Supplementary Materials

Revisit the Environmental Impact of Artificial Intelligence: The Overlooked Carbon Emission Source?

Yang Yu [a], Jiahui Wang [a], Yu Liu [b], Pingfeng Yu [a, c, d], Dongsheng Wang [a, c, d], Ping Zheng [a], Meng Zhang [a, c, d*]

[a] *Department of Environmental Engineering, College of Environmental & Resource Sciences, Zhejiang University, Hangzhou, China*

[b] *Engineering Laboratory of Low-Carbon Unconventional Water Resources Utilization and Water Quality Assurance, College of Environmental Science and Engineering, Nankai University, Tianjin 300350, China*

[c] *Zhejiang Province Key Laboratory for Water Pollution Control and Environmental Safety, Hangzhou, China*

[d] *Innovation Center of Yangtze River Delta, Zhejiang University, Jiashan 314100, China*

[*] Corresponding author: Meng Zhang

Tel: +86-0571-88982819; Fax: +86-0571-88982819

E-mail address: zhangm_environment@zju.edu.cn

This Supplementary material contains 1 figure and 1 table.

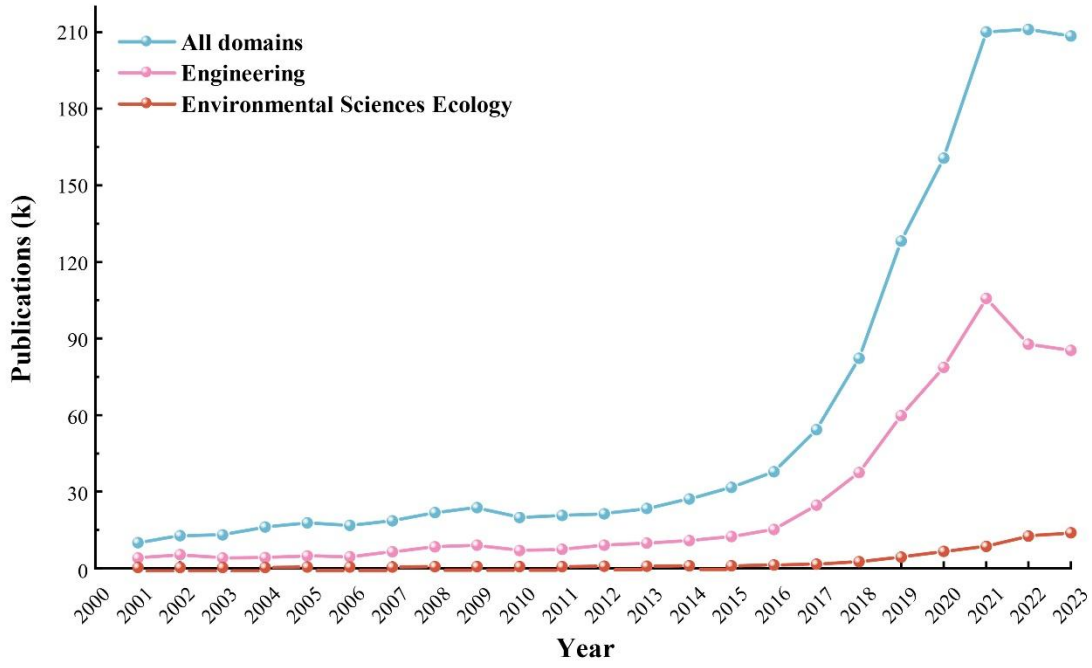

Fig. S1 Publication of AI related studies (including journal publications and patents) in different domains (Source: Clarivate Web of Science; Searched with the TS (Topic, which consists of titles, abstracts, and keywords) = (“Artificial Intelligence” OR “AI”); Accessed 24 April 2024.)

Table S1 Calculated CO_2eq of compute-notable AI systems.

