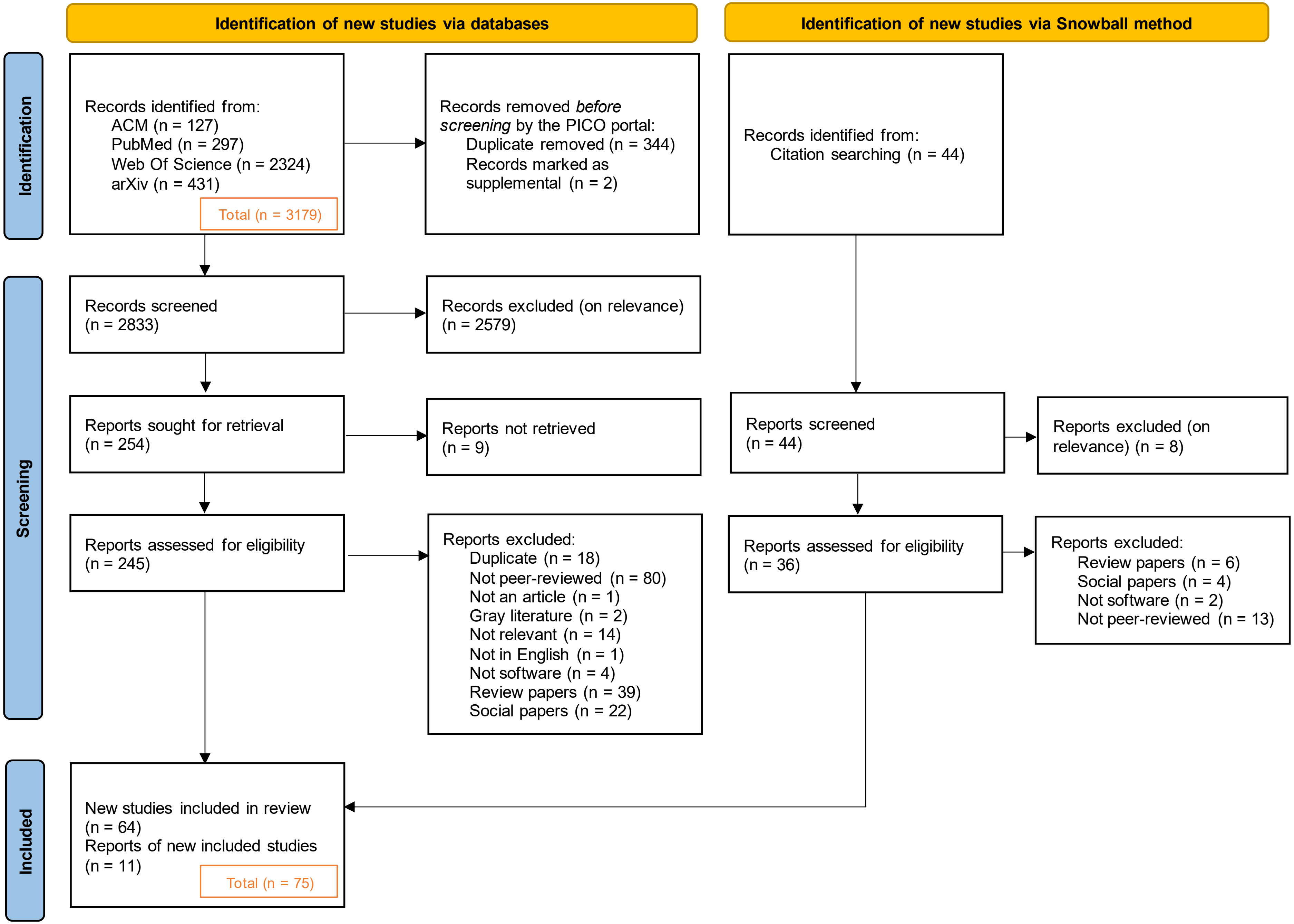
PRISMA flow diagram for "Green AI: Systematic Review and Guidelines for Sustainability in Artificial Intelligence"



“Green AI: Systematic Review and Guidelines for Sustainability in Artificial Intelligence”

Data Extracted from 75 Publications:

Paper	Type of Study	AI Lifecycle Stage	Type of AI Model	Learning Paradigm	Computing Infrastructure	Hardware Used	Software Libraries & Framework	Consumption-related Measuring Tools	Environmental Metrics	Performance Metrics	Energy Saving
Theme 1: Energy Consumption Tracking Tools											
Budennyy et al. 2022 [1]	Empirical	Training	Transformer *(Malevich, Kandinsky)*	Single-site	Cloud; HPC	CPU *(AMD EPYC)*; GPU *(NVIDIA A100)*	PyTorch	PyNVML; psutil	Energy *(kWh)*; Carbon *(gCO2e)*	Accuracy *(via loss function)*	17%
Jääskeläinen et al. 2023 [2]	Theoretical	Inference	-	-	Local	GPU	-	-	-	-	-
Jesse et al. 2022 [3]	Empirical	Inference	LLM *(BERT)*	Single-site	Cloud	GPU *(NVIDIA TITAN)*	-	-	Energy *(kWh)*; Carbon *(gCO2e)*	-	-
Justus et al. 2018 [4]	Empirical	Training	CNN	Single-site	Cloud; Local	CPU; GPU *(NVIDIA V100, P100, K80, M60, GTX 1080 Ti)*	TensorFlow	-	Indirect *(time)*	RMSE	-
Kannan et al. 2024 [5]	Empirical	Inference	-	Single-site	-	-	PyTorch	-	Indirect *(num. of computations)*	Accuracy	-
Li et al. 2022 [6]	Empirical	Training	CNN	Single-site	Local	CPU *(Intel Core i9, AMD Ryzen 9)*; GPU *(NVIDIA RTX 3080)*	PyTorch	-	Energy *(J)*	RMSE; R-squared	-
Lannelongue et al. 2021 [7]	Empirical	Inference	LLM *(BERT)*	Single-site	HPC	CPU *(Intel Xeon)*; GPU *(NVIDIA V100)*; TPU *(TPU-v3 Pod)*	-	Green Algorithms	Energy *(kWh)*; Carbon *(gCO2e)*	-	-
Ma et al. 2022 [8]	Empirical	Training	-	Single-site	HPC; Local	-	-	-	Power *(W)*	Mean relative deviation	-
Santos et al. 2022 [9]	Empirical	Inference	CNN *(MobileNet, Inception,*	Single-site	HPC	CPU *(Intel Xeon)*; GPU *(NVIDIA GTX*	TensorFlow	AI Benchmark Alpha;	Energy *(kWh)*; Carbon *(gCO2e)*	Accuracy	-

