Balancing GreenAI: Efficiency and Impact of LLMs in Climate-Vulnerable Communities

Anonymous ACL submission

Abstract

Large language models (LLMs) are revolutionizing natural language processing (NLP), but their creation comes at a significant environmental cost. This research investigates the carbon emissions produced by pre-training BERT-based language models and situates these findings within the broader context of global carbon emissions. Climate change disproportionately affects low- and middle-income countries (LMICs), so we weigh LLMs' impact within the context of these disadvantaged communities. We explore methodologies for estimating the carbon footprint of model training. We contemplate trade-offs between model efficiency and potential bias, considering how such side effects could exacerbate existing inequalities, particularly in LMICs. Furthermore, this research emphasizes the necessity of transparency and accountability in NLP. LLMs should be developed with clear purposes and a focus on both efficiency and mitigating potential harms, particularly in climate-vulnerable communities.

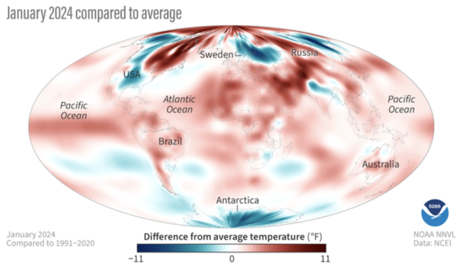

Figure 1: Global surface temperatures in January 2024 compared to the 1991-2020 average. Most areas exhibited warmer-than-average temperatures (depicted in red), with scattered colder-than-average regions (depicted in blue)[1].

1 Introduction

LLMs are gaining prominence in both NLP research and everyday use. However, creating these models comes with significant computational costs. Industry rivals continuously invest in developing LLMs, which incurs financial expenses and contributes to environmental degradation through carbon emissions. This environmental impact is often overlooked.

Global climate change is a crisis, and carbon emissions actively contribute to it. Consequently, when developing LLMs, researchers should think of both accuracy and efficiency, as the energy required to train models can be traced to greenhouse gas emissions.

This research aims to estimate the carbon emissions produced by pretraining widely used LLMs, and to ground these findings in the broader context of carbon emissions. We will determine how much the LLMs emit and provide insights into current industry practices and LLM applications to recommend action in the wake of climate change.

Alongside all other carbon-intensive industries, LLM development contributes to global climate change. We focus on how such climate consequences disproportionately harm historically disadvantaged communities worldwide. NLP's climate impact is felt globally, but its benefits do not follow suit.

2 Related Work

2.1 Green AI

The method for calculating model pretraining carbon emissions is derived from Strubell et al. (2019), where the authors reported the training carbon impacts of four LLMs: T2T, BERT, ELMo, and GPT-2. We train a different set of models. We follow the same methodology of querying for machine power usage, averaging such results, and computing carbon dioxide and monetary estimates. However, we diverge in the qualitative aspects of this research and seek greater context to such carbon

[1] Source:https://www.climate.gov/news-features/understanding-climate/global-climate-summary-january-2024

emission estimates, considering global systems for valuating emissions and the ultimate goals of training LLMs. We emphasize interdisciplinary evaluation on whether LLMs' carbon footprints are appropriately offset by their impact, specifically in healthcare applications in the communities most impacted by climate change.

Schwartz et al. (2019) discuss Green AI in the exponential growth of computational requirements for deep learning research and the resulting environmental and financial costs. The authors advocate for integrating efficiency, alongside accuracy and other relevant metrics as an evaluation criteria for research.

Kaack et al. (2022) discuss the Information and Communication Sector (ICT)'s environmental impact and emphasize the need for AI research to align with climate change mitigation efforts through transparent, comprehensive reporting on compute costs and emissions. They address concerns that efficiency gains in machine learning may lead to rebound effects, where reduced emissions per-run could be countered by increased consumption, undermining environmental benefits.

2.2 Emissions & Climate Change

Climate change has driven concerns about the impact of data centers and their ever-growing compute demands. A single data center can consume as much electricity, and emit subsequent carbon, as 50,000 homes (Monserrate, 2022). Siddik et al. (2021) calculate the carbon and water footprints of data centers in the U.S, referencing how U.S. data centers use 1.8% of national electricity and account for about 25% of global data centers.