System	Publication date	Training compute (FLOP)	Training hardware	TDP (W)	Computing power (FLOP/s)	Training chip-hours	Hardware utilization	Reference link	Energy consumption (kW·h)	CO_2eq (t)
Gemini Ultra	2023/12/6	5.00E+25	Google TPU v4	300	2.75E+14	1.32E+08	-	https://doi.org/10.48550/arXiv.2312.11805	39600000	37620
GPT-4	2023/3/15	2.10E+25	NVIDIA A100 SXM4 40 GB	400	3.12E+14	5.70E+07	0.34	https://doi.org/10.48550/arXiv.2303.08774	22800000	21660
Mistral Large	2024/2/26	2.00E+25	NVIDIA H100 PCIe	350	9.89E+14	1.87E+07	-	https://mistral.ai/news/mistral-large/	6553571	6226
Inflection-2	2023/11/22	1.00E+25	NVIDIA H100 SXM5	700	7.56E+14	1.23E+07	-	https://inflection.ai/inflection-2	8581962	8153
PaLM 2	2023/5/10	7.34E+24	Google TPU v4	300	2.75E+14	2.47E+07	-	https://arxiv.org/abs/2305.10403	7414141	7043
Claude 2	2023/7/11	3.87E+24	NVIDIA A100	250	3.12E+14	1.15E+07	-	https://www.anthropic.com/index/claude-2	2868293	2725
Falcon-180B	2023/9/6	3.76E+24	NVIDIA A100 SXM4 40 GB	250	3.12E+14	1.77E+07	0.19	https://arxiv.org/abs/2311.16867	4423680	4202
Minerva (540B)	2022/6/29	2.74E+24	Google TPU v4	300	2.75E+14	7.13E+05	-	https://arxiv.org/abs/2206.14858	213811	203
DBRX	2024/3/27	2.60E+24	NVIDIA H100 SXM5	700	7.56E+14	3.18E+06	-	https://www.databricks.com/blog/introducing-dbrx-new-state-art-open-llm	2229081	2118

GPT-3.5 (text-davinci-003)	2022/11/28	2.58E+24	NVIDIA A100 SXM4 40 GB	250	3.12E+14	7.65E+06	-	https://platform.openai.com/docs/models/gpt-3-5	1912690	1817
U-PaLM (540B)	2022/10/20	2.53E+24	Google TPU v4	300	2.75E+14	6.14E+04	-	https://arxiv.org/abs/2210.11399	18432	18
PaLM (540B)	2022/4/4	2.53E+24	Google TPU v4	300	2.75E+14	8.40E+06	0.46	https://arxiv.org/abs/2204.02311	2521498	2395
Qwen-72B	2023/11/30	1.30E+24	NVIDIA A100	250	3.12E+14	3.86E+06	-	https://huggingface.co/Qwen/Qwen-72B	964506	916
XVERSE-65B-2	2023/12/8	1.25E+24	NVIDIA A100	250	3.12E+14	3.70E+06	-	https://github.com/xverse-ai/XVERSE-65B/blob/main/README_EN.md	925926	880
Code Llama-70B	2024/1/29	1.23E+24	NVIDIA A100 SXM4 80 GB	400	3.12E+14	2.59E+06	0.44	https://ai.meta.com/research/publications/code-llama-open-foundation-models-for-code/ https://arxiv.org/abs/2308.12950	1036800	985
Megatron-Turing NLG 530B	2021/10/11	1.17E+24	NVIDIA A100 SXM4 80 GB	400	3.12E+14	3.45E+06	0.30	https://arxiv.org/abs/2201.11990	1379840	1311
ChatGLM3	2023/10/27	1.09E+24	NVIDIA A100	250	3.12E+14	3.24E+06	-	https://www.zhipuai.cn/en/news/76	810185	770
ERNIE 3.0 Titan	2021/12/23	1.04E+24	Huawei Ascend 910,NVIDIA	250	1.12E+14	8.62E+06	-	https://arxiv.org/abs/2112.12731	2153811	2046