			ResNet, Inception-ResNet, VGG-16)			*1080 Ti)*; HPC *(Dell PowerEdge R730)*		Experiment Impact Tracker			
Shaikh et al. 2021 [10]	Empirical	Training	-	-	-	-	-	Intel RAPL; NVIDIA-SMI	-	-	-
Taewon et al. 2023 [11]	Empirical	Inference	CNN	Single-site	-	CPU; GPU	TensorFlow	CodeCarbon	Carbon *(gCO2e)*	-	-
Tien-Ju et al. 2017 [12]	Empirical	Training	CNN	Single-site	-	-	-	-	-	Accuracy	-
Tiutiulnikov et al. 2023 [13]	Empirical	Training	CNN	Single-site	Cloud	CPU; GPU *(NVIDIA V100)*	PyTorch	-	Carbon intensity *(gCO2e/kWh)*	-	17.8% (CE)
Theme 2: Evaluates Environmental Impact of AI methods											
Caspart et al. 2022 [14]	Empirical	Training	RNN *(LSTM)*	Single-site	HPC	CPU; GPU *(NVIDIA A100)*	PyTorch	NVIDIA Management Library *(NVML)*	Energy *(J)*; Power *(W)*	Accuracy; Mean absolute percentage error *(MAPE)*	-
Castanyer et al. 2024 [15]	Empirical	Inference	CNN *(ResNet)*; RNN *(GRU, LSTM)*; Transformer	Single-site	-	CPU; GPU *(NVIDIA P100)*; Mobile *(Xiaomi Poco X3 NFC)*	PyTorch; TensorFlow; ONNX; TensorFlow Lite; Unity3D; Android Studio	Android Profiler; Unity Profiler	Indirect *(time)*	Accuracy	-
Dan et al. 2023 [16]	Empirical	Inference	LLM *(BERT, LLaMA 65B)*; CNN *(ResNet)*; Graph Neural Network *(DimeNet)*	Single-site	HPC	GPU *(NVIDIA A100, V100)*	Slurm	NVIDIA-SMI	Power *(W)*	training time; inference time*(latency)*	10-24%
del Rey et al. 2023 [17]	Empirical	Training	CNN *(MobileNetV2, NASNet Mobile, Xception, ResNet50, VGG16)*	Single-site	Cloud; Local	CPU; GPU *(NVIDIA GTX 750 Ti, RTX 3070, RTX 3090)*	TensorFlow	NVIDIA-SMI	Power *(W)*	F1 score	81%

Desislavov et al. 2023 [18]	Empirical	Inference	CNN *(AlexNet, VGG, ResNet, EfficientNet)*; Transformer*(BERT, MobileBERT)*	Single-site	-	GPU	PyTorch; TensorFlow	ptflops	Energy *(J)*	Accuracy	-
Emma et al. 2020 [19]	Empirical	Training	LLM *(BERT, GPT-2)*; RNN*(ELMo)*	Single-site	Cloud; HPC	-	TensorFlow ; Hyperopt	NVIDIA-SMI; Intel RAPL	Energy *(kWh)*; Power *(W)*; Carbon intensity *(gCO2e/kWh)*	-	-
Ferro et al. 2023 [20]	Empirical	Training	Traditional ML *(Decision Tree, Random Forest, XGBoost)*	Single-site	Local	CPU *(Intel Core i7)*; GPU *(NVIDIA RTX 2080 Ti)*	scikit-learn; pandas; NumPy	perf; PyRAPL	Energy *(J)*	Accuracy; MSE	53%, 62%, 135%, 162%
Gomez-Carmona et al. 2020 [21]	Empirical	Optimization	Traditional ML *(Logistic Regression, Random Forest, KNN, Naive Bayes, Linear SVM, Decision Trees)*; MLP	Single-site *(Edge)*	Local	CPU *(Intel Core i7, Rasberry Pi 3, Rasberry i Zero W)*	scikit-learn	-	Indirect *(time)*	F1 score; Precision; Recall; Accuracy	80% (time)
Jean-Quartier et al. 2023 [22]	Empirical	Optimization	CNN *(YOLO)*; Traditional ML *(Random Forest, Decision Tree)*	Single-site	Local	CPU	PyTorch; scikit-learn; SHAP; SciPy	CodeCarbon	Energy *(kWh)*	Accuracy; MSE; R-squared	-
Yu et al. 2022 [23]	Empirical	Inference	Traditional ML *(LR, KNN, SVM, RF, XGB)*; Neural Networks *(NN1, NN5, Pruned NN, Quantized NN)*	Single-site	Local	CPU *(Intel Core i5)*	PyTorch; scikit-learn	Intel Power Gadget	Power *(W)*	Accuracy; AUROC	-