In 2021, electricity and heat generation made up almost 44% of global carbon dioxide emissions from fuel combustion, and coal accounted for 73% of such emissions (IEA, 2023). Fossil fuels, such as coal, oil, and natural gas, are carbon-based, nonrenewable energy sources that release carbon into the atmosphere upon combustion (Eurostat).

The Paris Agreement, established in 2015, seeks to limit global warming to 1.5°C above pre-industrial levels to mitigate severe climate impacts (Nations). Climate change results in rising sea levels, intensified storms and droughts, and dwindling sea ice. These effects cascade through ecosystems and human activities, impacting health, agriculture, and infrastructure (NASA, 2024). To reach the Paris Agreement's goal, global emissions must decrease by 45% by 2030 and reach net-zero by 2050 (Nations). Carbon emissions must be reduced because they contribute to the greenhouse effect, which traps heat in the atmosphere and warms the planet (NRDC, 2023).

2.3 Racial Justice & Climate Change

Climate change affects the entire globe, but some regions are impacted more. Developing countries, or LMIC, often face the greatest climate-related risks and are the least equipped to manage such events.

For instance, Haiti, which lacks the infrastructure for adequate disaster preparedness, has faced great devastation from earthquakes and hurricanes, where the 2010 earthquake cost $8 billion– more than the country's GDP. The population is heavily reliant on agriculture, but as sea levels rise, saltwater may permeate farmland and fresh water, leaving a large portion of the country without a livelihood (Law, 2019).

Additionally, rising sea levels threaten Kiribati's existence, as the country's islands are only six feet above sea level. Rising sea levels may contaminate fresh water supplies and soil, and ocean acidification and rising ocean temperatures harm Kiribati's vital fishing industry (Law, 2019).

In the face of climate change, equitable climate resilience is necessary. Environmental destruction disproportionately harms certain racial, ethnic, and national groups. Historical abuse and colonialism have left these communities most vulnerable to and least equipped to handle such events (Achiume, 2022).

To consider human health specifically, Berberian et al. (2022) report that U.S. communities of color, including Black, LatinX, Native American, Pacific Islander, and Asian communities, face higher risks of climate-related health impacts than Whites. These disparities manifest in increased mortality, respiratory and cardiovascular diseases, mental health issues, and heat-related illnesses. Climate change exacerbates existing social and economic inequalities, compounding racial health disparities in marginalized communities.

2.4 AI & Healthcare in LMIC

Ciecierski-Holmes et al. (2022) write how artificial intelligence, including LLMs, has the capacity to improve health systems by supporting and standardizing clinical judgments and applying healthcare processes with an objective, data-oriented approach. This power would be particularly beneficial in LMIC, where medical care is often limited, by improving the efficiency of existing labor and reducing training needed when expanding the workforce (Weissglass, 2022). However, the effective deployment of AI in LMICs faces significant challenges. Ciecierski-Holmes et al. (2022) reference how training data from LMIC is limited and may not be accessible, current, or expansive enough. Further, in low-resource settings, AI tools are only effective if they can be used or integrated into the existing infrastructure. Yet, these low-resource settings may not have the means to support such tools; for instance, some have no or unstable internet access. AI's promise in LMIC healthcare hinges on addressing data gaps, infrastructure hurdles, and integration challenges.

2.5 Bias in Model Compression

Ramesh et al. (2023) investigate how model compression techniques such as pruning, quantization, and distillation affect bias in language models. It explores intrinsic and extrinsic measures of fairness evaluation in language models and identifies the impact of compression techniques on fairness.

Gonçalves and Strubell (2023) study the propagation of social biases and tradeoffs between bias reduction and efficiency in LLMs, highlighting the challenges of addressing biases in pretrained models due to the computational expense of retraining.

3 Data

We use a subset of the AllenAI C4 English dataset. The AllenAI C4 dataset is a cleaned version of Common Crawl's web crawl corpus and is commonly used for training NLP models due to its fast coverage and quality. The C4 English dataset is 305 GB and 364,868,892 entries, consisting of an entry's URL, text, and timestamp (for AI, 2024).

Due to memory constraints, we use a small subset of the dataset. As we will discuss later, the experiment focuses on the act of training, not the training outcomes themselves, as we limit pretraining time to 8 hours. The subset was preprocessed to use only the text with each sentence formatted to the models' criteria.