			Tesla V100 DGXS 32 GB							
TigerBot-70B	2023/9/6	1.02E+24	NVIDIA A100	250	3.12E+14	3.03E+06	-	https://github.com/TigerResearch/TigerBot/blob/main/README_en.md, https://arxiv.org/abs/2312.08688	756766	719
Inflection-1	2023/6/23	1.00E+24	NVIDIA H100 SXM5	700	7.56E+14	1.22E+06	-	https://inflection.ai/assets/Inflection-1.pdf	857425	815
Llama 2-70B	2023/7/18	8.10E+23	NVIDIA A100 SXM4 80 GB	400	3.12E+14	2.16E+06	0.44	https://ai.meta.com/research/publications/llama-2-open-foundation-and-fine-tuned-chat-models/ https://arxiv.org/abs/2307.09288	864000	821
DeepSeek LLM 67B	2024/1/5	8.04E+23	NVIDIA A100	250	3.12E+14	2.39E+06	-	https://arxiv.org/abs/2401.02954, https://github.com/deepseek-ai/DeepSeek-LLM	596510	567
Gopher (280B)	2021/12/8	6.31E+23	Google TPU v3	283	-	3.77E+06	0.38	https://arxiv.org/abs/2112.11446	1066435	1013
Yi-34B	2023/11/2	6.10E+23	NVIDIA A100	250	3.12E+14	1.81E+06	-	https://arxiv.org/abs/2403.04652	452576	430

xTrimoPGLM -100B	2023/7/6	6.00E+23	NVIDIA A100 SXM4 40 GB	250	3.12E+14	2.35E+03	-	https://www.biorxiv.org/content/10.1101/2023.07.05.547496v1	588	1
BLOOM-176B	2022/11/8	5.77E+23	NVIDIA A100 SXM4 80 GB	400	3.12E+14	1.08E+06	0.48	https://arxiv.org/abs/2211.05100	431309	410
Chinchilla	2022/3/29	5.76E+23	Google TPU v4,Google TPU v3	300	2.75E+14	1.94E+06	-	https://arxiv.org/abs/2203.15556	581818	553
BIG-G 137B	2022/6/9	5.60E+23	NVIDIA A100	250	3.12E+14	1.66E+06	-	https://arxiv.org/abs/2206.04615	415480	395
LLaMA-65B	2023/2/24	5.50E+23	NVIDIA A100	400	3.12E+14	1.02E+06	0.47	https://arxiv.org/abs/2302.13971	409600	389
Code Llama-34B	2023/8/14	5.30E+23	NVIDIA A100 SXM4 80 GB	400	3.12E+14	1.57E+06	-	https://arxiv.org/abs/2308.12950	629155	598
PanGu	2023/3/20	4.67E+23	Huawei Ascend 910	310	-	1.23E+06	-	https://arxiv.org/abs/2303.10845	380928	362
BlenderBot 3	2022/8/10	4.30E+23	NVIDIA A100 SXM4 40 GB	250	3.12E+14	1.28E+06	-	https://arxiv.org/abs/2208.03188	319029	303
OPT-175B	2022/5/2	4.30E+23	NVIDIA A100 SXM4 80 GB	400	3.12E+14	8.13E+05	0.47	https://arxiv.org/abs/2205.01068	325018	309
Llama 2-34B	2023/7/18	4.08E+23	NVIDIA A100 SXM4 80 GB	400	3.12E+14	1.21E+06	-	https://arxiv.org/abs/2307.09288	484330	460
ViT-22B	2023/2/10	4.00E+23	Google TPU v4	300	2.75E+14	1.35E+06	-	https://arxiv.org/abs/2302.05442v1	404051	384

Parti	2022/6/22	3.96E+23	Google TPU v4	300	2.75E+14	1.33E+06	-	https://arxiv.org/abs/2206.10789v1	400292	380
DeepSeek Coder 33B	2023/11/2	3.96E+23	NVIDIA A100	250	3.12E+14	1.18E+06	-	https://github.com/deepseek-ai/DeepSeek-Coder	293803	279
StarCoder 2 15B	2024/2/29	3.87E+23	NVIDIA A100	250	3.12E+14	1.15E+06	-	https://arxiv.org/abs/2402.19173	287126	273
GLM-130B	2022/8/4	3.78E+23	NVIDIA A100 SXM4 40 GB	250	3.12E+14	1.11E+06	0.43	https://keg.cs.tsinghua.edu.cn/glm-130b/posts/glm-130b/	276480	263
GLaM	2021/12/13	3.74E+23	Google TPU v4	300	2.75E+14	1.40E+06	-	https://arxiv.org/abs/2112.06905	419635	399
Jurassic-1-Jumbo	2021/8/11	3.70E+23	NVIDIA A100	250	3.12E+14	1.10E+06	-	https://uploads-ssl.webflow.com/60fd4503684b466578c0d307/61138924626a6981ee09caf6_jurassic_tech_paper.pdf	274513	261
LaMDA	2022/2/10	3.55E+23	Google TPU v3	283	-	1.42E+06	0.57	https://arxiv.org/abs/2201.08239	401362	381
Yuan 1.0	2021/10/12	3.54E+23	NVIDIA A100	250	3.12E+14	7.00E+05	0.45	https://arxiv.org/abs/2110.04725	174996	166
CodeFuse-13B	2023/10/10	3.30E+23	NVIDIA A100 SXM4 80 GB	400	3.12E+14	9.79E+05	-	https://arxiv.org/abs/2310.06266	391738	372