Mazurek et al. 2024 [24]	Empirical	Data Management	CNN *(UNet)*	Single-site	HPC	CPU *(AMD EPYC 7742)*; GPU *(NVIDIA A100)*	PyTorch; MONAI; Lightning-Bagua; Weights & Biases	CodeCarbon	Energy *(J)*	DICE	400% (time)
McIntosh et al. 2019 [25]	Empirical	Training	Traditional ML *(Naive Bayes, J48, SMO, Logistic Regression, Random Forest, KNN, ZeroR)*; MLP	Single-site	-	Mobile *(Galaxy Nexus Android 4.2.2)*	WEKA;Neuroph; Android SDK; Dalvik VM	-	Energy *(J)*; Power *(W)*	Accuracy; RMSE	-
Nesrine et al. 2021 [26]	Empirical	Inference	LLM	Single-site	HPC; Local	CPU; GPU *(GTX 1080 Ti, V100)*	Hugging Face Transformers	CarbonTracker; Green Algorithms; ML CO2 Impact; Experiment Impact Tracker; Cumulator; Energy Usage	Energy *(kWh)*; Carbon *(gCO2e)*	-	-
Ollivier et al. 2023 [27]	Empirical	Inference	CNN *(AlexNet, VGG-16)*	Single-site *(Edge)*	Local	CPU; GPU; Supercomputer *(NVIDIA Jetson NX)*	PyTorch; TensorFlow	-	Power *(W)*; Carbon *(gCO2e)*	Throughput	-
Omar et al. 2024 [28]	Empirical	Data Management	Traditional ML *(SVM, Naive Bayes, Decision Tree, Hoeffding Tree, KNN, Random Forest, AdaBoost, Bagging Classifier)*	Single-site	Local	CPU *(Intel Xeon)*	scikit-multiflow; scikit-learn	CodeCarbon	Energy *(J)*	Drift detection closeness; True alarm rate	-

Rakshit et al. 2021 [29]	Empirical	Training	LLM *(BERT)*	Single-site	Local	CPU; GPU	PyTorch	CodeCarbon	Energy *(kWh)*; Carbon *(gCO2e)*	Accuracy	-
Raluca Maria et al. 2022 [30]	Empirical	Finetuning	-	Single-site	Local	CPU *(Rasberry Pi 3 and 4)*	ONNX; Docker; WebAssembly *(Wasm)*	High Voltage Power Monitor *(HVPM)* hardware by Monsoon	Energy *(J)*	Execution time	-
Tomlinson et al. 2024 [31]	Theoretical	Training	LLM	Single-site	-	GPU *(NVIDIA A100)*	-	-	Carbon *(gCO2e)* per task	-	-
Yokoyama et al. 2023 [32]	Empirical	Optimization	CNN; Traditional ML *(XGBoost, Random Forest, Decision Trees, K-means)*; MLP	Single-site	Local	CPU; GPU *(NVIDIA RTX)*; TPU; Supercomputer *(NVIDIA Jetson TX-2)*	Dask; scikit-learn; pandas; NumPy	perf; NVIDIA-SMI	Energy *(J)*; Carbon *(gCO2e)*	Accuracy	24%
Castellanos-Nieves et al. 2023 [33]	Empirical	Training	Traditional ML *(SGD Classifier, Logistic Regression, KNN, Random Forest, Gradient Boosting, Decision Tree, Linear & Non-Linear SVM, Gaussian Naive Bayes)*	Single-site	Local	CPU *(Intel Core i7)*; GPU	scikit-learn	mathematical model; psutil	Energy *(kWh)*; Carbon intesity *(gCO2e/kWh)*	Accuracy; F1 score	-
Dagoberto et al. 2024 [34]	Empirical	Training	CNN *(EfficientNet)*; RNN *(BiLSTM)*	Single-site	Cloud; Local	CPU; GPU	Keras; PyTorch; Ray Tune	-	Power *(W)*; Carbon *(gCO2e/kWh)*	Accuracy	28%

Toscano-Durán et al. 2024 [35]	Empirical	Data Management	CNN *(YOLO)*; MLP	Single-site	Local	CPU *(Intel Xeon Silver 4210)*; GPU *(NVIDIA Quadro RTX 4000)*	PyTorch	CodeCarbon	Indirect *(time)*; Carbon *(gCO2e)*	Accuracy; Precision; recall; F1 score; Mean Average Precision *(MAP)*	40% (CE)
Yarally et al. 2023 [36]	Empirical	Training	-	Single-site	Local	GPU *(NVIDIA GTX 1080)*	PyTorch	-	Energy *(J)*	Accuracy	-
Theme 3: Green Frameworks in Federated Learning											
Albaseer et al. 2024 [37]	Empirical	Optimization	CNN	Federated	Distributed	CPU; GPU	PyTorch	-	Energy *(J)*	-	17%
Fontenla-Romero et al. 2023 [38]	Empirical	Training	Neural Network *(Single-Layer Perceptron)*	Federated	Distributed	CPU *(Intel Core i7)*	-	-	Power *(Wh)*	Accuracy	87% (time)
Hsu et al. 2022 [39]	Empirical	Optimization	-	Federated	Distributed	CPU	-	Mathematical model	Energy *(J)*	RMSE	69%
Hu et al. 2022 [40]	Empirical	Training	CNN; MLP	Federated *(Edge)*	Distributed	CPU	-	Mathematical model	Energy *(J)*	Accuracy	-
Kim et al. 2024 [41]	Empirical	Training	CNN	Federated	Distributed	-	-	Energy model of processing chip	Energy *(J)*	Accuracy; Precision	70%
Kuswiradyo et al. 2024 [42]	Empirical	Training	-	Federated	Distributed	CPU *(Intel Core i5)*	IBM ILOG CPLEX Optimization Studio	-	Reduction in Energy *(%)*	Latency	11-27%
Maryam Ben et al. 2024 [43]	Empirical	Optimization	CNN	Federated	Distributed	GPU *(NVIDIA T4)*	Keras; TensorFlow	-	Energy *(J)*	Accuracy	43%
Peichun et al. 2021 [44]	Empirical	Data Management	CNN *(VGG-9)*	Federated *(Edge)*	Distributed	CPU	-	Mathematical model	Energy *(J)*	Accuracy	32%, 57%
Wang et al. 2023 [45]	Empirical	Training	Quantized Neural Network *(QNN)*	Federated	Distributed	-	MATLAB	Mathematical model	Energy *(J)*	Fronthaul Capacity *(Gbps)*; Level of precision *(bit)*	50%
Zhou et al. 2022 [46]	Empirical	Inference	-	Federated	Distributed	CPU	MATLAB	-	Energy *(J)*	-	-
Lee et al. 2023 [47]	Empirical	Training	CNN	Federated *(Edge)*	Distributed	CPU; GPU	PyTorch	-	Energy *(J)*	Structural Similarity	-