4 Models

Devlin et al. (2019)'s Bidirectional Encoder Representations from Transformers (BERT) leverages a deep multi-layered Transformer encoder with self-attention to analyze all words in a sentence bidirectionally, capturing richer context compared to unidirectional models. Pre-training with masked language modeling further strengthens BERT's ability to understand word relationships and meaning. The original authors trained BERT-base on 16 TPU chips for 4 days. In 2019, AWS reportedly trained BERT-base in 62 minutes on 2,048 32GB V100 GPUs (Bindal et al., 2019).

Sanh et al. (2020)'s DistilBERT inherits the core architecture of BERT but achieves a smaller size and faster inference by employing knowledge distillation. This technique trains a smaller "student" model (DistilBERT) to mimic the outputs of a larger, pre-trained "teacher" model (BERT). DistilBERT achieves this through a combination of a reduced number of layers and a specialized loss function that incorporates knowledge distillation alongside traditional language modeling objectives. The authors trained DistilBERT for 90 hours on 8 16GB V100 GPUs.

Lan et al. (2020)'s ALBERT builds upon BERT's architecture, maintaining strong performance, but reduces parameters and training time via two key modifications. First, ALBERT employs a factorized parameterization technique, which breaks down the self-attention mechanism into smaller, more manageable components. Second, ALBERT uses a cross-layer parameter sharing, where the weights of specific layers are shared across the entire encoder. AWS reportedly trained ALBERT-base in 20 hours on 64 32GB V100 GPUs, or in 39 hours on 64 16GB V100 GPUs (Nielsen et al., 2020).

5 Method

We assess the models' environmental impacts by quantifying their total carbon and financial costs from training. Our process and calculations are adapted from Strubell et al. (2019).

We pre-train the previously referenced models, BERT-base, DistilBERT, and ALBERT, using "out-of-box," default settings for a set period. Each model was trained for 8 hours on 1 32GB V100 GPU. While training, we repeatedly query for GPU power draw in watts using nvidia-smi. nvidia-smi reports the last measured power-draw for the entire board, which includes the GPU chip, memory modules, power delivery circuitry, cooling solution, connectors, and supporting components, and

Model	Hardware	n_{gpu}	t (hours)	p_{gpu}	$P_{totalpower}$	C_{CO_2} (lbs)	C_{fin}
BERT-base	32GB V100 GPU	2048	1.03333	0.268854	898.968	770.416	6722.67
DistilBERT	16GB V100 GPU	8	90	0.154449	175.701	150.576	2349.6
ALBERT	16GB V100 GPU	64	39	0.198559	783.054	671.077	8145.28
ALBERT	32GB V100 GPU	64	20	0.198559	401.566	344.142	4066.13

Table 1: Estimated carbon and financial costs of complete training for three models. Listed hardware, t, and n_gpu are derived from instances of complete training. p_{gpu} is calculated from experiments run on 1 32GBV100 GPU.

is accurate within +/- 5 watts[2].

After training, we estimate the total power draw, in kilowatt-hours (kWh), of training each model P to completion using our experimental average power draws and the training specifications reported in instances of complete training. The experimental average GPU power-draw in kilowatts (kW), p_g, is derived from the data points collected using nvidia-smi during training. We multiply this value by the global PUE coefficient e and the hours t and number of GPUs n_{gpu} reported in instances of complete training. The PUE coefficient measures data center efficiency, where 1.0 is the ideal, indicating that all energy consumed by the data center is used to power computing devices. Energy used for non-computing components, such as lighting and cooling, increases PUE above 1.0. The 2023 global average PUE is 1.58 (Bizo, 2023).

$$P = e \times p_g \times n_{gpu} \times t$$

After calculating the total power P consumed in training each model, we can convert it to pounds of carbon dioxide C_{CO2}. For each kWh of electricity generated in the U.S., an average of 0.857 pounds of CO2 is released at the power plant (Center for Sustainable Systems, University of Michigan, 2023). So, we calculate C_{CO2} as follows.

$$C_{CO2} = P \times 0.857$$

Then, we can estimate the financial cost of training each model using cloud computing resources, $C_{financial}$. We approximate the average hourly compute cost C_{cloud} of one 16GB or 32GB V100 GPU across three major cloud providers: AWS, Google Cloud Platform, and Microsoft Azure[3]. Then, depending on the hardware reported in complete trainings, we calculate the financial cost using the referenced GPU's hourly rates and the quantity of GPUs employed n_{gpu}.

$$C_{financial} = n_{gpu} \times t \times c_{cloud}$$

6 Results

Table 1 lists the estimated carbon and financial costs of training the models referenced in 4. Our experimental average power draw, p_g, is derived from a 32GB V100 GPU, while some instances of complete training, which contain n_{gpu} and t, were trained on 16GB V100 GPU. We observed that BERT-base is the most carbon-intensive model, and DistilBERT is the least.

