accelerate the rendering of graphics in video games given their parallel processing capabilities. They are now the most commonly-used type of hardware for training and deploying AI models.

**Greenhouse gases (GHGs):** gases that exist in the Earth's atmosphere that trap heat within it, thereby contributing towards the greenhouse effect (i.e. raising the surface temperature), e.g., carbon dioxide, methane, nitrous oxide.

**Large language model (LLM):** there is no single, agreed-upon definition of large language models (also called 'frontier models' or 'foundation models'), but they are largely defined as computational models that are capable of taking natural language as input or producing it as output, sometimes alongside other modalities such as images. See [Rogers and Luccioni \(2024\)](#) for a more in-depth discussion.

**Model architectures:** Depending on the task at hand and the amount of data available, there are different architectures of AI models that can be used – ranging from simpler ones such as decision trees to more complex ones such as transformers and convolutional neural networks. See [Goodfellow et al \(2016\)](#) for more details.

## Acknowledgments

Thank you to Brigitte Tousignant for her help in editing this primer, and Philipp Hacker and Lynn Kaack for their invaluable comments and suggestions.