Researching Responsible and Trustworthy Natural Language Processing

Sustainability

Emily Allaway

23 February 2026

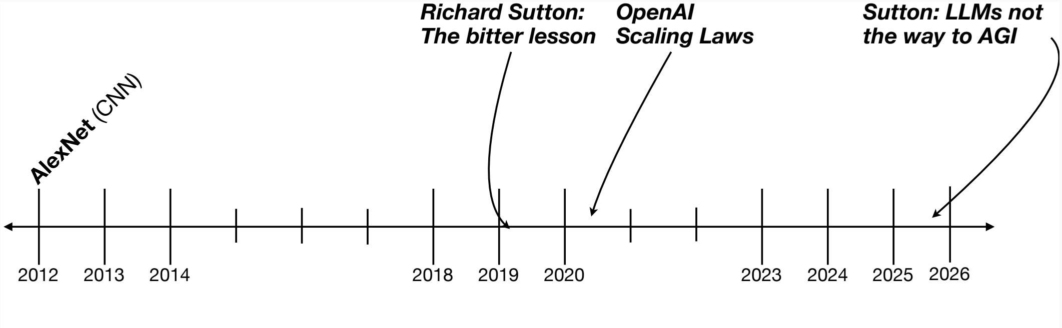
School of Informatics
University of Edinburgh
eallaway@ed.ac.uk

Overview

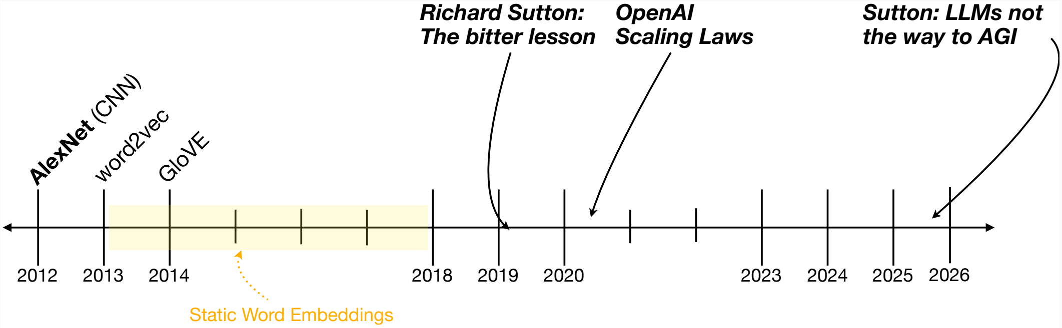
The rise of scaling and LLMs

- AlexNet (Krizhevsky et al., 2012)
  "All of our experiments suggest that our results can be improved simply by waiting for faster GPUs and bigger datasets to become available."
- The bitter lesson (Sutton, 2019)
  "...the great power of general purpose methods, of methods that continue to scale with increased computation"
- OpenAI Scaling Laws (Kaplan et al., 2020)