7 Discussion around contextualizing CO2 emissions

7.1 Carbon Valuation

Existing cap-and-trade systems worldwide attribute a monetary value to an "allowance" of one metric tonne of carbon dioxide emissions or the equivalent of other greenhouse gases. This market-driven approach curtails pollution by establishing an annual ceiling, or "cap," on the greenhouse gas emissions within a given region. Participating entities are allotted specific carbon allowances, within which they must confine their emissions. They can trade surplus allowances via auctions, akin to a carbon open market. The system's cap diminishes annually, directing companies to reduce their carbon footprints and invest in cleaner technologies (Comission).

The European Union's Emissions Trading System (ETS), the Regional Greenhouse Gas Initiative (RGGI), and California's Cap-and-Trade Program are three implemented Cap-and-Trade programs. The ETS applies to electricity and heat, energy-intensive (oil refineries, etc.), aviation, and maritime transport industries in the European Union (Comission). In 2024, one allowance is traded

[2]Source: https://developer.download.nvidia.com/compute/DCGM/docs/nvidia-smi-367.38.pdf

[3]Average C_{cloud} for one 32GB V100 GPU is 3.18, and average C_{cloud} for one 16GB V100 GPU is 3.06. Source: https://www.paperspace.com/gpu-cloud-comparison

at an average of €89.60 (approximately \$95.48) (Chen, 2023). Less robust, the RGGI applies to fossil-fuel-fired power generators in the northeastern U.S.[4](RGGI, 2024a). In March 2024, one allowance cleared for an average of \$16 (RGGI, 2024b). Additionally, California's Cap-and-Trade Program encompasses 80% of the state's emissions (CA.gov, 2024). In February 2024, one allowance was auctioned at an average of \$41.76.

With these systematic valuations of carbon emissions, we can examine LLMs' carbon footprints' costs. In our experiment (see section 6, we estimate the financial cost of training the models on cloud resources, but we do not consider the financial cost of the emissions produced. Using the ETS allowance value, emissions from training the most expensive model, BERT-base, would cost $C_{co2} * 95.48 = 33.37$. Under RGGI, BERT-base emissions would cost $C_{co2} * 16.00 = 5.59$. Under California's Cap-and-Trade Program, BERT-base emissions would cost $C_{co2} * 41.76 = 14.59$

So, if LLM developers participated in such Cap-and-Trade systems, they would have to pay a respective \$33.37, \$5.59, or \$14.59 to account for their carbon emissions. Imposing Cap-and-Trade rates could act as a deterrent to inefficient training or finetuning for LLMs. However, one tonne of carbon dioxide is not consistently valued high enough to provoke a significant change in action. If we were to train BERT-base in Massachusetts, using our experimental averages, it would cost \$6723 using cloud resources, and the proposed additional emissions fee of \$5.59 would be merely 0.083% of the estimated compute cost.

Incorporating a standardized system for assigning prices to tonnes of greenhouse gas emissions could spur more energy-efficient LLMs, but this solution is dependent on how such systems set the price. Too low of an evaluation would have a negligible impact on LLM development.

7.2 Impact of LLMs in LMIC

As previously mentioned in Section 2, climate change is a global crisis, and it disproportionately affects historically disadvantaged groups. To consider only one aspect, Berberian et al. (2022) describes how communities of color in the U.S. face a greater risk of climate-related health impacts than Whites. To expand internationally, inequalities in the distribution of basic resources expose the economically disadvantaged to disability, disease, and premature death at a higher rate than the wealthy (Weissglass, 2022). So, there is a need for greater access to healthcare, and LLMs could be beneficial tools in labor and resource-lacking environments. However, implementing such tools faces fundamental challenges including access to training data and infrastructure in LMIC.

In the wake of climate change, there is an even greater need for healthcare in LMIC. LLMs are a promising solution, but we first must consider potential side effects.