Galactica	2022/11/16	3.24E+23	NVIDIA A100 SXM4 80 GB	400	3.12E+14	9.62E+05	-	https://arxiv.org/abs/2211.09085	384615	365
GPT-3 175B (davinci)	2020/5/28	3.14E+23	NVIDIA Tesla V100 DGXS 32 GB	300	1.12E+14	3.55E+06	0.22	https://arxiv.org/abs/2005.14165	1065600	1012
Luminous-supreme	2022/8/15	2.80E+23	NVIDIA A100 SXM4 40 GB,NVIDIA A100 SXM4 80 GB	325	3.12E+14	8.31E+05	-	https://docs.aleph-alpha.com/docs/introduction/model-card/	270062	257
LLaMA-33B	2023/2/27	2.73E+23	NVIDIA A100	250	3.12E+14	8.10E+05	-	https://arxiv.org/abs/2302.13971	202546	192
Whisper v3	2023/11/6	2.70E+23	NVIDIA A100	250	3.12E+14	8.01E+05	-	https://huggingface.co/openai/whisper-large-v3	200321	190
Flamingo	2022/4/29	2.70E+23	Google TPU v4	300	2.75E+14	5.53E+05	-	https://arxiv.org/abs/2204.14198	165888	158
Gemma 7B	2024/2/21	2.52E+23	Google TPU v5e	300	1.97E+14	1.18E+06	-	https://storage.googleapis.com/deepmind-media/gemma/gemma-report.pdf	355330	338
Skywork-13B	2023/10/30	2.50E+23	NVIDIA A800	300	3.12E+14	4.84E+05	0.46	https://arxiv.org/abs/2310.19341	145160	138
Qwen-14B	2023/9/28	2.50E+23	NVIDIA A100	250	3.12E+14	7.42E+05	-	https://arxiv.org/abs/2309.16609	185482	176
Granite 13B	2023/11/30	2.44E+23	NVIDIA A100	400	3.12E+14	7.24E+05	-	https://www.ibm.com/	289649	275

								downloads/cas/X9W4O6BM		
Nanbeige-16B	2023/11/1	2.40E+23	NVIDIA A100	250	3.12E+14	7.12E+05	-	https://github.com/Nanbeige/Nanbeige/blob/main/README_EN.md	178063	169
LightOn Mini	2023/3/21	2.40E+23	A100	250	3.12E+14	7.12E+05	-	https://www.lighton.ai/blog/lighton-s-blog-4/lighton-s-large-language-model-of-40-billion-parameters-mini-19	178063	169
Falcon-40B	2023/3/15	2.40E+23	NVIDIA A100	400	3.12E+14	5.53E+05	0.39	https://arxiv.org/abs/2311.16867	221184	210
BloombergGPT	2023/3/30	2.36E+23	NVIDIA A100	400	3.12E+14	6.50E+05	0.32	https://arxiv.org/abs/2303.17564	260096	247
YaLM	2022/6/23	2.20E+23	NVIDIA A100	250	3.12E+14	1.25E+06		https://medium.com/yandex/yandex-publishes-yalm-100b-its-the-largest-gpt-like-neural-network-in-open-source-d1df53d0e9a6	312000	296
AlexaTM 20B	2022/8/2	2.04E+23	NVIDIA A100	250	3.12E+14	3.69E+05	0.49	https://arxiv.org/abs/2208.01448	92160	88
Baichuan2-13B	2023/9/6	2.03E+23	NVIDIA A100	250	3.12E+14	6.02E+05	-	https://arxiv.org/abs/2309.10305	150611	143