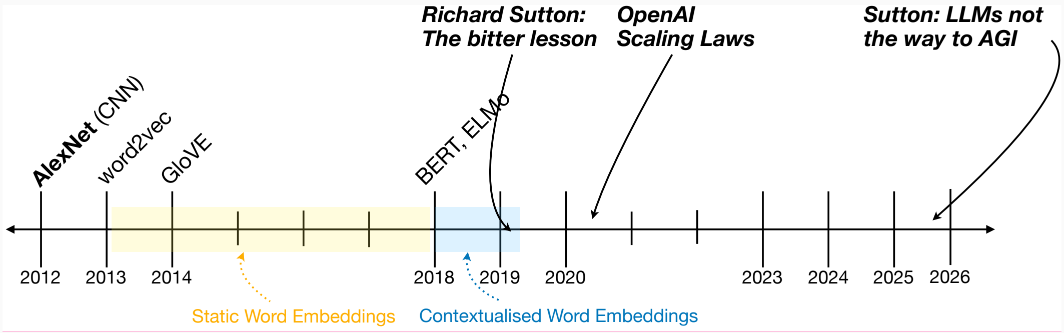
Timeline of LMs



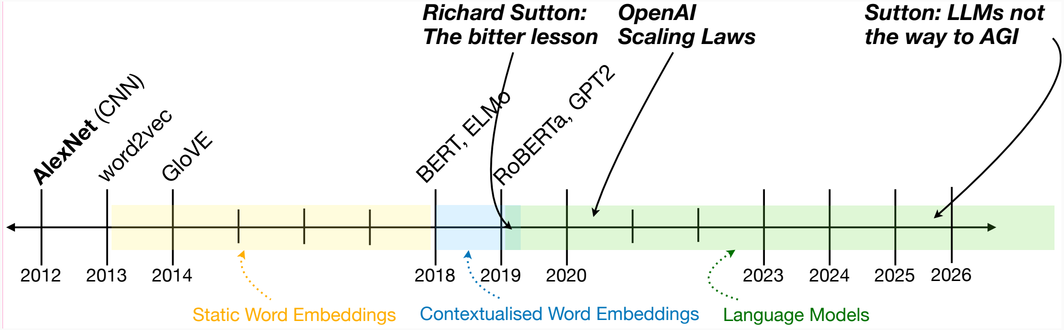
Timeline of LMs



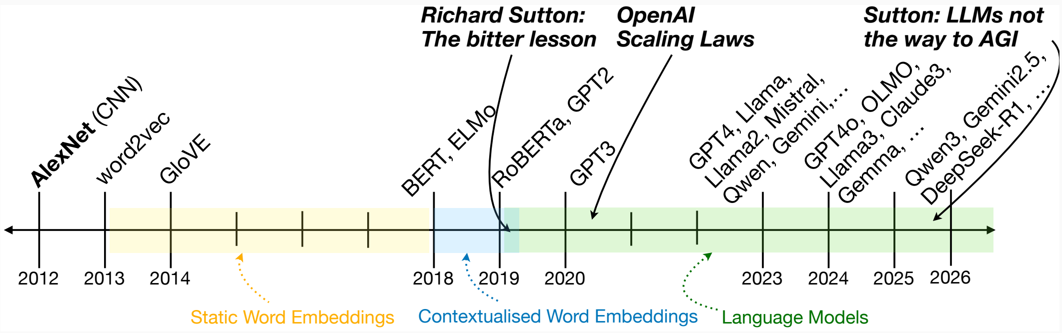
Timeline of LMs



Timeline of LMs



Timeline of LMs

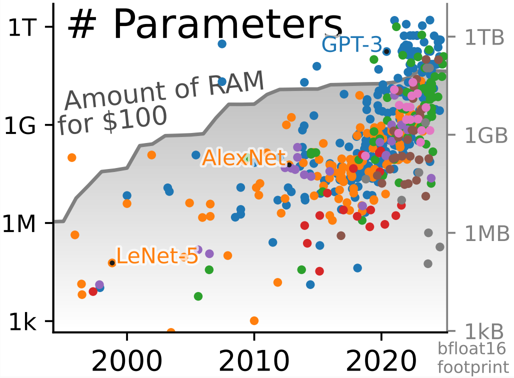


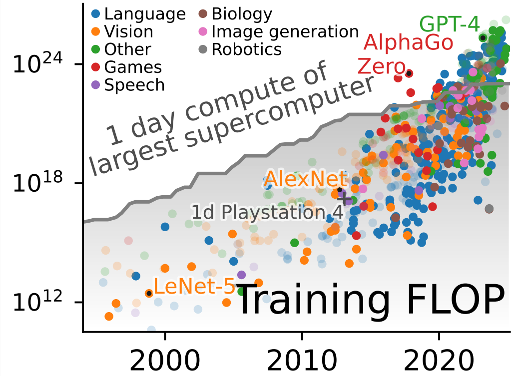
Jevons paradox

Economics:	AI:
1. Efficiency of general use technology ↑	AI usage of GPUs
2. Cost of that technology ↓	GPUs of a certain efficiency
3. Demand for the cheaper technology ↑	More people using GPUs
4. Overall resource usage ↑	More GPUs needed

From Luccioni et al. (2025).

