A Systematic Review of Green AI

Roberto Verdecchia
University of Florence
Florence, Italy
roberto.verdecchia@unifi.it

June Sallou
TU Delft
Delft, The Netherlands
j.sallou@tudelft.nl

Luís Cruz
TU Delft
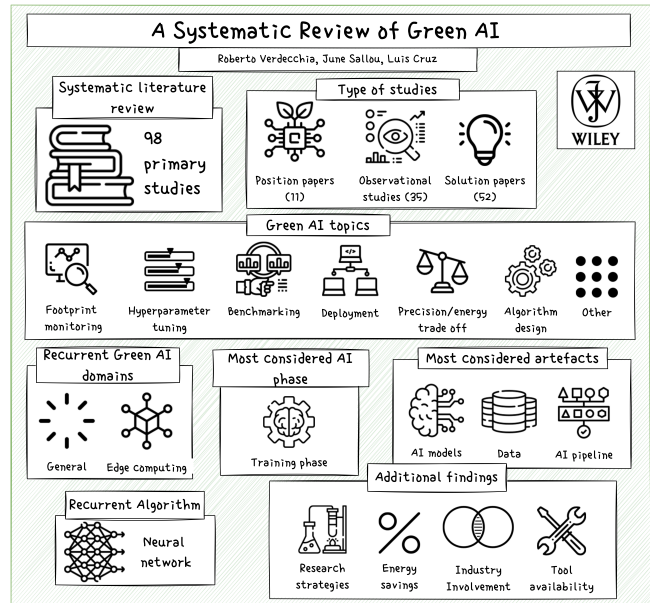Delft, The Netherlands
l.cruz@tudelft.nl

Abstract

With the ever-growing adoption of AI-based systems, the carbon footprint of AI is no longer negligible. AI researchers and practitioners are therefore urged to hold themselves accountable for the carbon emissions of the AI models they design and use. This led in recent years to the appearance of researches tackling AI environmental sustainability, a field referred to as Green AI. Despite the rapid growth of interest in the topic, a comprehensive overview of Green AI research is to date still missing. To address this gap, in this paper, we present a systematic review of the Green AI literature. From the analysis of 98 primary studies, different patterns emerge. The topic experienced a considerable growth from 2020 onward. Most studies consider monitoring AI model footprint, tuning hyperparameters to improve model sustainability, or benchmarking models. A mix of position papers, observational studies, and solution papers are present. Most papers focus on the training phase, are algorithm-agnostic or study neural networks, and use image data. Laboratory experiments are the most common research strategy. Reported Green AI energy savings go up to 115%, with savings over 50% being rather common. Industrial parties are involved in Green AI studies, albeit most target academic readers. Green AI tool provisioning is scarce. As a conclusion, the Green AI research field results to have reached a considerable level of maturity. Therefore, from this review emerges that the time is suitable to adopt other Green AI research strategies, and port the numerous promising academic results to industrial practice.

Graphical Abstract: From a systematic review of the Green AI literature, Green AI results to focus on solutions, and is often not bound to a specific context or algorithm. The Green AI research field results to be mature, *i.e.*, the moment is suitable to port results from academic research to industrial practice.

1 Introduction

In recent years, the Artificial Intelligence (AI) community has been challenged to bring the carbon footprint of AI models to the top of their research agenda. The iconic paper by Strubell et al. [97] analyzes the carbon impact of training their own state-of-the-art models. Results lead to the conclusion that we need to reduce the carbon footprint of developing and running AI models.

This self-reflection was an eye-opener to the AI research community. Many papers followed, calling for a new research direction that would consider this problem. Schwartz et al. coined the term Green AI as *"AI research that yields novel results while taking into account the computational cost"* [92]. Bender et al. published a position paper highlighting the consequences of continuously increasing the size of AI models [31]. A natural question that is posed is whether we are doing enough as a research community to mitigate the carbon impact of developing and running AI-based software.

AI systems are significantly complex and, to achieve Green AI, we need a joint effort that targets all the different stages of an AI

system's lifecycle (e.g, data collection, training, monitoring), different artifacts (*e.g.*, data, model, pipeline, architecture, hardware), etc [3].

Given the heterogeneity of the field, it is also difficult to have a broad view of all the Green AI literature that has been published in the past years. To understand the existing research, we conduct a systematic literature review on Green AI. We provide an overview and characterization of the existing research in this field. Moreover, we study how the field has been evolving over the years, pinpoint the main topics, approaches, artifacts, and so on.

This literature review shows that there has been a significant growth in Green AI publications – 76% of the papers have been published since 2020. The most popular topics revolve around monitoring, hyperparameter tuning, deployment, and model benchmarking. We also highlight other emerging topics that might lead to interesting solutions – namely, *Data Centric* Green AI, *Precision/Energy Trade-off* analysis. The current body of research has already showcased promising results with energy savings from 13% up to 115%. Still, most of the existing work focuses on the training stage of the AI model. Moreover, we observe that there is little involvement

of the industry (23%) and that most studies revolve around laboratory experiments. We argue that the field is growing to a level of maturity in which involvement of the industry is quintessential to enable the overarching goal of Green AI: harness the full potential of AI without a negative impact in our planet.

To encourage open science and the reproducibility of this study, we provide all data and scripts in a replication package available online with an open-source license[1].

The remainder of this paper is structured as follows. In Section 2, we describe the methodology used to collect and analyze Green AI literature. In Section 3, we present all the results yielded by our methodology. Section 4 discusses findings and reflects on the impact of our results in the research community. In Section 5, we reflect on the potential threats to the validity of this study. Following, Section 6 describes related work and pinpoints the differences with our study. The main conclusions and future work are presented in Section 7.

2 Methodology

In this section, we document the research design, which was rigorously adhered to during the planning and execution of the study. We primarily followed the guidelines for conducting SLRs in software engineering research presented by Kitchenham [6].

2.1 Research Objective and Question

The goal of this review is to understand the characteristics of existing Green AI research. By utilizing the Goal-Question-Metric method [1], this objective can be described more formally as follows:

Analyze Green AI literature
For the purpose of knowledge collection and categorization
With respect to AI
From the viewpoint of researchers and practitioners
In the context of environmental sustainability.

The goal of this research can be directly translated into a research question (RQ), which states as follows:

RQ: *What are the characteristics of Green AI state-of-the-art research?*

By answering our research question, we aim at gaining a systematic overview of the Green AI body of knowledge, starting from an outline of the general publication trends, to a detailed analysis of the past and current Green AI research activities and their characteristics.

2.2 Research Process

An overview of the research process followed is depicted in Figure 1. The process starts with the execution of a conservative automated search query via the digital libraries and indexing platforms *Google Scholar*, *Scopus*, and *Web of Science*, complemented by a subsequent iterative bidirectional snowballing process, which is conducted until the achievement of theoretical saturation. Including multiple literature indexing platforms to execute the automated search allows us to conduct an encompassing search of the literature based on multiple sources, hence allowing us to mitigate potential threats

to external validity, as further documented in Section 5. Following, the details of each step of our research process are documented in detail.

2.2.1 Automated Initial Search. To identify a preliminary set of potentially relevant research works, we design an encompassing automated query to be executed on three different literature indexers, namely *Google Scholar*, *Scopus*, and *Web of Science*. The automated query targeting publication titles states as follows:

Listing 1: Automated search query

```
1  INTITLE("green" OR "sustainab*") AND
2  INTITLE("AI" OR "ML" OR "artificial_intelligence"
3  OR "machine_learning" OR "deep_learning")
```

The query is designed to retrieve literature with titles containing keywords related to sustainability, identified by the keywords *green* or *sustainability* and its variations, *e.g.,* "sustainable" (Listing 1, Lines 1). The second part of the query instead is used to retrieve literature concerning AI, or related synonyms and acronyms (Listing 1, Lines 2-3). The query is executed on the three aforementioned literature libraries and indexes on the 18th of July 2022, and led to the identification of 190 potentially relevant studies. In order to be as comprehensive as possible, and avoid potential threats to external validity, the year of publication is left unbounded in the automated search.

2.2.2 Application of Selection Criteria. Subsequent to the identification of the initial potentially relevant studies, we execute the manual selection of the studies via a set of selection criteria defined *a priori*. A paper is confirmed as primary study if it adheres to all inclusion criteria, and none of the exclusion ones. The following inclusion (I) and exclusion (E) criteria are used:

I1- The study regards AI
I2- The study regards environmental sustainability
I3- The study regards the environmental sustainability of AI
I4- The study regards the software level
E1- The study is not written in English
E2- The study is not available
E3- The study is a duplicate or extensions of an already included study
E4- The study is a secondary or tertiary study
E5- The study is in the form of editorials, tutorials, books, extended abstracts, etc.
E6- The study is a non-scientific publication or grey literature

With the first three inclusion criteria (I1-I3), we ensure that the primary studies focus on Green AI (I1, I2), and that the studies regard the environmental sustainability *of* AI, rather than the improvement of environmental sustainability *through* AI. With the fourth inclusion criterion instead (I4), we ensure that the primary studies focus on software-centric Green AI. This latter criterion is used to exclude studies focusing on hardware-specific Green AI techniques, *e.g.,* the use of *ad hoc* implemented hardware components, which we consider out of reach for most researchers/practitioners interested in Green AI, and is only marginal to the definition of Green AI itself [92].