Nemotron-3-8B	2023/11/15	1.80E+23	NVIDIA A100	400	3.12E+14	4.71E+05	0.34	https://developer.nvidia.com/blog/nvidia-ai-foundation-models-build-custom-enterprise-chatbots-and-co-pilots-with-production-ready-llms/	188537	179
MPT-30B	2023/6/22	1.80E+23	NVIDIA A100 SXM4 40 GB,NVIDIA H100 SXM5	475	3.12E+14	5.34E+05	-	https://huggingface.co/mosaicml/mpt-30b	253739	241
Llama 2-13B	2023/7/18	1.60E+23	NVIDIA A100 SXM4 80 GB	400	3.12E+14	4.75E+05	-	https://arxiv.org/abs/2307.09288	189934	180
SparseOPT-175B	2023/1/2	1.58E+23	NVIDIA A100	250	3.12E+14	4.69E+05	-	https://arxiv.org/abs/2301.00774	117225	111
AlphaCode	2022/2/2	1.57E+23	Google TPU v4,Google TPU v4i	300	2.75E+14	5.28E+05	-	https://arxiv.org/abs/2203.07814	158400	150
StarCoder 2 7B	2024/2/29	1.55E+23	NVIDIA H100 SXM5	350	7.56E+14	1.90E+05	-	https://arxiv.org/abs/2402.19173	66444	63

Poro34B (700B token checkpoint)	2023/12/14	1.53E+23	AMD Instinct MI250X	560	3.83E+14	3.70E+05	-	https://huggingface.co/LumiOpen/Poro-34B	207137	197
HyperCLOVA	2021/9/10	1.48E+23	NVIDIA A100	250	3.12E+14	6.59E+05	0.20	https://arxiv.org/abs/2109.04650	164659	156
UL2	2022/5/10	1.20E+23	Google TPU v4	300	2.75E+14	3.81E+05	0.32	https://arxiv.org/abs/2205.05131v1	114278	109
PLaMo-13B	2023/9/28	1.17E+23	NVIDIA A100	250	3.12E+14	3.47E+05	-	https://huggingface.co/pfnet/plamo-13b	86806	82
IDEFICS	2023/8/22	1.16E+23	NVIDIA A100	250	3.12E+14	3.44E+05	-	https://huggingface.co/blog/idefics	86016	82
StarCoder	2023/5/9	1.12E+23	NVIDIA A100 SXM4 80 GB	400	3.12E+14	3.20E+05		https://arxiv.org/abs/2305.06161	128102	122
Meena	2020/1/28	1.12E+23	Google TPU v3	300	1.23E+14	7.37E+05	0.34	https://arxiv.org/abs/2001.09977	221184	210
OPT-66B	2022/6/21	1.10E+23	NVIDIA A100	250	3.12E+14	3.26E+05	0.30	https://huggingface.co/facebook/opt-66b	81612	78
Code Llama-7B	2023/8/14	1.10E+23	NVIDIA A100 SXM4 80 GB	400	3.12E+14	3.26E+05	0.30	https://ai.meta.com/research/publications/code-llama-open-foundation-models-for-code/ https://arxiv.org/abs/2308.12950	130579	124

Whisper v2	2022/12/5	1.10E+23	NVIDIA A100	250	3.12E+14	3.26E+05	0.30	https://huggingface.co/openai/whisper-large-v2	81612	78
BlueLM 13B	2023/10/31	1.09E+23	NVIDIA A100	250	3.12E+14	3.24E+05	0.30	https://github.com/vivo-ai-lab/BlueLM	81019	77
Qwen-7B	2023/9/28	1.01E+23	NVIDIA A100	250	3.12E+14	3.00E+05	0.30	https://arxiv.org/abs/2309.16609	74935	71
Total	-	-	-	-	-	-	-	-	125392236	119123

Note that the Energy consumption and CO_2eq are calculated results, and other information can be sourced in Epoch AI (2024).

References

Epoch AI (2024). Compute Trends Across Three Eras of Machine Learning. Available online at https://epochai.org/blog/compute-trends. (accessed April 24, 2024).