										Index Measure *(SSIM)*; Membership inference attack accuracy	
Qi et al. 2024 [48]	Empirical	Data Management	CNN	Federated	Distributed	CPU	PyTorch	-	Indirect *(overhead, latency)*	Accuracy	80% (comm. overhead)
Xia et al. 2024 [49]	Empirical	Training	CNN *(LeNet-5, AlexNet, MobileNetV)*	Federated	Distributed	CPU; GPU	-	-	Indirect *(computational complexity and overhead)*	Accuracy	-
Theme 4: Novel Techniques for Green AI											
Acmali et al. 2024 [50]	Empirical	Training	CNN	Single-site	Local	CPU; GPU	-	-	Indirect *(FLOPs)*	Accuracy	50% (model param)
Álvaro Domingo et al. 2024 [51]	Empirical	Training	CNN; Auto-Regressive Prediction Model	Single-site	Local	CPU; GPU	TensorFlow	NVIDIA-SMI	Energy *(kWh)*	Accuracy	56%
Candelieri et al. 2021 [52]	Empirical	Optimization	Traditional ML *(SVM)*	Single-site	HPC	CPU	-	-	Indirect *(time)*	Accuracy	66% (time)
Gille et al. 2022 [53]	Empirical	Data Management	CNN *(Convolutional AutoEncoder)*	Single-site	HPC	CPU; GPU *(NVIDIA A100)*	PyTorch	-	Reduction in Energy *(%)*	Peak SNR *(PSNR)*; Structural Similarity Index Measure *(SSIM)*	25%
Huang et al. 2023 [54]	Empirical	Finetuning	Transformer-Based LLM *(OPT, BLOOMZ, FLAN-T5)*	Single-site	-	GPU *(NVIDIA A100)*	PyTorch; TensorFlow	Torch Profiler	Indirect *(FLOPs)*	ROGUE scores	64% (FLOPs)
Lazzaro et al. 2023 [55]	Empirical	Training	-	Single-site	Local	CPU; GPU	-	-	Reduction in Energy *(%)*	-	7-31%
Liu et al. 2024 [56]	Empirical	Training	-	Single-site	Local	GPU *(NVIDIA RTX 3090)*	-	-	Carbon *(gCO2e)*	AUC; Mean Reciprocal Rank *(MRR)*; Normalized Discounted Cumulative Gain *(NDCG)*	87.6% (CO2)

Luis et al. 2024 [57]	Empirical	Training	MLP	Single-site	Local	-	sktime	-	Indirect *(num. of parameters)*	Mean Absolute Error *(MAE)*; RMSE; Mean Absolute Percentage Error *(MAPE)*; R-squared; Mean Squared Logarithmic Error *(MSLE)*; Median Absolute Percentage Error *(MdAPE)*; DIstortion Loss including shApe and TimE *(DILATE)*	20% (number of model parameters)
Nijkamp et al. 2024 [58]	Empirical	Inference	-	Single-site	Cloud	CPU	Datadog	-	Indirect *(CPU usage)*	Precision; Recall; F1 score	14%, 57%
Pirnat et al. 2022 [59]	Empirical	Training	CNN *(ResNet18)*	Single-site	Cloud	GPU	-	-	Energy (*J*); Carbon (*gCO2e*)	Mean Distance Error *(MDE)*; RMSE	94%
Ryan et al. 2017 [60]	Empirical	Training	MLP	Single-site	Local	CPU *(Intel Core i7, Xeon)*	-	-	Indirect *(num. of computations)*	Accuracy	95% (compute cost)
Shi et al. 2024 [61]	Empirical	Inference	LLM *(CodeBERT, GraphCodeBERT)*	Single-site	Local	CPU; GPU	Hugging Face Transformers	MLCO2 Impact	Energy (*kWh*); Carbon (*gCO2e*)	Accuracy; latency	99%
Thomas et al. 2020 [62]	Empirical	Training	Generative Models *(VAE, Flow)*	Single-site	Local	-	PyTorch	-	-	Bits per dimension *(BPD)*	90-94% (model size) del
Xiangyu et al. 2020 [63]	Theoretical	Inference	CNN *(GoogLeNet)*	Single-site	-	-	-	-	Power (*W*)	Signal-to-Interference-plus-Noise Ratio *(SINR)*	67%
Xiaokai et al. 2023 [64]	Empirical	Optimization	LLM	Single-site	Cloud	CPU; GPU	PyTorch	-	Carbon (*gCO2e*)	BLEU *(BiLingual Evaluation Understudy)*	28.8%, 48.4%, 55% (CO2)