LLMs are powerful tools, and they require a lot of power to build them. Our research experiment aims to assess just how much energy is required in training and to calculate the subsequent carbon emissions. These greenhouse gas emissions contribute to climate change, which in turn harms LMIC. To justify the carbon-cost of training, the LLMs produced must be usable and beneficial to their users.

Thus, compressed, more efficient models are preferred when developers aim to limit emissions. However, such compression techniques may introduce bias in LLMs (Ramesh et al., 2023). By chasing efficient models, we risk deploying biased tools in disadvantaged communities that have already been systematically subjected to and damaged by societal prejudices and biases. To this, we pose but do not answer the question: is poor help better than no help?

To limit contributions to climate change, LLMs must be built with energy, and therefore carbon, efficiency in mind. However, we warn against tunnel vision in pursuing this metric, as an efficient model will have no use in a setting that cannot support basic technology, and a biased yet efficient model may have a greater capacity for harm than good.

8 Conclusion

In our experiment, we propose a method for calculating the carbon impact of training BERT-based LLMs.

Recently, there have been greater efforts for transparency and reporting on the carbon impacts of LLMs and other computer science applications in the industry. Researchers more frequently report their pretraining carbon emissions (Meta, 2024; Luccioni et al., 2022). Further, major cloud providers include an account's carbon emissions insights (AWS, 2024; Microsoft, 2024; Cloud, 2024).

[4] CT, ME, MD, MA, NH, NJ, NY, PA, RI, VT

Model	Hardware	n_{gpu}	t (hours)	p_{gpu}	$P_{totalpower}$	C_{CO_2} (lbs)	C_{fin}
BERT-base	32GB V100 GPU	1	8	0.268854	3.39832	2.91236	25.4133
DistilBERT	16GB V100 GPU	1	8	0.154449	1.95223	1.67307	26.1067
ALBERT	16GB V100 GPU	1	8	0.198559	2.50979	2.15089	26.1067
ALBERT	32GB V100 GPU	1	8	0.198559	2.50979	2.15089	25.4133

Table 2: Estimated carbon and financial costs of experimental training for three models. Our experiments were run for 8 hours on 1 32GB V100 GPU. Costs were calculated as described in Section 5.

However, industry leaders like OpenAI still do not report their carbon emissions and overall lack transparency in their models (de Bolle, 2024).

While the community is moving in the right direction, the lack of transparency in company emissions leads to a lack of accountability. We emphasize the need for regulations to mandate emissions reporting and, ideally, limit it. As industry researchers compete to build the best LLMs, models are growing in size, and their energy use follows (Sundberg, 2023).

The National Institute of Standards and Technology's AI Risk Management Framework outlines how the environmental impact of AI model training and management should be assessed and documented (NIST, 2024). Most radically, we suggest not only strict enforcement of the NIST Framework, but to enact regulations on researchers' total carbon emissions, including any emissions companies claim to have offset. While this aim for transparency regulation and emissions constraints may be extreme, the sentiment is a core value.

Climate change is a global crisis that disproportionately affects disadvantaged groups, both within the U.S. and internationally. We assess LLMs' climate impact with the understanding that not all people are affected equally. This problem is complex and highly interdisciplinary; merely building "green" models will not help LMICs that lack basic resources. To truly benefit disadvantaged communities, LLMs must be developed with carbon efficiency, fairness, and context in mind, avoiding the pitfalls of deploying biased models. LLMs are powerful tools, but many communities are not equipped to use them, rendering these tools powerless.

In tandem with promoting efficiency, transparency, and accountability, we emphasize the necessity of reflection and careful evaluation in developing LLMs. Developing and finetuning LLMs to serve unclear purposes will contribute to climate change, while the model may sit unused. LLMs are powerful tools with a great capacity for good, but aimless, expensive development should not become commonplace.

9 Ethics Statement

We acknowledge that our experiments are not exempt from carbon emissions. So, we report the carbon and financial costs of our 8-hour experimental trainings (Table 2) to promote transparency and accountability.

10 Limitations

We were constrained by GPU memory and uptime while obtaining our experimental average power draws. The original models were trained on far larger datasets, which may impact energy usage. Further, we were unable to record the CPU and dynamic random access memory (DRAM) statistics, which contribute to overall power usage, so our estimated power consumption is incomplete.