The exclusion criteria are designed to ensure that data can be extracted from the papers (E1, E2), do not represent duplication or

---

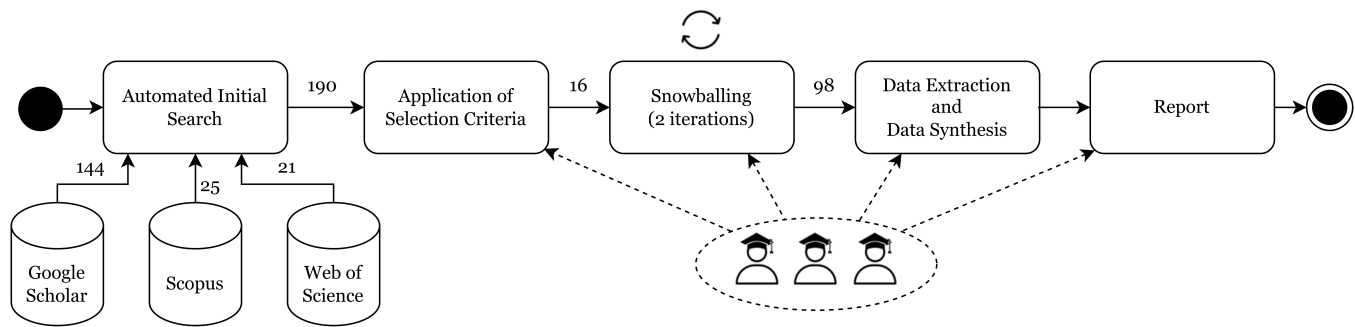[1]Replication package: https://github.com/luiscruz/slr-green-ai

Figure 1: Systematic literature review process overview.

redundancy with respect to other primary studies (E3, E4), and are provided in the form of scientific studies (E5, E6).

To ease the primary study selection process, adaptive reading depth [13] is used to efficiently assess potentially relevant studies. In order to mitigate subjective biases and interpretations, the three authors independently utilized the selection criteria to scrutinize 63-64 candidate studies. Weekly meeting are held during the selection process to jointly discuss examples, doubts, and align the selection process between the three researchers.

The application of the selection criteria concludes with the identification of 16 primary studies, which constitute the starting set for the subsequent snowballing process.

2.2.3 Snowballing. In order to enrich the set of selected primary studies, and ensure that the primary study comprehensively represents the Green AI body of literature, the automated search results are complemented with a recursive bidirectional snowballing process [21]. This step entails the scrutiny of all studies either citing or cited by the already included primary studies. As for the application of selection criteria, three researchers are involved in the snowballing. During each snowballing round, the researchers independently snowball different primary studies, and propose new primary studies to be included, *i.e.,* the new identified studies which adhere to the selection criteria. During each snowballing round, examples, doubts, and divergences are jointly revisited and resolved, and the next snowballing iteration is started. A total of two rounds of backward and forward snowballing are executed before no new studies are identified, *i.e.,* when theoretical saturation is reached. The snowballing process terminates with the inclusion of 82 new primaries studies, leading to a total of 98 primary studies which are considered in the literature review reported in this research.

2.2.4 Data Extraction. In order to achieve the intended goal of this study and answer our RQ (see Section 2.1), we proceed to systematically extract data from the primary studies. The data extraction process consisted of two subsequent phases.

The first phase consists of a data exploration process, which terminates with the establishment of the data extraction framework of this study. Specifically, during this first phase, the three authors of this review independently scan the identified primary studies, and annotate the characteristics of the studies which are relevant to answer our RQ. The identified characteristics are then jointly discussed and refined, leading to the consolidation of the fields constituting the data extraction framework of this review.

In the second data extraction phase, the primary studies are thoroughly analyzed, and the data is extracted from the studies according to the data extraction framework.

The fields of the data extraction framework utilized for this literature review on Green AI are the following.

- *Green AI Definition:* the level of abstraction used in the paper to quantify the impact of AI in the surrounding environment: energy efficiency [18], carbon footprint [20], or ecological footprint [9].
- *Study type:* The overarching type of study, which could be either presenting a position on Green AI, a Green AI solution, or an observational study on Green AI;
- *Topic:* The Green AI topic considered in the study, *e.g.,* hyperparameter-tuning to achieve energy efficiency of an AI algorithm;
- *Domain:* The domain considered in the study, *e.g.,* edge or mobile computing;
- *Type of data:* The type of data utilized by AI in the study, *e.g.,* text or images;
- *Artifact considered:* The AI artifact considered in the study, *e.g.,* the data used by AI models, the AI models themselves, or the AI deployment pipeline.
- *Considered phase:* If the study focused on the AI training phase, the AI inference phase, or both.
- *Research strategy:* The research strategy, as defined in [16], used to support the claims reported in the study;
- *Dataset size:* The size of the dataset, in number of data points, considered in the study (if any);
- *Energy Savings:* The reported percentage energy savings achieved by solutions reported in the study (if any is documented);
- *Industry involvement:* Industry involvement in the authorship of the study, which could be either academic-only authorship, industrial-only authorship, or mixed authorship;
- *Intended reader:* If the study is primarily intended for academic readers, industrial readers, or the general public.
- *Tool availability:* The availability of the tool(s) to address Green AI presented in the study (if any).

2.2.5 Data Synthesis. During the data extraction process, the data was harmonized by relying on the constant comparison [2] of extracted keywords, breaking up keywords into more specific ones when their semantic depth required it, or merging very similar keywords to avoid redundancy. This analysis process relied on open coding [4] to systematically identify recurrent concepts, followed by axial coding [4] to manage the increasing complexity of some emerging concepts.

The only exceptions were made for the *research strategy*, *industry involvement*, and *tool availability* fields of the extraction framework (see Section 2.2.4), for which provisional coding was used [4]. Specifically, coding of the *research strategy* relied on the research strategy categories reported by Stol *et al.* [16] was used. The industry involvement instead relied on three pre-defined fields, namely "academic-only authorship", "industrial-only authorship", or "mixed authorship". Finally, *tool availability* could only assume one of two pre-defined values, namely "Yes" (if the tool is available) or "No" (if the tool is not available, or none is presented in the primary study).

During the data extraction and synthesis phase, emerging codes are continuously discussed among the three authors of the review. This process ensures that the emerging codes and their abstraction level are kept consistent among researchers, and are aligned with the research goal and question of the study.

3 Results

In this section, we present the results collected with our SLR on Green AI.

3.1 Publication Years

The literature spans from 2015 with the first publication on the topic to this present year (*i.e.*, 2022). Figure 2 presents the distribution of the literature papers regarding the publication year. We observe a global increase following the years. Furthermore, a spike in the number of publications is seen in 2020, going from 7 publications in 2019 to 20 in 2020. As the automated initial search was launched in 2022, the publication trends reported in this review might not be representative of the actual research output of 2022 (see also Section 2.2.1).
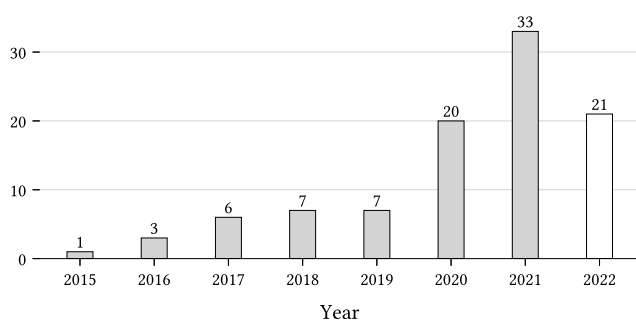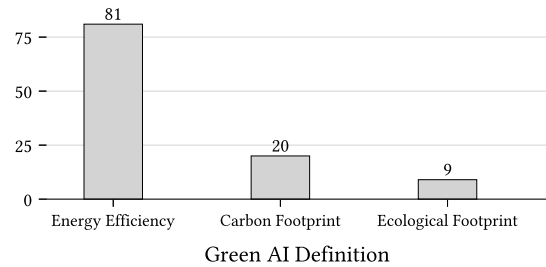
Figure 2: Number of publications per year.

Figure 3: Number of publications per type of Green AI definition.

3.2 Venue Types

Publications are particularly concentrated on conferences (▷ *47 out of 98 papers.*) and journals (▷ *39 out of 98 papers.*). Only 12 out of the 98 publications are associated with a workshop. Conferences being treated as an equal publishing venue as journals follows the trends observed in the computer science research field [5, 19].