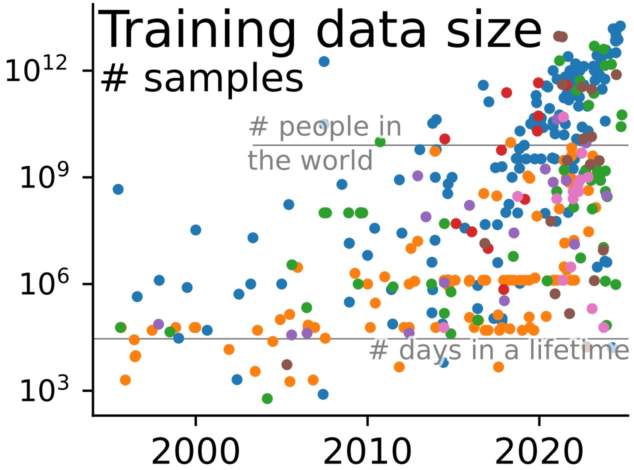
Bigger requires more compute





From Varoquaux et al. (2025).

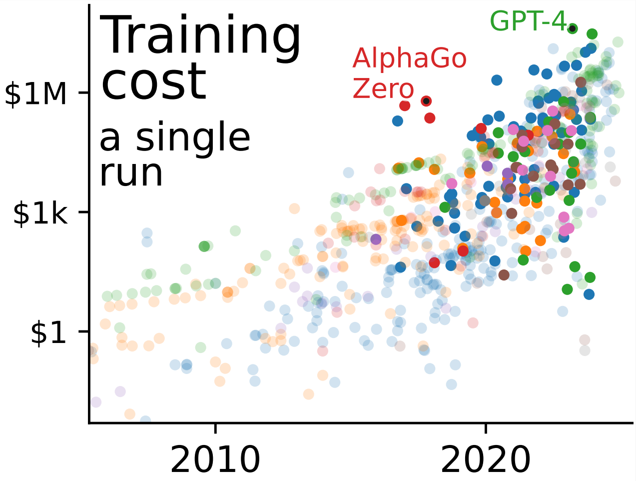
Data requirements



One consequence: companies turning to potentially immoral/illegal data gathering

From Varoquaux et al. (2025).

Cost and academia



From Varoquaux et al. (2025).

Training time for BLOOM

From Luccioni et al. (2023)

Total training time	118 days, 5 hours, 41 min
Total number of GPU hours	1,082,990 hours
Total energy used	433,196 kWh
GPU models used	Nvidia A100 80GB
Carbon intensity of the energy grid	57 gCO_2eq/kWh

Model: 176B parameters — Training time $\approx$ 4 months

Training data: 1.6TB (Laurençon et al., 2022) $\approx$ 379B tokens (Workshop et al., 2022)

GPU hours $\approx$ 123 *years* — $\rightarrow$ so used 384 GPUs

Carbon emissions $\approx$ 27 tonnes — $\approx$ driving a car NYC to San Francisco 31 times

Overview

Carbon emissions

Model name	Number of parameters	Datacenter PUE	Carbon intensity of grid used	Power consumption	CO_2eq emissions	CO_2eq emissions $\times$ PUE
GPT-3	175B	1.1	429 gCO_2eq/kWh	1,287 MWh	*502 tonnes*	552 tonnes
Gopher	280B	1.08	330 gCO_2eq/kWh	*1,066 MWh*	*352 tonnes*	380 tonnes
OPT	175B	*1.09* [2]	*231 gCO_2eq/kWh*	*324 MWh*	70 tonnes	*76.3 tonnes* [3]
BLOOM	176B	1.2	57 gCO_2eq/kWh	433 MWh	25 tonnes	30 tonnes

502 tonnes ≈ driving a car NYC to San Francisco 438 times

352 tonnes ≈ 401 times

70 tonnes ≈ 80 times

25 tonnes ≈ 31 times

From Luccioni et al. (2023) and Climate Impact Partners (2025)

Emissions breakdown for BLOOM

Process	**CO_2emissions (CO_2eq)**	**Percentage of total emissions**
Embodied emissions	11.2 tonnes	22.2 %
Dynamic consumption	24.69 tonnes	48.9 %
Idle consumption	14.6 tonnes	28.9 %
Total	**50.5 tonnes**	**100.00%**