										score; Accuracy	
Yang et al. 2023 [65]	Empirical	Data Management	CNN *(ResNet-20, ResNet-18, ResNet-50)*; LLM *(RoBERTa)*	Single-site	Local	GPU *(NVIDIA RTX A6000)*	-	-	-	Accuracy	60% (time)
Yin et al. 2024 [66]	Empirical	Data Management	CNN *(ResNet, DenseNet)*	Single-site	HPC; Local	GPU *(NVIDIA A100, RTX 3080)*	-	-	Indirect *(time)*	Accuracy	47%, 52%, 62%
Theme 5: New Carbon-aware AI frameworks											
Erik Johannes et al. 2024 [67]	Empirical	Optimization	CNN; RNN *(LSTM)*; Traditional ML *(Decision Tree, Random Forest, XGBoost, Gradient Boost)*	Single-site	Cloud; Local	CPU; GPU	PyTorch; TensorFlow	CodeCarbon	Energy *(kWh)*; Carbon *(gCO2e)*	F1 score; R-squared	-
Gutierrez et al. 2023 [68]	Empirical	Optimization	CNN; RNN *(LSTM)*; Traditional ML *(Decision Tree, Random Forest, XGBoost, Gradient Boost)*	Single-site	Local	CPU	Scikit	Energy Efficiency Tester *(hardware)*	Power *(W/s)*	-	12%
Li et al. 2023 [69]	Empirical	Inference	LLM *(ALBERT)*; CNN *(EfficientNet, YOLOv5)*	Single-site	Cloud; HPC	CPU *(ADM EPYC)*; GPU *(NVIDIA A100)*	PyTorch; Flask; NetworkX	Carbon Tracker	Carbon *(gCO2e)* per day	Accuracy; Latency	80%
Meilin et al. 2023 [70]	Theoretical	Training	-	-	-	-	-	-	-	-	-
Henderson et al. 2021 [71]	Empirical	Inference	-	Single-site	HPC	CPU *(Intel Xeon)*; GPU *(NVIDIA A100, V100)*	PyTorch; TensorFlow	RAPL; Intel Power Gadget; NVIDIA-SMI; psutil	Energy *(kWh)*; Carbon intesity *(gCO2e/kWh)*	BLEU *(BiLingual Evaluation Understudy)* score	-

Roberto et al. 2024 [72]	Empirical	Finetuning	Traditional ML *(Linear Regression, Decision Tree, Random Forest, XGBoost)*; Neural Network	Single-site	Local	CPU	-	-	Carbon *(gCO2e)*	Accuracy	-
Salh et al. 2023 [73]	Empirical	Training	-	Federated	Distributed	CPU	-	-	Energy *(J)*	Time	-
Sunxuan et al. 2024 [74]	Empirical	Training	-	Federated	Distributed	CPU	-	-	Energy *(J)*	Accuracy *(via loss function)*; Differential privacy cost	11%
Zhaohui et al. 2020 [75]	Theoretical	Training	-	Federated	Distributed	-	-	-	Reduction in Energy *(%)*	Maximum average transmit power in dB	59%

References:

[1] Budennyy, S. A. et al. eco2AI: Carbon Emissions Tracking of Machine Learning Models as the First Step Towards Sustainable AI. Doklady Mathematics 106, S118–S128 (2022).

[2] Jääskeläinen, P. Explainable Sustainability for AI in the Arts. in The 1st International Workshop on Explainable AI for the Arts, ACM Creativity and Cognition Conference (2023).

[3] Dodge, J. et al. Measuring the Carbon Intensity of AI in Cloud Instances. in ACM Conference on Fairness Accountability and Transparency 1877–1894 (ACM, 2022). doi:10.1145/3531146.3533234.

[4] Justus, D., Brennan, J., Bonner, S. & McGough, A. S. Predicting the Computational Cost of Deep Learning Models. in IEEE International Conference on Big Data (Big Data) 3873–3882 (IEEE, 2018). doi:10.1109/BigData.2018.8622396.

[5] Kannan, J., Barnett, S., Simmons, A., Selvi, T. & Cruz, L. Green Runner: A Tool for Efficient Deep Learning Component Selection. in Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering - Software Engineering for AI 112–117 (ACM, 2024). doi:10.1145/3644815.3644942.

[6] Li, C., Tsourdos, A. & Guo, W. A Transistor Operations Model for Deep Learning Energy Consumption Scaling Law. IEEE Transactions on Artificial Intelligence 5, 192–204 (2024).

[7] Lannelongue, L., Grealey, J. & Inouye, M. Green Algorithms: Quantifying the Carbon Footprint of Computation. Advanced Science 8, (2021).

[8] Ma, H. & Ding, A. Method for evaluation on energy consumption of cloud computing data center based on deep reinforcement learning. Electric Power Systems Research 208, 107899 (2022).