> **Green AI publication trends**
>
> ✏ The topic of Green AI is experiencing an increasing trend of popularity, with a considerable growth in publications from 2020 onward. Most studies are published in conferences and journals, while only a minor portion in workshops.

3.3 Green AI Definition

The distribution of publications across different Green AI definitions is presented in Figure 3. Most literature addresses Green AI at the level of energy efficiency (81 papers). Higher-level definitions, namely carbon and ecological footprint, are only addressed in 20 and 9 publications respectively. Note that a primary study might be mapped to more than one definition, if more than one is used in the paper at hand.

3.4 Study Types

Existing literature on Green AI spans across three types of studies, namely *observational*, *solution*, and *position* papers (see also Section 2.2.4). As shown in Figure 4, from the 98 papers covered in this review, the most common are solution papers, with 51 entries, followed by observational with 35, and position papers with 12. Note that study types are mutually exclusive, *i.e.,* a single paper has only one study type.

3.5 Green AI Topics

From our analysis we identify 13 main topics being addressed by the Green AI literature. Figure 5 depicts the distribution of publications across the different topics. The most popular topic is *Monitoring*, addressed by 28 papers, followed by *Hyperparameter Tuning* (18), *Model Benchmarking* (17), *Deployment* (17), and *Model Comparison* (17). Since papers are not exclusive to a single topic, these top-4
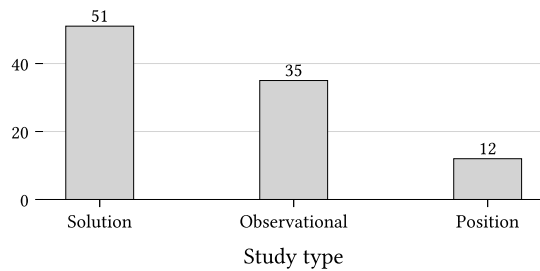
Figure 4: Number of publications per study type.

topics alone cover 61% of the papers in this review. Below, we pinpoint each topic with a short summary and the respective number of publications.

Monitoring ▷ ***28*** *out of 98 papers.* Covering monitoring approaches to study the energy and/or carbon footprint of AI models.

In this topic, papers report and reflect on the energy footprint of state-of-the-art models throughout their lifecycle. For example, Wu *et al.* [110] provide a landscape of the carbon footprint of AI models across Facebook. Findings showcase that, typically, throughout the lifetime of AI models, 50% of their carbon cost lies in the embodied carbon footprint of the hardware used to develop these models. However, the paper shows that the vast majority of training workflows under-utilizes GPUs at 30–50% of their full capacity.

Other papers within this topic focus on solutions to make carbon monitoring feasible in any AI project [48]. As an example, the Carbontracker offers a toolset to track and predict the energy and carbon footprint of training DL models [26]. These studies argue that it is quintessential to report the energy and carbon footprint of model development and training alongside performance metrics.

Hyperparameter Tuning ▷ ***18*** *out of 98 papers.* Improving or assessing the impact on the energy consumption of optimizing hyperparameters when training an AI model.

Many publications are motivated by the fact that tuning parameters leads to significant energy costs – it requires retraining a model multiple times in order to find the optimal set of hyperparameter values. Hence, most publications within this topic focus on identifying alternative strategies that reduce the number of iterations required to tune hyperparameters [95].

On a different perspective, Chavannes *et al.* [87] explore how hyperparameter tuning can help deliver more energy-efficient models by adding power consumption to the set of parameters being optimized.

Model Benchmarking. ▷ ***17*** *out of 98 papers.* Studies that contribute with benchmarks to compare the energy footprint of different models or training techniques.

Benchmarks help the community understand how the state of the art behaves w.r.t. given performance indicators. Ultimately, they help create baselines so that new approaches can be properly validated and compared to the state of the art. As example of publications within this category, Asperti et al. [27] evaluate the energy cost of different variational autoencoders. Another study, by Yu *et al.* [117], compares the energy efficiency of common machine learning algorithms when applied to clinical laboratorial datasets.

Deployment ▷ ***17*** *out of 98 papers.* Addressing the deployment stage of the lifecycle of an AI model.

Typically, publications in this topic discuss the problem of deploying AI models in a real scenario or in a scenario with peculiar constraints that challenge a standard approach. For example, deployment publications showcase the challenges of deploying energy-efficient AI in FPGA [101], in Edge devices [51, 68], in mobile devices [62, 79, 104], and so on.

Precision/Energy Trade Off ▷ ***11*** *out of 98 papers.* There is a turning point where to increase a very small fraction of the model performance, it is required to endure an energy-intensive training loop. Within this topic, papers address the Pareto trade-off between having optimal accuracy and/or optimal energy efficiency.

Zhang *et al.* [118] study how removing neurons from neural networks affects both accuracy and energy consumption. Results indicate that a good portion of neurons are redundant and can be removed to reduce energy consumption without a significant impact on accuracy. At the same time, it shows that there is a turning point where removing neurons improves energy efficiency but significantly reduces accuracy. Hence, the two parameters always need to be analyzed together. Other works opt for optimizing energy while keeping accuracy loss within a negligible margin [106].

Algorithm Design ▷ ***10*** *out of 98 papers.* Design of new training algorithms that produce models that are significantly more energy-efficient than the state of the art.

Some works propose small changes to the algorithms that make a big difference in the final energy consumption. For example, Garcia-Martin *et al.* [46] approximate the splitting criteria by selecting branches that require less computational effort. Results showcase decision trees that are up to 31% more energy efficient and with minimal impact on accuracy. Other examples include Espnetv2 [81], a lightweight convolutional neural network designed with power-efficiency in mind.

Libraries ▷ ***8*** *out of 98 papers.* Our choice of libraries have an impact on the final carbon footprint of AI systems. Studies within this topic provide some sort of evaluation of different AI libraries and how they contribute to energy efficiency.

This category shows that software engineering studies play an important role in enabling Green AI. Georgiou *et al.* [50] compare the energy footprint of deep learning frameworks. Results showcase that PyTorch is more energy-efficient than Tensorflow at the training stage. However, Tensorflow tends to be more energy-efficient at the inference stage. The study delves into the framework's different API methods and highlights code in the frameworks that should be optimized to reduce energy consumption. Finally, the authors motivate the importance of reporting and discussing energy efficiency in the documentation of deep learning frameworks.

Data Centric ▷ ***6*** *out of 98 papers.* Typically, the AI community has looked into coming up with better model training strategies. However, there is a new trend in AI that is raising the importance of developing better data collection and processing techniques as a more effective way to deliver better AI models. This line of thought within Green AI aims at reducing the carbon footprint of AI by tackling the problem at the data level.

Data-centric approaches for Green AI show that feature selection and subsampling techniques can significantly reduce the energy consumption of training machine learning models [102]. Subsampling strategies can be more sophisticated by removing data points that are expected to be redundant in terms of knowledge acquisition [39].

Network Architecture ▷ *6 out of 98 papers.* The impact of a distributed network on the energy efficiency of AI. AI models are often deployed in a distributed context – *e.g.,* IoT, edge computing, etc. Hence the design and architecture of the network plays an important role in leveraging sustainable models.

For example, Kim and Wu [68] propose an adaptive execution engine that selects the inference strategy according to the signal strength of the network in different devices, as it is known to affect the energy efficiency of the edge mobile system.

Estimation ▷ *5 out of 98 papers.* Collecting and making sense of energy or climate data is far from trivial – many different factors contribute to the final estimation [48]. This topic revolves around understanding ways of estimating the energy consumption or carbon footprint of models.

Existing solutions to estimate energy consumption for software fail to provide meaningful insight about energy consumption that can be mapped to a machine learning model's structure. IrEne creates a graph that breaks down NLP models into low-level machine learning primitives and provides energy estimations at the primitive level [37].

Emissions ▷ *4 out of 98 papers.* Papers that focus on understanding the carbon impact of creating and/or consuming AI systems. Dhar [40] flags the importance of being able to quantify carbon impact and the lack of tools and data available. Fraga-Lamas *et al.* [44] go beyond reporting the energy consumption of an AI-enabled IoT scenario and present how much carbon would be emitted in different countries and different energy sources.

Policy ▷ *3 out of 98 papers.* Studies within this topic address and discuss strategies on how we should handle the carbon footprint of AI as a society.

Perucica and Andjelkovic [85] reflect on the environmental policies implemented by the European Union, discussing whether they fit the AI era or new regulations are needed. Rhode *et al.* [89] call out for the unclear dilemma between the impact of existing/upcoming AI technologies and the commitment to achieve the 1.5℃ climate change goal as expressed in the UNFCCC Paris Declaration.

Ethics ▷ *3 out of 98 papers.* Papers that focus on the ethical implications of the growing carbon footprint of AI. Tamburrini [100] discusses the responsibilities of AI scientists, AI infrastructure providers, and other stakeholders in enabling Green AI. The paper questions whether it is ethically justified to create massive AI pipelines to improve accuracy.