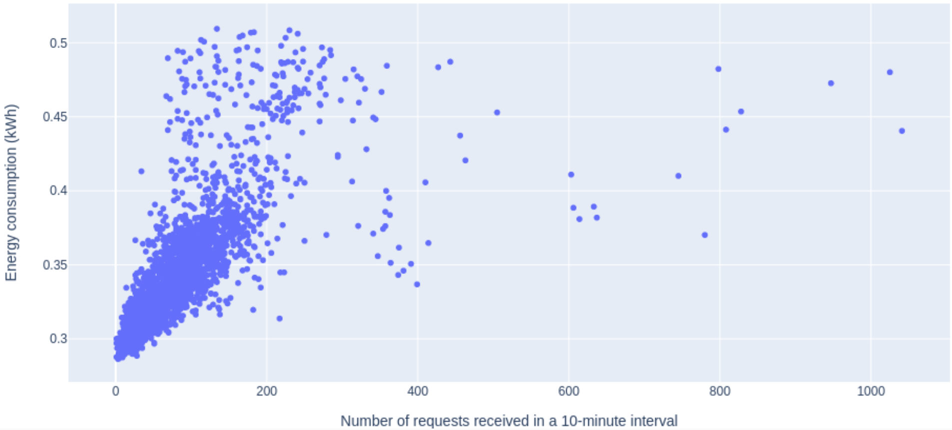
Embodied: producing the computing equipment

Dynamic: power needed for training

Idle: the broader infrastructure staying on

From Luccioni et al. (2023).

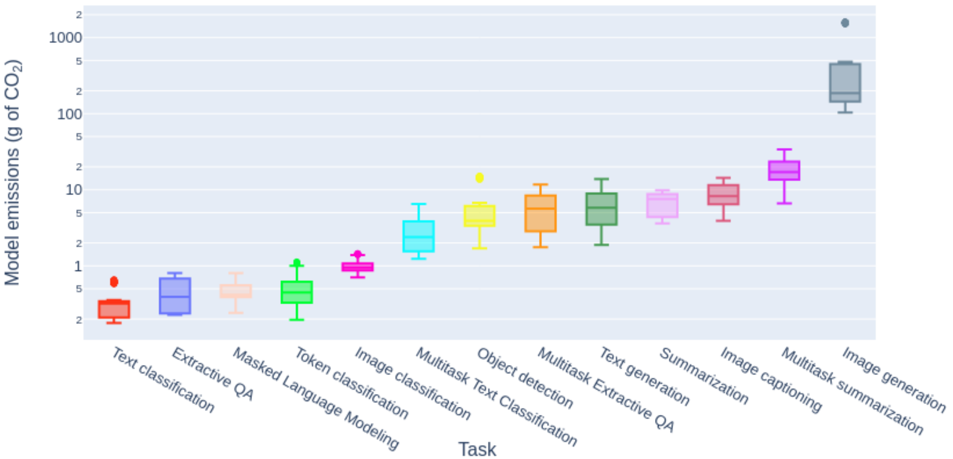
Inference energy usage for BLOOM API



Deployed on Google Cloud

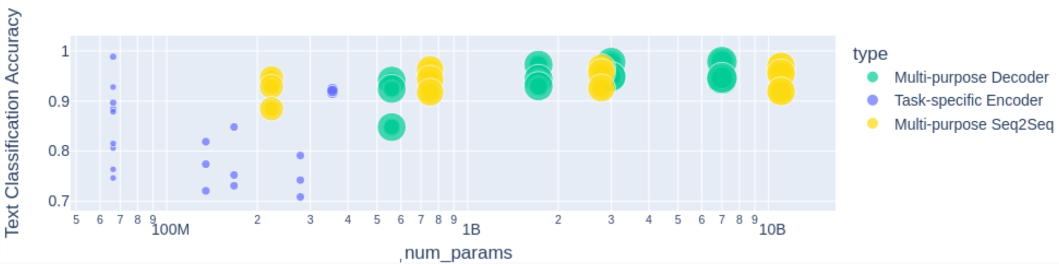
From Luccioni et al. (2023).

Inference emissions across tasks



From Luccioni et al. (2024).