[9] Carastan-Santos, D. & Pham, T. H. T. Understanding the Energy Consumption of HPC Scale Artificial Intelligence. in High Performance Computing. CARLA 2022. Communications in Computer and Information Science, vol 1660 131–144 (Springer, Cham, 2022). doi:10.1007/978-3-031-23821-5_10.
[10] Shaikh, O. et al. EnergyVis: Interactively Tracking and Exploring Energy Consumption for ML Models. in Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems 1–7 (ACM, 2021). doi:10.1145/3411763.3451780.
[11] Yoo, T., Lee, H., Oh, S., Kwon, H. & Jung, H. Visualizing the Carbon Intensity of Machine Learning Inference for Image Analysis on TensorFlow Hub. in Computer Supported Cooperative Work and Social Computing 206–211 (ACM, 2023). doi:10.1145/3584931.3606959.
[12] Yang, T.-J., Chen, Y.-H., Emer, J. & Sze, V. A method to estimate the energy consumption of deep neural networks. in 51st Asilomar Conference on Signals, Systems, and Computers 1916–1920 (IEEE, 2017). doi:10.1109/ACSSC.2017.8335698.
[13] Tiutiulnikov, M. et al. eco4cast: Bridging Predictive Scheduling and Cloud Computing for Reduction of Carbon Emissions for ML Models Training. Doklady Mathematics 108, S443–S455 (2023).
[14] Caspart, R. et al. Precise Energy Consumption Measurements of Heterogeneous Artificial Intelligence Workloads. in International Conference on High Performance Computing 108–121 (2022). doi:10.1007/978-3-031-23220-6_8.
[15] Castanyer, R. C., Martínez-Fernández, S. & Franch, X. Which design decisions in AI-enabled mobile applications contribute to greener AI? Empirical Software Engineering 29, 2 (2024).
[16] Zhao, D. et al. Sustainable Supercomputing for AI. in Proceedings of the 2023 ACM Symposium on Cloud Computing 588–596 (ACM, 2023). doi:10.1145/3620678.3624793.
[17] del Rey, S., Martínez-Fernández, S., Cruz, L. & Franch, X. Do DL models and training environments have an impact on energy consumption? in 49th Euromicro Conference on Software Engineering and Advanced Applications (SEAA) 150–158 (IEEE, 2023). doi:10.1109/SEAA60479.2023.00031.
[18] Desislavov, R., Martínez-Plumed, F. & Hernández-Orallo, J. Trends in AI inference energy consumption: Beyond the performance-vs-parameter laws of deep learning. Sustainable Computing: Informatics and Systems 38, 100857 (2023).
[19] Strubell, E., Ganesh, A. & McCallum, A. Energy and Policy Considerations for Modern Deep Learning Research. in Proceedings of the AAAI Conference on Artificial Intelligence vol. 34 13693–13696 (2020).
[20] Ferro, M., Silva, G. D., de Paula, F. B., Vieira, V. & Schulze, B. Towards a sustainable artificial intelligence: A case study of energy efficiency in decision tree algorithms. Concurrency and Computation: Practice and Experience 35, (2023).
[21] Gómez-Carmona, O., Casado-Mansilla, D., Kraemer, F. A., López-de-Ipiña, D. & García-Zubia, J. Exploring the computational cost of machine learning at the edge for human-centric Internet of Things. Future Generation Computer Systems 112, 670–683 (2020).
[22] Jean-Quartier, C. et al. The Cost of Understanding—XAI Algorithms towards Sustainable ML in the View of Computational Cost. Computation 11, 92 (2023).
[23] Yu, J.-R. et al. Energy Efficiency of Inference Algorithms for Clinical Laboratory Data Sets: Green Artificial Intelligence Study. Journal of Medical Internet Research 24, e28036 (2022).
[24] Mazurek, S., Pytlarz, M., Malec, S. & Crimi, A. Investigation of Energy-Efficient AI Model Architectures and Compression Techniques for “Green” Fetal Brain Segmentation. in Computational Science – ICCS 2024. ICCS 2024. Lecture Notes in Computer Science, vol 14835 61–74 (Springer, Cham, 2024). doi:10.1007/978-3-031-63772-8_5.
[25] McIntosh, A., Hassan, S. & Hindle, A. What can Android mobile app developers do about the energy consumption of machine learning? Empirical Software Engineering 24, 562–601 (2019).
[26] Bannour, N., Ghannay, S., Névéol, A. & Ligozat, A.-L. Evaluating the carbon footprint of NLP methods: a survey and analysis of existing tools. in Proceedings of the Second Workshop on Simple and Efficient Natural Language Processing 11–21 (Association for Computational Linguistics, 2021). doi:10.18653/v1/2021.sustainlp-1.2.
[27] Ollivier, S. et al. Sustainable AI Processing at the Edge. IEEE Micro 43, 19–28 (2023).

[28] Omar, R. et al. How to Sustainably Monitor ML-Enabled Systems? Accuracy and Energy Efficiency Tradeoffs in Concept Drift Detection. in 10th International Conference on ICT for Sustainability (ICT4S) 172–182 (IEEE, 2024). doi:10.1109/ICT4S64576.2024.00026.
[29] Naidu, R. et al. Towards Quantifying the Carbon Emissions of Differentially Private Machine Learning. in Workshop on Socially Responsible Machine Learning, The International Conference on Machine Learning (ICML) (2021). doi:10.48550/arxiv.2107.06946.
[30] Hampau, R. M., Kaptein, M., van Emden, R., Rost, T. & Malavolta, I. An empirical study on the Performance and Energy Consumption of AI Containerization Strategies for Computer-Vision Tasks on the Edge. in The International Conference on Evaluation and Assessment in Software Engineering 2022 50–59 (ACM, 2022). doi:10.1145/3530019.3530025.
[31] Tomlinson, B., Black, R. W., Patterson, D. J. & Torrance, A. W. The carbon emissions of writing and illustrating are lower for AI than for humans. Scientific Reports 14, 3732 (2024).
[32] Yokoyama, A. M., Ferro, M., de Paula, F. B., Vieira, V. G. & Schulze, B. Investigating hardware and software aspects in the energy consumption of machine learning: A green AI-centric analysis. Concurrency and Computation: Practice and Experience 35, (2023).
[33] Castellanos-Nieves, D. & García-Forte, L. Improving Automated Machine-Learning Systems through Green AI. Applied Sciences 13, 11583 (2023).
[34] Castellanos-Nieves, D. & García-Forte, L. Strategies of Automated Machine Learning for Energy Sustainability in Green Artificial Intelligence. Applied Sciences 14, 6196 (2024).
[35] Perera-Lago, J. et al. An in-depth analysis of data reduction methods for sustainable deep learning. Open Research Europe 4, 101 (2024).
[36] Yarally, T., Cruz, L., Feitosa, D., Sallou, J. & van Deursen, A. Uncovering Energy-Efficient Practices in Deep Learning Training: Preliminary Steps Towards Green AI. in IEEE/ACM 2nd International Conference on AI Engineering – Software Engineering for AI (CAIN) 25–36 (IEEE, 2023). doi:10.1109/CAIN58948.2023.00012.
[37] Albaseer, A., Seid, A. M., Abdallah, M., Al-Fuqaha, A. & Erbad, A. Novel Approach for Curbing Unfair Energy Consumption and Biased Model in Federated Edge Learning. IEEE Transactions on Green Communications and Networking 8, 865–877 (2024).
[38] Fontenla-Romero, O., Guijarro-Berdiñas, B., Hernandez-Pereira, E. & Perez-Sanchez, B. An effective and efficient green federated learning method for one-layer neural networks. in Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing 1050–1052 (ACM, 2024). doi:10.1145/3605098.3636144.
[39] Hsu, Y.-L., Liu, C.-F., Wei, H.-Y. & Bennis, M. Optimized Data Sampling and Energy Consumption in IIoT: A Federated Learning Approach. IEEE Transactions on Communications 70, 7915–7931 (2022).
[40] Hu, Y., Huang, H. & Yu, N. Resource Optimization and Device Scheduling for Flexible Federated Edge Learning with Tradeoff Between Energy Consumption and Model Performance. Mobile Networks and Applications 27, 2118–2137 (2022).
[41] Kim, M., Saad, W., Mozaffari, M. & Debbah, M. Green, Quantized Federated Learning Over Wireless Networks: An Energy-Efficient Design. IEEE Transactions on Wireless Communications 23, 1386–1402 (2024).
[42] Kuswiradyo, P., Kar, B. & Shen, S.-H. Optimizing the energy consumption in three-tier cloud–edge–fog federated systems with omnidirectional offloading. Computer Networks 250, 110578 (2024).
[43] Driss, M. B., Sabir, E., Elbiaze, H., Diallo, A. & Sadik, M. A Green Multi-Attribute Client Selection for Over-The-Air Federated Learning: A Grey-Wolf-Optimizer Approach. ACM Transactions on Modeling and Performance Evaluation of Computing Systems 10, 1–24 (2025).
[44] Li, P., Huang, X., Pan, M. & Yu, R. FedGreen: Federated Learning with Fine-Grained Gradient Compression for Green Mobile Edge Computing. in IEEE Global Communications Conference (GLOBECOM) 1–6 (IEEE, 2021). doi:10.1109/GLOBECOM46510.2021.9685582.
[45] Wang, J., Mao, Y., Wang, T. & Shi, Y. Green Federated Learning Over Cloud-RAN With Limited Fronthaul Capacity and Quantized Neural Networks. IEEE Transactions on Wireless Communications 23, 4300–4314 (2024).
[46] Zhou, X., Zhao, J., Han, H. & Guet, C. Joint Optimization of Energy Consumption and Completion Time in Federated Learning. in IEEE 42nd International Conference on Distributed Computing Systems (ICDCS) 1005–1017 (IEEE, 2022). doi:10.1109/ICDCS54860.2022.00101.
[47] Lee, J., Seif, M., Cho, J. & Poor, H. V. Exploring the Privacy-Energy Consumption Tradeoff for Split Federated Learning. IEEE Network 38, 388–395 (2024).