Other ▷ *5 out of 98 papers.* Studies addressing a relevant topic with only a single publication in total: User values [69], Scheduling [120], Rebound Effects [109], Security [93], Energy Capping [70].

> **Green AI topics by study type**
>
> 🍃 There are 13 main topics on Green AI. The majority (61%) of the publications focuses on Monitoring, Hyperparameter-tuning, Model Benchmarking, and Deployment. Despite being important, topics such as Data-Centric, Estimation, and Emissions are underrepresented in the scientific literature.

3.6 Green AI Topics by Study Type

We further investigate the distribution of papers across different topics per category. Figure 6 presents a bubble plot that draws a bubble for each pair topic (x-axis) and study type (y-axis). The size of the bubble is proportional to the number of papers published in each pair. The plot enables a few observations.

Most topics adhere to the general pattern observed earlier in Section 3.4: the majority of papers consist of solution studies, followed by observational and then position. However, the topics of *Model Benchmarking* and *Libraries* do not follow this pattern, being mostly covered by observational papers. This is expected as these topics revolve around comparing different libraries and models to provide insight on the energy efficiency of different design decisions.

Moreover, papers from the least represented topics *Ethics*, *Policy*, and *Emissions* tend to be position papers. From the ten studies in these three topics, only one is observational and none is solution.

Also worth noticing is the fact that the majority of the position studies in Green AI only cover the smallest topics. Considering the top-10 topics – from Monitoring to Estimation – only 6 are position papers. In contrast, the bottom-4 topics (including Other) are covered by 10 position papers.

> **Green AI topics by study type**
>
> 🍃 Most publications on Ethics, Policy, and Emissions are position studies calling for more research in these topics.

3.7 Domains

Figure 7 presents the distribution of the publications according to the domain they cover. The majority of the publications (*i.e.,* ▷ *58 out of 98 papers.*) do not devote their studies to a specific domain, but tackle the energy efficiency of AI in a general context. Regarding the most specific studies, the most covered domains are:

Edge Regarding Internet of Things and Edge Computing, which are usually associated with distributed systems and networks. ▷ *24 out of 98 papers.*

Computer Vision Regarding image recognition. ▷ *6 out of 98 papers.*

Cloud ▷ *5 out of 98 papers.*

Mobile ▷ *4 out of 98 papers.*

The **Other** category gathers publications about a specific domain, being covered only once, among Health, Autonomous Driving, Smart cities, Human Activity, Wearables, and Embedded Systems.

> **Green AI domains**
>
> 🍃 The majority of Green AI studies does not focus a specific domain. Among specific domains, edge computing results to be the most recurrent one.
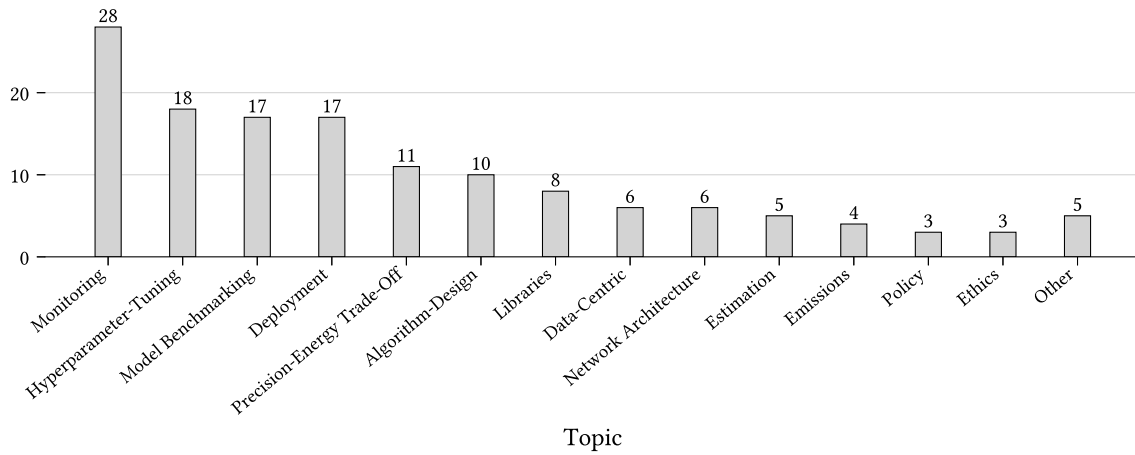
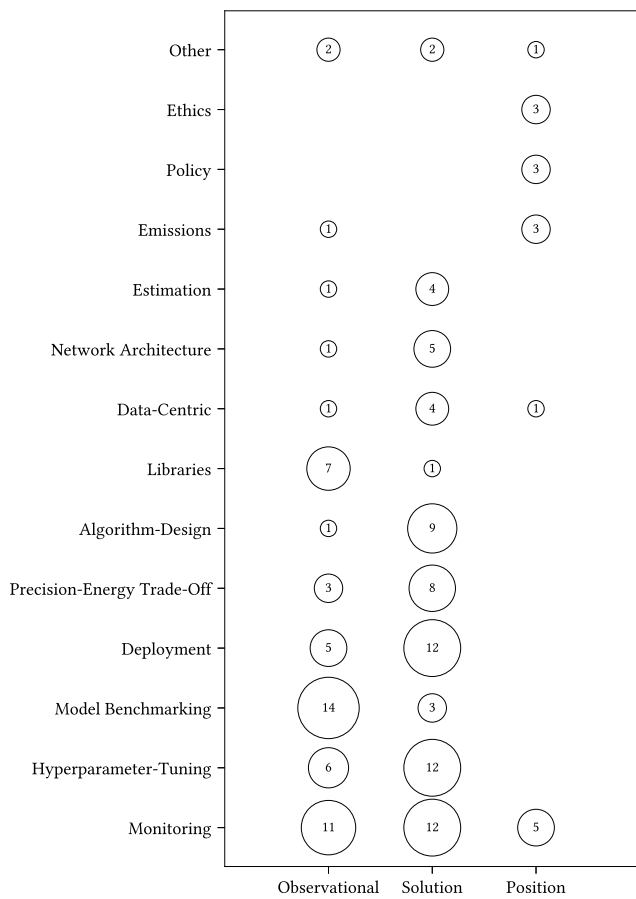**Figure 5: Number of papers per Green AI topic.**

Figure 6: Number of publications by topic and study type.

Figure 7: Number of publications per study domain.

is used to make predictions from new data. Thus, we classify the papers according to 3 categories: **training**, **inference**, and **all**. The **all** category translates the fact that the paper does not consider a particular phase, but the whole pipeline.

As depicted in Figure 8, we find that most of the publications on the topics of Green AI focus on the training phase (▷ *49 out of 98 papers.*). In comparison, fewer papers direct their studies at the inference phase (▷ *17 out of 98 papers.*) or on the overall process (▷ *32 out of 98 papers.*).

Figure 8: Number of publications per studied phase of AI.

3.8 AI Pipeline Phases

The AI pipeline is divided into two major phases: the training, when the AI model is built, and the inference, when the model
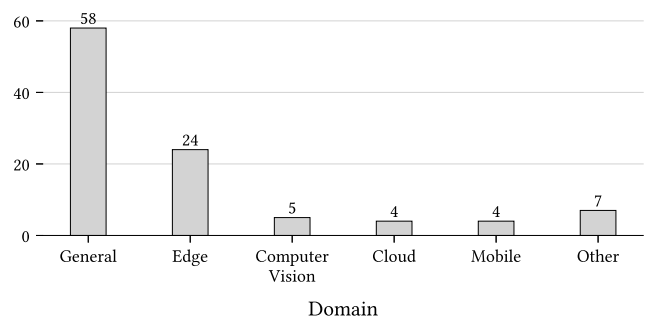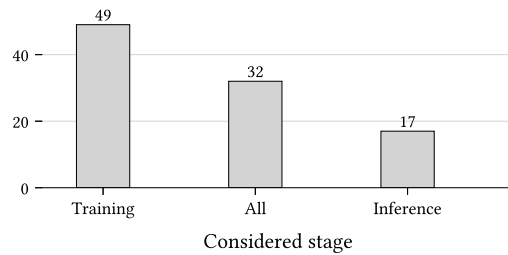
Figure 10: Occurrence of data types used in the Green AI literature.