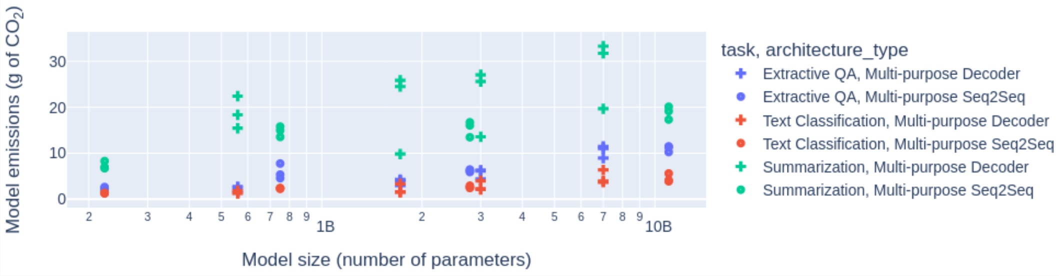
Inference emissions from general-purpose vs. task-specific models



Dot size indicates quantity of emissions (log scale)

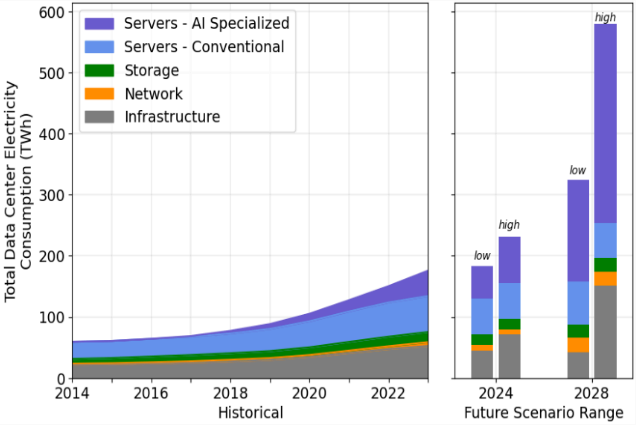
From Luccioni et al. (2024).

General-purpose model inference emissions by task



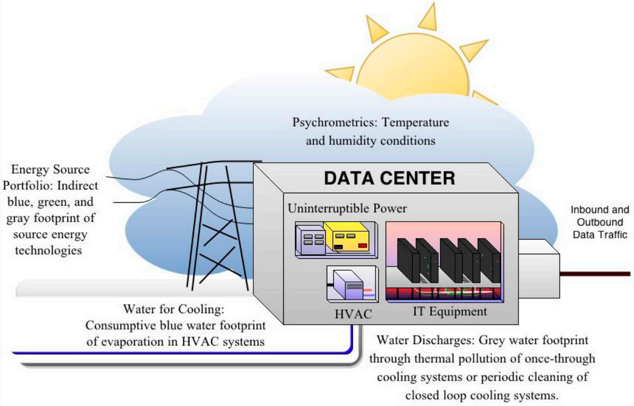
From Luccioni et al. (2024).

US Data Center Electricity Consumption 2023



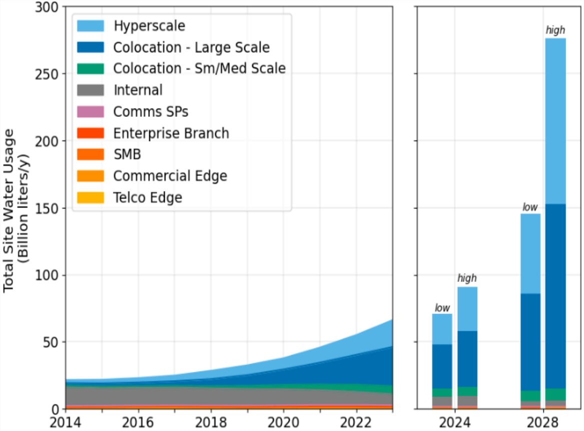
From Shehabi et al. (2024).

Water consumption



From Ristic et al. (2015).

US Data Center Water Consumption 2023



Usage in 2023: 66B liters $\approx 26k$ olympic-size swimming pools

From Shehabi et al. (2024).

Scaling and generalisation

ACL 2025 Theme: **Generalization of NLP Models**

- Question to discussion panel:
  If I worked with SLMs, would you accept my papers?