[48] Qi, Y. & Hossain, M. S. Harnessing federated generative learning for green and sustainable Internet of Things. Journal of Network and Computer Applications 222, 103812 (2024).

[49] Xia, C., Tsang, D. H. K. & Lau, V. K. N. Structured Bayesian Federated Learning for Green AI: A Decentralized Model Compression Using Turbo-VBI-Based Approach. IEEE Internet of Things Journal 11, 12783–12798 (2024).

[50] Acmali, S. S., Ortakci, Y. & Seker, H. Green AI-Driven Concept for the Development of Cost-Effective and Energy-Efficient Deep Learning Method: Application in the Detection of Eimeria Parasites as a Case Study. Advanced Intelligent Systems 6, (2024).

[51] Reguero, Á. D., Martínez-Fernández, S. & Verdecchia, R. Energy-efficient neural network training through runtime layer freezing, model quantization, and early stopping. Computer Standards & Interfaces 92, 103906 (2025).

[52] Candelieri, A., Perego, R. & Archetti, F. Green machine learning via augmented Gaussian processes and multi-information source optimization. Soft Computing 25, 12591–12603 (2021).

[53] Gille, C., Guyard, F., Antonini, M. & Barlaud, M. Learning Sparse auto-Encoders for Green AI image coding. in ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 1–5 (IEEE, 2023). doi:10.1109/ICASSP49357.2023.10096638.

[54] Huang, K., Yin, H., Huang, H. & Gao, W. Towards Green AI in Fine-tuning Large Language Models via Adaptive Backpropagation. in The Twelfth International Conference on Learning Representations (2024). doi:10.48550/arXiv.2309.13192.

[55] Lazzaro, D. et al. Minimizing Energy Consumption of Deep Learning Models by Energy-Aware Training. in Image Analysis and Processing – ICIAP 2023. Lecture Notes in Computer Science, vol 14234 515–526 (Springer, Cham, 2023). doi:10.1007/978-3-031-43153-1_43.

[56] Liu, Q., Zhu, J., Dai, Q. & Wu, X.-M. Benchmarking News Recommendation in the Era of Green AI. in Companion Proceedings of the ACM Web Conference 2024 971–974 (ACM, 2024). doi:10.1145/3589335.3651472.

[57] Balderas, L., Lastra, M. & Benítez, J. M. An Efficient Green AI Approach to Time Series Forecasting Based on Deep Learning. Big Data and Cognitive Computing 8, 120 (2024).

[58] Nijkamp, N., Sallou, J., van der Heijden, N. & Cruz, L. Green AI in Action: Strategic Model Selection for Ensembles in Production. in Proceedings of the 1st ACM International Conference on AI-Powered Software 50–58 (ACM, 2024). doi:10.1145/3664646.3664763.

[59] Pirnat, A. et al. Towards Sustainable Deep Learning for Wireless Fingerprinting Localization. in ICC 2022 - IEEE International Conference on Communications 3208–3213 (IEEE, 2022). doi:10.1109/ICC45855.2022.9838464.

[60] Spring, R. & Shrivastava, A. Scalable and Sustainable Deep Learning via Randomized Hashing. in Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining vol. Part F129685 445–454 (ACM, 2017).

[61] Shi, J. et al. Greening Large Language Models of Code. in Proceedings of the 46th International Conference on Software Engineering: Software Engineering in Society 142–153 (ACM, 2024). doi:10.1145/3639475.3640097.

[62] Bird, T., Kingma, F. H. & Barber, D. Reducing the Computational Cost of Deep Generative Models with Binary Neural Networks. in International Conference on Learning Representations (ICLR) (2021). doi:10.48550/arXiv.2010.13476.