3.9 Considered Artifacts

AI systems are based on several artifacts, and tackling the energy efficiency of such systems can thus involve multiple of those artifacts (*e.g.,* data, model, pipeline) or different related artifacts (*e.g.,* architecture, framework, CPU). A distribution of the artifacts considered in the primary studies is documented in Figure 9. The categories of artifacts are:

Model The publications within this category focus on the model and/or associated algorithm to tackle the energy efficiency of AI. ▷ *63 out of 98 papers.*

Data Papers that address energy efficiency through the study of the data used in the AI pipeline. ▷ *8 out of 98 papers.*

Pipeline Studies looking at the whole AI pipeline. ▷ *3 out of 98 papers.*

Other Publications dealing with CPU, architecture, and framework. ▷ *4 out of 98 papers.*

General The papers do not specify a particular artifact and address AI systems as a whole. ▷ *24 out of 98 papers.*
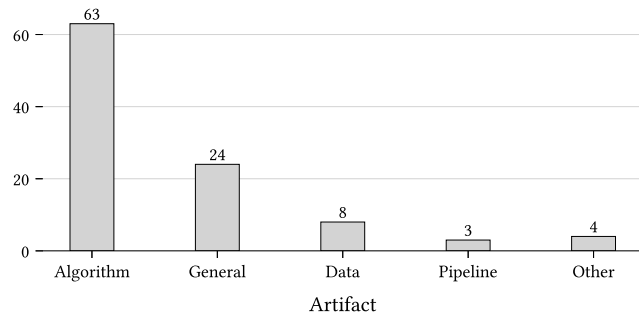

Figure 9: Number of publications per studied artifact.

3.10 Algorithm Types

By considering the primary studies which focus on a specific algorithm (▷ *51 out of 98 papers.*), we note that the vast majority focus on **neural networks** (▷ *41 out of 98 papers.*). Only a much smaller fraction focuses on algorithms of different nature, such as decision trees (▷ *5 out of 98 papers.*), genetic algorithms (▷ *1 out of 98 papers.*), or logistic regression models (▷ *5 out of 98 papers.*).

Regarding the deep neural network algorithms, we also note a further characterization of this field, with 8 studies focusing on **convolutional neural networks**, one on **transformers**, and one on **spiking neural networks**. We also observe three algorithms that appear only once in the Green AI literature (**Other** category, ▷ *3 out of 98 papers.*), namely *genetic algorithms*, *logic regression algorithms*, and *stochastic gradient descent algorithms*.
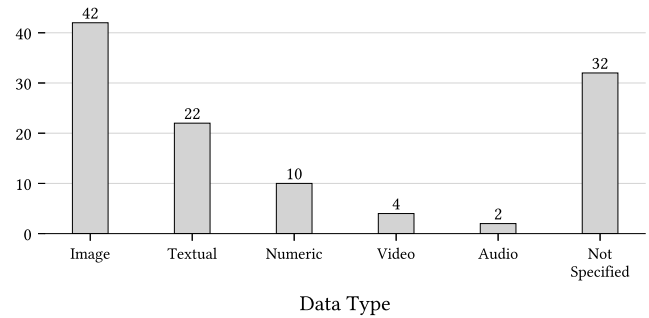
3.11 Data Types Used

Regarding the types of data used in the Green AI body of literature, an overview of their distribution is reported in Figure 10. From the figure, we can observe that the recurrence of data types across primary studies is:

Image data ▷ *42 out of 98 papers.*
Textual data ▷ *22 out of 98 papers.*
Numeric data ▷ *10 out of 98 papers.*
Video data ▷ *4 out of 98 papers.*
Audio data ▷ *2 out of 98 papers.*

From the distribution of data types, we notice that image data is by far the most used one, and is utilized by almost half of the studies in the body of literature. The second most utilized data type is textual data, which nevertheless appears approximately half as often as the image one. Other types of data result to be less recurrent, with only few studies utilizing audio data (*e.g.,* Lenherr et al. present a metric to measure the sustainability of Green AI by considering as case study the Intel MovidiusX processor, an embedded video processor with a Neural Engine for video processing and object detection [74]).

A rather high number of primary studies does not specify any kind of data (**Not specified** category, ▷ *32 out of 98 papers.*). This finding has to be primarily attributed to the position and theoretical papers included in the review (see also Section 3.4 and Section 3.5).

3.12 Dataset sizes

Regarding the size of the datasets used in the papers, approximately half of the primary studies (▷ *48 out of 98 papers.*)) directly reference the number of data points used. By inspecting such numbers, we note that the number of data points used to study and to evaluate Green AI algorithms and approaches varies greatly, and ranges from
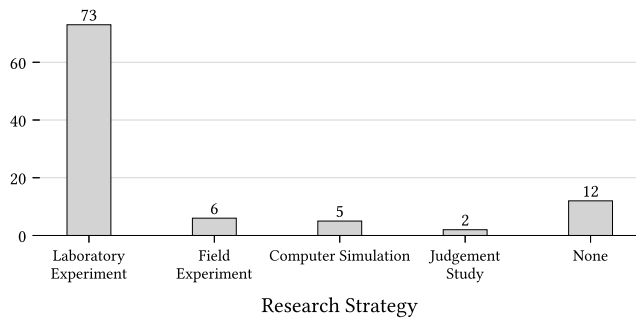
Figure 11: Occurrence of research strategies used in the Green AI literature.

1k data points [51] to **40M data points** [45]. Almost half of the studies reporting the number of data points (▷ *25 out of 48 papers*) utilize data points in the order of thousands ($1k \leq \#datapoints \leq 70k$), while the remaining (▷ *23 out of 48 papers*) use one million data points or more ($1M \leq \#datapoints \leq 40M$).

> **Green AI dataset sizes**
>
> 🌿 Dataset sizes range from 1k to 40M data points, with approximately half of the studies utilizing 1M or more data points.

3.13 Research Strategies

By considering the research strategies [16] utilized in the Green AI literature, the distribution of the various strategies, according to the collected primary studies, is reported in Figure 11.

The majority of paper results adopt **laboratory experiments** (▷ *73 out of 98 papers.*), while only a fraction uses other research strategies, such as **field experiments** (▷ *6 out of 98 papers.*), *i.e.,* experiments conducted in pre-existing settings and **computer simulations**, *i.e.,* "in silico" simulations conducted in a nonempirical setting (▷ *5 out of 98 papers.*). As examples, Liu *et al.* [77] use a field study to assess a green software stack for computer vision of autonomous robots, while Yosuf *et al.* [116] leverage computer simulations to study how virtualized cloud fog networks can be used to improve AI energy efficiency. The 12 papers not displaying any research strategy correspond to the position papers (cf. the "None" category in Figure 11).

> **Green AI Research Strategies**
>
> 🌿 Most Green AI studies use laboratory experiments, while only a minority adopt other research strategies, such as field experiments and computer simulations.

3.14 Energy savings

By considering the energy savings reported achievable *via* Green AI strategies, we note that only approximately a third of the primary studies explicitly document them (▷ *27 out of 98 papers.*). Out of all Green AI strategies, among the ones which report concrete saving percentages, a technique based on structure simplification for deep neural networks results to save more energy, amounting to **115% energy savings** [118]. The other techniques which result
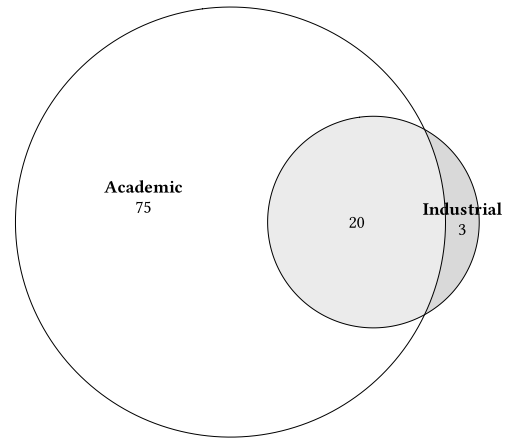

Figure 12: Industry involvement.

to optimize energy the most are based on quantizing the inputs of decision trees [23] (97% energy savings), using data-centric Green AI techniques [102] (92% energy savings), and leveraging efficient deployment of AI algorithms *via* virtualized cloud fog networks (91% energy savings) [116]. Overall, more than half of the papers explicitly reporting energy saving percentages report a saving of at least 50% (▷ *17 out of 27 papers*), while only a minor number savings between 13% and 49%.

> 🌿 **Green AI energy savings**
>
> Studies report energy savings between 13% and 115% energy savings, with more than half of the papers reporting savings of at least 50%.

3.15 Industry involvement

Regarding the industry involvement in Green AI scientific publications (see also Section 2.2.4), an overview of the authorship of the Green AI primary papers is depicted in Figure 12.

From the figure, we can note that most Green AI studies are authored exclusively by academic researchers (▷ *75 out of 98 papers.*), while also a considerable portion, amounting almost to a fourth of all primary studies, are authored by a mix of academic and industrial researchers (▷ *20 out of 98 papers.*). Green AI studies written exclusively by industrial authors appear only in rare instances (▷ *3 out of 98 papers.*).

> 🌿 **Industry involvement**
>
> Most studies are written by academic authors, while a minor portion by a mix of academic and industrial authors. Green AI studies written exclusively by academic authors are very rare.