Recommended Reading & Resources i

1. Hype, Sustainability, and the Price of the Bigger-is-Better Paradigm in AI (Varoquaux et al., 2025)
    - General overview of many of the issues
2. Power Hungry Processing: Watts Driving the Cost of AI Deployment (Luccioni et al., 2024)
    - Analysis of inference-time energy usage and emissions
3. BLOOM and the BigScience initiative
    - BLOOM: A 176B-parameter Open-Access Multilingual Language Model (Workshop et al., 2022)
    - The BigScience Roots Corpus: A 1.6 TB Composite Multilingual Dataset (Laurençon et al., 2022)

Recommended Reading & Resources ii

- Estimating the carbon footprint of BLOOM, a 176B parameter Language Model (Luccioni et al., 2023)

4. From Efficiency Gains to Rebound Effects: The Problem of Jevons' Paradox in AI's Polarized Environmental Debate (Luccioni et al., 2025)
   - Interesting discussion on many impacts of AI on environment, economy, and society
5. Mapping 1,000+ Language Models via the Log-Likelihood Vector (Oyama et al., 2025)
   - Table 10 in the Appendix lists 1018 LMs along with their size and number of downloads (popularity)
6. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? (Bender et al., 2021)
   - Seminal paper on issues with scaling and LMs

Recommended Reading & Resources iii

7. Gary Marcus' blog on AI
    - Game over for pure LLMs. Even Turing Award Winner Rich Sutton has gotten off the bus. https://garymarcus.substack.com/p/game-over-for-pure-llms-even-turing
    - "Scale is All You Need" is dead https://garymarcus.substack.com/p/breaking-news-scale-is-all-you-need
    - How Generative AI is destroying society https://garymarcus.substack.com/p/how-generative-ai-is-destroying-society
8. Epoch AI Table of Models: https://epoch.ai/data/ai-models

References i

Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 610–623.

Website: Climate Impact Partners. 2025. The carbon footprint of ai.

Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. 2020. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.

Hugo Laurençon, Lucile Saulnier, Thomas Wang, Christopher Akiki, Albert Villanova del Moral, Teven Le Scao, Leandro Von Werra, Chenghao Mou, Eduardo González Ponferrada, Huu Nguyen, et al. 2022. The bigscience roots corpus: A 1.6 tb composite multilingual dataset. *Advances in Neural Information Processing Systems*, 35:31809–31826.

Alexandra Sasha Luccioni, Emma Strubell, and Kate Crawford. 2025. From efficiency gains to rebound effects: The problem of jevons' paradox in ai's polarized environmental debate. In *Proceedings of the 2025 ACM conference on fairness, accountability, and transparency*, pages 76–88.

Alexandra Sasha Luccioni, Sylvain Viguier, and Anne-Laure Ligozat. 2023. Estimating the carbon footprint of bloom, a 176b parameter language model. *Journal of machine learning research*, 24(253):1–15.

Sasha Luccioni, Yacine Jernite, and Emma Strubell. 2024. Power hungry processing: Watts driving the cost of ai deployment? In *Proceedings of the 2024 ACM conference on fairness, accountability, and transparency*, pages 85–99.

Momose Oyama, Hiroaki Yamagiwa, Yusuke Takase, and Hidetoshi Shimodaira. 2025. Mapping 1,000+ language models via the log-likelihood vector. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 32983–33038, Vienna, Austria. Association for Computational Linguistics.

Bora Ristic, Kaveh Madani, and Zen Makuch. 2015. The water footprint of data centers. *Sustainability*, 7(8):11260–11284.

A Shehabi, A Newkirk, S Smith, A Hubbard, N Lei, M Siddik, B Holecek, J Koomey, E Masanet, and D Sartor. 2024. 2024 united states data center energy usage report.

Richard Sutton. 2019. The bitter lesson. Incomplete ideas.

Gaël Varoquaux, Sasha Luccioni, and Meredith Whittaker. 2025. Hype, sustainability, and the price of the bigger-is-better paradigm in ai. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, pages 61–75.

BigScience Workshop, Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilić, Daniel Hesslow, Roman Castagné, Alexandra Sasha Luccioni, François Yvon, et al. 2022. Bloom: A 176b-parameter open-access multilingual language model. *arXiv preprint arXiv:2211.05100*.