11 Future Work

Research could be expanded to consider additional power metrics, models, and hardware. Our experiment calculates carbon cost based solely on GPU power draw, but including CPU and DRAM statistics would provide a more comprehensive energy analysis. This experiment focuses only on BERT-based models, but we could include newer LLMs with different architectures. Additionally, we could explore differences in hardware, such as TPUs and A100 GPUs. Finally, we aim to consider the carbon footprint of downstream training.

References

E. Tendayi Achiume. 2022. The global climate crisis is a racial justice crisis: Un expert.

AWS. 2024. Customer carbon footprint tool.

Anahit G Berberian, Daniel JX Gonzalez, and Lara J Cushing. 2022. Racial disparities in climate change-related health effects in the united states. *Current Environmental Health Reports*, 9(3):451–464. Epub 2022 May 28.

Aditya Bindal, Kevin Haas, and Indu Thangakrishnan. 2019. Amazon web services achieves fastest training times for bert and mask r-cnn.

Daniel Bizo. 2023. Global pues — are they going anywhere?

CA.gov. 2024. Cap-and-trade program.

Center for Sustainable Systems, University of Michigan. 2023. Carbon footprint factsheet. Pub. No. CSS09-05.

Scott Chen. 2023. Bearish eu carbon prices to continue in 2024 on lower power emissions, oversupply.

Taylor Ciecierski-Holmes, Rajesh Singh, Malte Axt, et al. 2022. Artificial intelligence for strengthening healthcare systems in low- and middle-income countries: a systematic scoping review. *npj Digital Medicine*, 5:162.

Google Cloud. 2024. Carbon footprint.

European Comission. Emissions cap and allowances.

Monica de Bolle. 2024. Ai's carbon footprint appears likely to be alarming.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding.

Eurostat. Glossary:fossil fuel.

Allen Institute for AI. 2024. C4.

Gustavo Gonçalves and Emma Strubell. 2023. Understanding the effect of model compression on social bias in large language models.

IEA. 2023. Greenhouse gas emissions from energy data explorer.

Lynn H. Kaack, Priya L. Donti, Emma Strubell, et al. 2022. Aligning artificial intelligence with climate change mitigation. *Nat. Clim. Chang.*, 12:518–527.

Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2020. Albert: A lite bert for self-supervised learning of language representations.

Tara Law. 2019. The climate crisis is global, but these 6 places face the most severe consequences.

Sasha Luccioni, Zachary Mueller, and Nate Raw. 2022. Co2 emissions and the hub: Leading the charge.

Meta. 2024. Meta llama 3.

Microsoft. 2024. Emissions impact dashboard.

Steven Gonzalez Monserrate. 2022. The Cloud Is Material: On the Environmental Impacts of Computation and Data Storage. *MIT Case Studies in Social and Ethical Responsibilities of Computing*, (Winter 2022). Https://mit-serc.pubpub.org/pub/the-cloud-is-material.

NASA. 2024. The effects of climate change.

United Nations. The paris agreement.

Jared Nielsen, Aditya Bindal, and Derya Cavdar. 2020. Train albert for natural language processing with tensorflow on amazon sagemaker.

NIST. 2024. Ai risk management framework.

NRDC. 2023. Greenhouse effect 101.

Krithika Ramesh, Arnav Chavan, Shrey Pandit, and Sunayana Sitaram. 2023. A comparative study on the impact of model compression techniques on fairness in language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15762–15782, Toronto, Canada. Association for Computational Linguistics.

RGGI. 2024a. Elements of rggi.

RGGI. 2024b. Market monitor report for auction 63.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2020. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter.

Roy Schwartz, Jesse Dodge, Noah A. Smith, and Oren Etzioni. 2019. Green AI. *CoRR*, abs/1907.10597.

Md Abu Bakar Siddik, Arman Shehabi, and Landon Marston. 2021. The environmental footprint of data centers in the united states. *Environmental Research Letters*, 16(6):064017.

Emma Strubell, Ananya Ganesh, and Andrew McCallum. 2019. Energy and policy considerations for deep learning in NLP. *CoRR*, abs/1906.02243.

Niklas Sundberg. 2023. Tackling ai's climate change problem.

Daphne E Weissglass. 2022. Contextual bias, the democratization of healthcare, and medical artificial intelligence in low- and middle-income countries. *Bioethics*, 36(2):201–209. Epub 2021 Aug 30.