[63] Yang, X. et al. Sparse Optimization for Green Edge AI Inference. Journal of Communications and Information Networks 5, 1–15 (2020).

[64] Wei, X. et al. Towards Greener Yet Powerful Code Generation via Quantization: An Empirical Study. in Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering 224–236 (ACM, 2023). doi:10.1145/3611643.3616302.

[65] Yang, Y., Kang, H. & Mirzasoleiman, B. Towards Sustainable Learning: Coresets for Data-efficient Deep Learning. in Proceedings of the 40th International Conference on Machine Learning, PMLR 202 39314–39330 (2023). doi:10.48550/arXiv.2306.01244.

[66] Yin, Z., Pu, J., Wan, R. & Xue, X. Embrace sustainable AI: Dynamic data subset selection for image classification. Pattern Recognition 151, 110392 (2024).

[67] Husom, E. J., Sen, S. & Goknil, A. Engineering Carbon Emission-aware Machine Learning Pipelines. in Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering - Software Engineering for AI 118–128 (ACM, 2024). doi:10.1145/3644815.3644943.

[68] Gutiérrez, M., Moraga, M. Á., García, F. & Calero, C. Green-IN Machine Learning at a Glance. Computer 56, 35–43 (2023).
[69] Li, B., Samsi, S., Gadepally, V. & Tiwari, D. Clover: Toward Sustainable AI with Carbon-Aware Machine Learning Inference Service. in Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis 1–15 (ACM, 2023). doi:10.1145/3581784.3607034.
[70] Shi, M. et al. Thinking Geographically about AI Sustainability. AGILE: GIScience Series 4, 1–7 (2023).
[71] Henderson, P. et al. Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning. The Journal of Machine Learning Research 21, 10039–10081 (2020).
[72] Pagliari, R. et al. A Comprehensive Sustainable Framework for Machine Learning and Artificial Intelligence. in 27th European Conference on Artificial Intelligence (volume 392) 834–841 (2024). doi:10.48550/arXiv.2407.12445.
[73] Salh, A. et al. Energy-Efficient Federated Learning With Resource Allocation for Green IoT Edge Intelligence in B5G. IEEE Access 11, 16353–16367 (2023).
[74] Zhang, S. et al. Differential Privacy-Aware Generative Adversarial Network-Assisted Resource Scheduling for Green Multi-Mode Power IoT. IEEE Transactions on Green Communications and Networking 8, 956–967 (2024).
[75] Yang, Z., Chen, M., Saad, W., Hong, C. S. & Shikh-Bahaei, M. Energy Efficient Federated Learning Over Wireless Communication Networks. IEEE Transactions on Wireless Communications 20, 1935–1949 (2021).

Tools mentioned in the Systematic Review

<table>
<tr><th>Tool</th><th>#Occurrences</th><th colspan="2">Function</th></tr>
<tr><td colspan="4">Experiment Tracking & Monitoring Tools</td></tr>
<tr><td>Ai Benchmark Alpha</td><td>1</td><td colspan="2">Benchmarking tool (python-based) for AI algorithms in HPC setting to compare energy and accuracy outcomes.</td></tr>
<tr><td>Code Carbon</td><td>7</td><td colspan="2">Python package that computes carbon emissions based on time, power (GPU + CPU + RAM), and carbon intensity.</td></tr>
<tr><td>ML CO2 Impact</td><td>2</td><td colspan="2">Online tool that computes carbon emissions based on user-provided information such as runtime, hardware, cloud provider and location.</td></tr>
<tr><td>Carbon Tracker</td><td>2</td><td colspan="2">Python package. (same formula as MLCO2)</td></tr>
<tr><td>Experiment Impact Tracker</td><td>2</td><td colspan="2">Python package. ((same formula as MLCO2)</td></tr>
<tr><td>Green Algorithms</td><td>2</td><td colspan="2">Convenient online tool. (same formula as MLCO2)</td></tr>
<tr><td>Cumulator</td><td>1</td><td colspan="2">Python package that estimates carbon emissions based on runtime, GPU load and carbon intensity as well as communication energy.</td></tr>
<tr><td>Energy usage tool</td><td>1</td><td colspan="2">Python package (possibly buggy)</td></tr>
<tr><td colspan="4">Profiling & Performance Tools</td></tr>
<tr><td>NVIDIA-SMI</td><td>7</td><td>command-line tool</td><td rowspan="3">Monitor NVIDIA GPUs</td></tr>
<tr><td>NVML</td><td>1</td><td>C-based API</td></tr>
<tr><td>pyNVML</td><td>1</td><td>Python library</td></tr>
<tr><td>Intel Power Gadget</td><td>2</td><td>Software app</td><td rowspan="3">Monitor Intel CPU power usage</td></tr>
<tr><td>Intel RAPL</td><td>3</td><td>Hardware feature</td></tr>
<tr><td>pyRAPL</td><td>1</td><td>Python library</td></tr>
<tr><td>perf</td><td>2</td><td colspan="2">Command-line profiler tool in Linux to monitor CPU performance and system events.</td></tr>
<tr><td>psutil</td><td>3</td><td colspan="2">Python library for retrieving information on system utilization (CPU, memory, disks etc.)</td></tr>
<tr><td>ptflops</td><td>1</td><td colspan="2">Python library for calculating the number of FLOPs/computations during inference only.</td></tr>
<tr><td>PyTorch Profiler</td><td>1</td><td colspan="2">Python library for profiling time and memory usage by CPU/GPU.</td></tr>
<tr><td>Android Profiler</td><td>1</td><td colspan="2">Built-in Android Studio tool for monitoring CPU usage, memory, device power consumption, etc.</td></tr>
<tr><td>Unity Profiler</td><td>1</td><td colspan="2">Built-in Unity engine tool to get performance information on app’s CPU, GPU, memory, and rendering.</td></tr>
</table>