3.16 Intended readers

By considering the intended readers of the Green AI scientific literature, the vast majority targets academic readers (▷ *85 out of 98 papers.*), while a much smaller portion both academic and industrial readers (▷ *8 out of 98 papers.*). Despite scientific papers targeting intuitively a specialized audience, among the Green AI literature, few studies are intended also for the general public (▷ *5*

out of 98 papers.). For example, Dhar *et al.* [40], present an intuitive yet thoroughly positioned article on the systemic effect of AI on carbon emissions. Interestingly, among the primary studies, few are intended also for policymakers, *i.e.,* aim to sensibilize government stakeholders to consider issues related to Green AI. For example, in a paper by Rohde *et al.* [89], how opportunities and risks for the environment, economy and society associated with AI can be governed are discussed.

> 🍃 **Intended readers**
>
> The vast majority of Green AI studies are targeting academic readers, while a much smaller portion targets both academic and industrial readers. A handful of studies, especially position papers, are intended for the general public.

3.17   Tool Provision

Among the primary studies collected for this literature review on Green AI, only a small fraction (▷ **15 out of 98 papers.**) makes tools available to tackle Green AI. The tools provided are of heterogeneous nature, and range from tools to monitor the resource efficiency of AI algorithms [52], to tools optimizing the energy efficiency for stochastic edge inference [68], and implementations of convolutional neural networks optimized for energy efficiency [81].

> 🍃 **Green AI Tool Provision**
>
> Albeit numerous studies provide solution to tackle Green AI, only a fraction of them makes tools based on the solutions readily available online as an implemented tool.

4   Discussion

The consolidated and still growing Green AI publication trend. From the analysis of the publication trends a clear picture emerges. The topic is gaining increasing traction in the academic community, especially if the latest years are considered (from 2020 onward). Despite being a quite new research topic (with the first paper on Green AI being published in 2015), the socio-environmental relevance of the topic seems to be reflected in its targeted publication venues. With conferences and journal being the most recurrent Green AI publication venues, the Green AI research field seems to have positioned and consolidated itself quite quickly within AI research communities.

A definition of Green AI. From the results regarding how the term "Green AI" is used in the literature a clear picture emerges. Most Green AI studies consider Green AI as exclusively related to energy efficiency. Only fewer studies examine the influence of AI on greenhouse gas emissions (CO_2), and an even minor fraction examines the holistic impact that AI has on the natural environment.

By considering the different levels of abstraction (namely *energy efficiency*, *carbon footprint*, and *environmental footprint*) the higher, more encompassing level, of environmental footprint seems best fitted to define the field of Green AI. In fact, as demonstrated in recent literature, reducing the environmental impact of AI exclusively to energy consumption has to be deemed as overly simplistic process [8]. Similarly, as green resources are sustainable but not infinite [17], the field of Green AI has to account also for the multifaceted environmental impact AI can have, other than CO_2

emissions alone. Based on these considerations, we define the field of Green AI as follows:

> *"Green AI regards practices aimed at utilizing AI to mitigate the impact that humans have on the natural environment in terms of natural resources utilized, and/or mitigating the impact that AI itself can have on the natural environment."*

On one hand, the definition above perfectly fits the studies focusing on the holistic impact that Green AI has on the natural environment. On the other hand, given its encompassing nature, the definition is also suited for studies focusing on lower abstraction levels of sustainability, such as Green AI CO_2 emissions and energy consumption. In the latter case however, the definition also acts as a word of warning: while studying the lower levels of Green AI is paramount, only by considering the totality of the heterogeneous natural resources utilized by AI can we really understand the environmental impact of AI.

The transdisciplinary topics of Green AI (with gaps). The 13 different topics we discover in this review emphasize that Green AI is a broad field that needs to be tackled as a transdisciplinary field. Some topics are naturally tied to training strategies (*e.g.*, monitoring, hyperparameter tuning, algorithm design). However, there are other topics that take Green AI outside the training realm.

This is the case for example of Deployment, Libraries, and Estimation that promise to be relevant in enabling Green AI. We argue that other disciplines need to be involved. For example, Software Engineering which has been dealing with these topics for traditional software systems. As highlighted by Cao *et al.* in their work on estimation [37], one cannot expect existing strategies for traditional software to address the new challenges of AI-based systems. Conversely, only a few Green AI papers [50, 56, 80] come from software engineering venues.

Our analysis also shows that the topics Estimation and Emissions are under-represented, with six and five papers, respectively. We argue that more work is quintessential in these topics to help scientists and practitioners report the carbon footprint of their AI models in a seamless way.

We showcase that papers under the topic Policy are only covered by position papers. We find this finding disconcerting: new policies to encourage Green AI within both industry and academic contexts need to be backed up with reliable evidence. Hence, we need more observational and solution papers that tackle this topic in the near future.

The same issue is present in Emissions – only one paper is observational and the remaining are position. It might be the case that computing the climate impact of AI is far from trivial and it is easier said then done. Again, this is a call for the community to take action. It is not enough to ask big companies to provide their data on carbon impact – we also need to provide strategies and solutions to make it standard and straightforward.

The fundamental Green AI research unbounded from application domains. From the collected results we deduce that, in order to improve the environmental sustainability of AI, it is often not necessary to focus on a specific domain. This implies that frequently fundamental aspects of Green AI are still open to

investigation, and results can then be ported from a generic setting to specific domains. However, from the obtained results, we also note that the increasing distribution of digital infrastructures to achieve environmental sustainability [17] might have played a role in Green AI research, with edge computing being the most considered specific domain.

The high emphasis on the AI training phase. The results regarding the AI pipeline phases considered in the literature unequivocally point to training as the most studied phase. Albeit the training phase is intuitively the most energy-greedy phase, this results calls for a word of caution. From recent results (*e.g.,* a study on data-centric Green AI [102]) the inference phase results to consume only a negligible fraction of the energy consumed in the training phase. Nevertheless, given the high execution rate of the inference phase, how the energy consumed by the infrequent execution of the training phase compares to one of the highly executed inference phase is still an open question. As a call for action, studies should be conducted by considering the energy consumed throughout the whole life cycle of AI models, from their training to inference phase, till their eventual deprecation.

Image datasets as primary Green AI data source. By considering the data types used in Green AI studies, we note that the vast majority of the literature uses image data. To the best of our knowledge, this choice is not guided by any specific research design choice (*e.g.,* AI models based on image data being the most used in practice, or being the most energy greedy ones). For this reason, we conjecture that the popularity of utilizing image data for Green AI data is mostly driven by convenience, either because past work focused on such data by chance, image datasets are more accessible/standardized with respect to other ones, or more off the shelf image AI models/libraries are currently available. Regardless of the cause, this result points to the need of utilizing more heterogeneous data types, rather than focusing primarily on image data, in order to gain a holistic understanding of Green AI.

Laboratory experiments guided till now Green AI. The most common research strategy adopted for Green AI studies clearly emerges from the literature as being laboratory experiments. Given the fast popularization and consolidation of the Green AI research field, from this review it seems as if the time is suitable to shift the focus to other research strategies, *e.g.,* field experiments and case studies. This would not only allow to change the considered context from an *in vitro* to an *in vivo* setting, but also to bridge potential gaps between academic research and industrial practice.

The highly promising energy savings of Green AI. From the results of this review, we deduce that the research field of Green AI is highly promising, with more than half of the papers reporting 50% or more energy savings. This study focuses on the state of the art of Green AI, rather than focusing on the state of practice. It would be therefore interesting to understand, as future work, the extent to which this encouraging results are transposed to industrial practice, and the potential impediments which hinder their adoption or full potential.

A noticeable industry involvement. Regarding industry involvement in Green AI studies, the results gathered from this review are promising. The authorship of Green AI literature results to be to a good extent shared between academic and industrial researchers/practitioners. This finding might highlight the sensibility

of industry towards Green AI concerns, and/or the importance of moving towards more environmentally sustainable AI practices.

As a potential impediment to the industrial adoption of Green AI research, our results point to a low recurrence of studies targeted towards practitioners. While numerous journals are explicitly aimed at practitioners, *e.g.,* IEEE Software[2], only few studies on Green AI included in our review target them. This result might point to the fact that the Green AI interest is still primarily focused towards academic activities, while the authorship showcases a rather high interest of industry. As take away, similar to the considerations made for the Green AI research strategies, it might be the right moment to consider a higher involvement of industry in Green AI, which results to date to be a research area still targeted primarily towards academic readers.

Green AI lacks tool support. Finally, from this review, we note that the current situation regarding the provisioning of Green AI tools is not bright. Albeit the majority of the studies present Green AI solutions, only a small fraction of them makes the solutions available as a tool. We conjecture that this result might either point towards (i) a fast-paced nature of Green AI research, in which results are rapidly deprecated, and hence tools are not meaningful, or (ii) an immaturity of the research field, which still requires a solid empirical foundation on which tools can be built upon.

5 Threats to Validity

In this section, we discuss the threats to validity of our study. To ensure the quality of the results, we established a well-defined research protocol to proceed with the data collection. In addition, throughout our study, we followed the recommendations of the guidelines for conducting a systematic literature review [6, 7, 10, 14, 21]. We designed and carried the different reviewing processes according to the rigorous protocol we established after the guidelines and described in Section 2. Nevertheless, some threats to validity can still exist even with our best efforts. In the following, we present the threats which could have influenced our study, jointly with the strategies we adopted to mitigate them.

External validity The main threat to external validity is that the literature collected and analysed in this study is not sufficiently representative. To avoid this situation, we surveyed three prominent literature indexers through an automatic query (*i.e.,* Google Scholar, Scopus, and Web of Science), and left the year of publication unbounded, to reduce the probability of missing any relevant publication. In addition, the search query was designed to target relevant literature directly with specific keywords, while allow for flexibility by considering similar, complementary, and variation of the keywords (*e.g.,* the keywords *green*, *sustainability*, and *sustainable*). We also mitigated the threat of having an incomplete set of studies, as well as the threat associated with the specificity of the terms used in the search query, by performing a complementary iterative bidirectional snowballing process of the query results. This latter search strategy allowed us to include literature related to our query that was not directly referencing any of the automated search keywords. We limited our review of the literature to peer-reviewed studies, to moderate the threat about the low quality of the set of primary studies. We deem that such practice does not constitute

---

[2]https://www.computer.org/csdl/magazine/so. Accessed 22nd December 2022.

an additional threat, as peer-review is a standard requirement of high-quality publications.

Internal validity To address potential threats to internal validity, we established a rigorous research protocol *a priori*, and we followed it to conduct all the research activities. Subjective biases and interpretations were mitigated by closely complying with the selection criteria to evaluate the studies. Moreover, weekly meeting were held during the selection process to jointly discuss examples, doubts, and to align the selection process between the three researchers.

Construct validity To ensure that the set of studies answered our research questions, we applied *a priori* carefully constructed inclusion and exclusion criteria to strictly control the manual selection of studies. We then used the bidirectional snowballing technique to expand the range of relevant primary studies to a more comprehensive set.

Conclusion validity Possible sources of bias arising from the data extraction and analysis phases were mitigated by strict compliance with an *a priori* defined protocol, explicitly tailored to collect the data needed to answer our research questions. In all, we followed the best practises of the standard guidelines for systematic literature reviews [6, 7, 10, 14, 21]. Lastly, we documented all the data throughout the whole review process and made them available for reproducibility and replicability purposes (see Section 1).

6 Related Work

Despite the growing interest around Green AI, the topic has been marginally considered only in a handful of reviews. The related work manly investigates the topic as an intersection of AI and environmental sustainability, or by defining it as a specific subdomain of software engineering. To the best of our knowledge, this review is the first aiming towards a comprehensive review of Green AI research and its characteristics.

In a recent publication, Natarajan *et al.* perform a systematic literature review on the topics of 'AI for Environmental Sustainability' as well as 'Environmental Sustainability of AI'. The authors present the affordances of the use of AI for sustainability that they extracted from the literature [12]. 'AI affordances' are introduced as the possible actions offered by AI artifacts to an organizational actor whose goal is to achieve environmental sustainability. The authors point out the focus of previous research on the technical side, and they advocate for a further exploration of the concept of sustainable AI affordances from a socio-technical perspective. The literature is exclusively analyzed with respect to building the AI affordances, and other characteristics of the state-of-the art of Green AI are considered nor discussed in the study. In contrast, our review focuses on the sustainability of AI, and maps the entirety of the Green AI literature. In our review, we aim at providing a detailed and comprehensive overview of the characteristics of the Green AI state-of-the-art research (*e.g.,* topic, domain, type of study, targeted artifact, overview of energy savings, tool provision, industrial involvement). Therefore, in contrast to the work of Natarajan *et al.* [12], we consider the different facets of Green AI, rather than exclusively on AI affordances, leading to a more holistic review of Green AI, and a higher number of considered primary studies (98 versus 41 papers). This difference could be explained by the fact that their review only includes papers involving consumer products and services and excludes papers dealing with non-commercial applications, whereas we provide an overview of the whole field of Green AI.

Previous literature reviews consider Green AI research by focusing exclusively on specific subdomains of AI and application subdomains of Software Engineering, *e.g.,* deep learning [22], information retrieval [15], or embedded systems [11]. In contrast, our research aims to review the entirety of the Green AI literature, regardless of the specific AI or software engineering subdomain it focuses on.

In the survey of Xu et al. [22], the authors provide an overview of the approaches aimed at improving the environmental sustainability of deep learning. The authors map the different approaches using a taxonomy of the deep learning life cycle stage and its related artifacts. In contrast to such study, in this review we target a higher number of Green AI characteristics (see Section 2.2.4), and target the entirety of Green AI literature, rather than exclusively the one on deep learning.

Scells et al. [15] provide a literature review on methods related to the domain of Green Information Retrieval. The authors explain that the domain of Information Retrieval (IR) produces relatively low emissions compared to other research domains, but they also warn that similar trends of costs and environmental impact may appear considering the growing development of new IR-focused deep learning models. Natural Language Processing and Machine Learning are also discussed, but only with respect to the Information Retrieval domain. Therefore, they are not addressing the whole field of AI, as done in this review.

Finally, the optimizations that can be made for the implementation of deep learning models on the specific platform of NVIDIA Jetson are reviewed with a focus on energy efficiency by Mittal [11]. The review covers studies at both the hardware and software level. Nevertheless, the review addresses only the Jetson platform [3]. We differentiate ourselves from this study by providing a holistic review of Green AI, rather than focusing exclusively on deep learning.

7 Conclusion

In this systematic literature review, we aimed at characterizing the existing body of research in Green AI. We identified 98 peer-reviewed publications that show a significant growth in this research field since 2020.

We provide an encompassing overview and characterization of the different topics being addressed by Green AI papers. We identified 13 different Green AI topics, showcasing that the spotlight falls on monitoring, hyperparameter-tuning, model benchmarking, and deployment. Less frequent topics – such as data-centric, estimation, and emissions – show less obvious approaches that deserve further research in the upcoming years.

The potential of Green AI cannot be disregarded: the majority of publications show significant energy savings, up to 115%, at little or no cost in accuracy. However, we argue that most publications revolve around laboratory studies. More field experiments are quintessential to help AI practitioners embrace green strategies that

---

[3]https://developer.nvidia.com/embedded-computing. Accessed 23th December 2022.

are effective, feasible, and mensurable. This is also reflected in the small participation of the industry in these studies – only 23% of publications involve industry partners.

At the same time, we conclude that the field seems to be reaching a considerable level of maturity. Hence, it is necessary to encourage the port of promising academic results to industrial practice. In other words, our study calls out for the importance of having reproducible research. Only a small fraction of solution papers offers a tool or software package that can be used by the community. We argue that Green AI is an urgent and necessary line of research that needs to grow fast and solid – non-replicable research can only slow us down.

This review also serves as a foundation for future research that ultimately aims to reduce the climate impact of AI. In this respect, we see potential in follow-up grey literature or interview studies to understand how AI professionals are currently addressing the issue.

References

[1] Victor R. Basili, Gianluigi Caldiera, and Dieter Rombach. 1994. The Goal Question Metric Approach. In *Encyclopedia of Software Engineering*. Wiley, 528–532.

[2] Barney G Glaser. 1965. The constant comparative method of qualitative analysis. *Social problems* 12, 4 (1965), 436–445.

[3] Mark Haakman, Luís Cruz, Hennie Huijgens, and Arie van Deursen. 2021. AI lifecycle models need to be revised. *Empirical Software Engineering* 26, 5 (2021), 1–29.

[4] Bryan Jenner, Uwe Flick, Ernst von Kardoff, and Ines Steinke. 2004. *A companion to qualitative research*. Sage, 271–275.

[5] Jinseok Kim. 2019. Author-based analysis of conference versus journal publication in computer science. *J. Assoc. Inf. Sci. Technol.* 70, 1 (Jan. 2019), 71–82. https://doi.org/10.1002/asi.24079

[6] Barbara Kitchenham. 2004. Procedures for performing systematic reviews. *Keele, UK, Keele University* 33, TR/SE-0401 (2004), 28.

[7] Barbara Kitchenham and Pearl Brereton. 2013. A systematic review of systematic review process research in software engineering. *Information and software technology* 55, 12 (2013), 2049–2075.

[8] Alexandra Sasha Luccioni and Alex Hernandez-Garcia. 2023. Counting Carbon: A Survey of Factors Influencing the Emissions of Machine Learning. *arXiv preprint arXiv:2302.08476* (2023).

[9] Jan Matuštík and Vladimír Kočí. 2021. What is a footprint? A conceptual analysis of environmental footprint indicators. *Journal of Cleaner Production* 285 (2021), 124833.

[10] Philipp Mayring et al. 2004. Qualitative content analysis. *A companion to qualitative research* 1, 2 (2004), 159–176.

[11] Sparsh Mittal. 2019. A Survey on optimized implementation of deep learning models on the NVIDIA Jetson platform. *J. Syst. Archit.* 97 (Aug. 2019), 428–442. https://doi.org/10.1016/j.sysarc.2019.01.011

[12] Harish Karthi Natarajan, Danielly de Paula, Christian Dremel, and Falk Uebernickel. 2022. A Theoretical Review on AI Affordances for Sustainability. In *Americas Conference on Information Systems*.

[13] Kai Petersen, Robert Feldt, Shahid Mujtaba, and Michael Mattsson. 2008. Systematic mapping studies in software engineering. *International Conference on Evaluation and Assessment in Software Engineering*, 68–77.

[14] Kai Petersen, Sairam Vakkalanka, and Ludwik Kuzniarz. 2015. Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and software technology* 64 (2015), 1–18.

[15] Harrisen Scells, Shengyao Zhuang, and Guido Zuccon. 2022. Reduce, Reuse, Recycle: Green Information Retrieval Research. In *SIGIR '22: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Association for Computing Machinery, New York, NY, USA, 2825–2837. https://doi.org/10.1145/3477495.3531766

[16] Klaas-Jan Stol and Brian Fitzgerald. 2018. The ABC of software engineering research. *ACM Transactions on Software Engineering and Methodology (TOSEM)* 27, 3 (2018), 1–51.

[17] Roberto Verdecchia, Patricia Lago, and Carol de Vries. 2022. The future of sustainable digital infrastructures: A landscape of solutions, adoption factors, impediments, open problems, and scenarios. *Sustainable Computing: Informatics and Systems* (2022), 100767.

[18] Roberto Verdecchia, Patricia Lago, Christof Ebert, and Carol De Vries. 2021. Green IT and Green Software. *IEEE Software* 38, 6 (2021), 7–15.

[19] George Vrettas and Mark Sanderson. 2015. Conferences versus journals in computer science. *J. Assoc. Inf. Sci. Technol.* 66, 12 (Dec. 2015), 2674–2684. https://doi.org/10.1002/asi.23349

[20] Thomas Wiedmann and Jan Minx. 2008. A definition of 'carbon footprint'. *Ecological economics research trends* 1, 2008 (2008), 1–11.

[21] Claes Wohlin. 2014. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *International Conference on Evaluation and Assessment in Software Engineering*. ACM Press, 1–10.

[22] Jingjing Xu, Wangchunshu Zhou, Zhiyi Fu, Hao Zhou, and Lei Li. 2021. A Survey on Green Deep Learning. *arXiv* (Nov. 2021). https://doi.org/10.48550/arXiv.2111.05193 arXiv:2111.05193

Primary Studies

[23] Brunno Abreu, Mateus Grellert, and Sergio Bampi. 2020. VLSI design of tree-based inference for low-power learning applications. In *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 1–5.

[24] Brunno Abreu, Mateus Grellert, and Sergio Bampi. 2022. A framework for designing power-efficient inference accelerators in tree-based learning applications. *Engineering Applications of Artificial Intelligence* 109 (2022), 104638.

[25] Phyllis Ang, Bhuwan Dhingra, and Lisa Wu Wills. 2022. Characterizing the Efficiency vs. Accuracy Trade-off for Long-Context NLP Models. In *Proceedings of NLP Power! The First Workshop on Efficient Benchmarking in NLP*. Association for Computational Linguistics, Dublin, Ireland, 113–121. https://doi.org/10.18653/v1/2022.nlppower-1.12

[26] Lasse F. Wolff Anthony, Benjamin Kanding, and Raghavendra Selvan. 2020. Carbontracker: Tracking and Predicting the Carbon Footprint of Training Deep Learning Models. ICML Workshop on Challenges in Deploying and monitoring Machine Learning Systems. arXiv:2007.03051.

[27] Andrea Asperti, Davide Evangelista, and Elena Loli Piccolomini. 2021. A Survey on Variational Autoencoders from a Green AI Perspective. *SN Comput. Sci.* 2, 4 (July 2021), 1–23. https://doi.org/10.1007/s42979-021-00702-9

[28] Nesrine Bannour, Sahar Ghannay, Aurélie Névéol, and Anne-Laure Ligozat. 2021. Evaluating the carbon footprint of NLP methods: a survey and analysis of existing tools. In *EMNLP, Workshop SustaiNLP*.

[29] Michel Barlaud and Frédéric Guyard. 2021. Learning sparse deep neural networks using efficient structured projections on convex constraints for green AI. In *2020 25th International Conference on Pattern Recognition (ICPR)*. 1566–1573. https://doi.org/10.1109/ICPR48806.2021.9412162

[30] Soroush Bateni, Husheng Zhou, Yuankun Zhu, and Cong Liu. 2018. Predjoule: A timing-predictable energy optimization framework for deep neural networks. In *2018 IEEE Real-Time Systems Symposium (RTSS)*. IEEE, 107–118.

[31] 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 *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery, New York, NY, USA, 610–623. https://doi.org/10.1145/3442188.3445922

[32] Alexander EI Brownlee, Jason Adair, Saemundur O Haraldsson, and John Jabbo. 2021. Exploring the accuracy–energy trade-off in machine learning. In *2021 IEEE/ACM International Workshop on Genetic Improvement (GI)*. IEEE, 11–18.

[33] Sevda Ozge Bursa, Ozlem Durmaz Incel, and Gulfem Isiklar Alptekin. 2022. Transforming Deep Learning Models for Resource-Efficient Activity Recognition on Mobile Devices. In *2022 5th Conference on Cloud and Internet of Things (CIoT)*. IEEE, 83–89.

[34] Ermao Cai, Da-Cheng Juan, Dimitrios Stamoulis, and Diana Marculescu. 2017. *NeuralPower*: Predict and Deploy Energy-Efficient Convolutional Neural Networks. In *Proceedings of the Ninth Asian Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 77)*, Min-Ling Zhang and Yung-Kyun Noh (Eds.). PMLR, Yonsei University, Seoul, Republic of Korea, 622–637. https://proceedings.mlr.press/v77/cai17a.html

[35] Antonio Candelieri, Riccardo Perego, and Francesco Archetti. 2021. Green machine learning via augmented Gaussian processes and multi-information source optimization. *Soft Comput.* 25, 19 (Oct. 2021), 12591–12603. https://doi.org/10.1007/s00500-021-05684-7

[36] Qingqing Cao, Aruna Balasubramanian, and Niranjan Balasubramanian. 2020. Towards Accurate and Reliable Energy Measurement of NLP Models. In *Proceedings of SustaiNLP: Workshop on Simple and Efficient Natural Language Processing*. Association for Computational Linguistics, Online, 141–148. https://doi.org/10.18653/v1/2020.sustainlp-1.19

[37] Qingqing Cao, Yash Kumar Lal, Harsh Trivedi, Aruna Balasubramanian, and Niranjan Balasubramanian. 2021. IrEne: Interpretable Energy Prediction for Transformers. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (2021). https://doi.org/10.18653/v1/2021.acl-long.167

[38] Francisco M Castro, Nicolás Guil, Manuel J Marín-Jiménez, Jesús Pérez-Serrano, and Manuel Ujaldón. 2019. Energy-based tuning of convolutional neural networks on multi-GPUs. *Concurrency and Computation: Practice and Experience* 31, 21 (2019), e4786.

[39] Priyadarshan Dhabe, Param Mirani, Rahul Chugwani, and Sadanand Gandewar. 2021. Data Set Reduction to Improve Computing Efficiency and Energy Consumption in Healthcare Domain. In *Digital Literacy and Socio-Cultural Acceptance of ICT in Developing Countries*. Springer, 53–64.

[40] Payal Dhar. 2020. The carbon impact of artificial intelligence. *Nat. Mach. Intell.* 2, 8 (2020), 423–425.

[41] Josefa Díaz-Álvarez, Pedro A Castillo, Francisco Fernández de Vega, Francisco Chávez, and Jorge Alvarado. 2022. Population size influence on the energy consumption of genetic programming. *Measurement and Control* (2022), 00202940211064471.

[42] Jesse Dodge, Taylor Prewitt, Remi Tachet des Combes, Erika Odmark, Roy Schwartz, Emma Strubell, Alexandra Sasha Luccioni, Noah A. Smith, Nicole DeCario, and Will Buchanan. 2022. Measuring the Carbon Intensity of AI in Cloud Instances. In *FAccT '22: 2022 ACM Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery, New York, NY, USA, 1877–1894. https://doi.org/10.1145/3531146.3533234

[43] Mariza Ferro, Gabrieli D. Silva, Felipe B. de Paula, Vitor Vieira, and Bruno Schulze. 2021. Towards a sustainable artificial intelligence: A case study of energy efficiency in decision tree algorithms. *Concurrency Computat. Pract. Exper.* n/a, n/a (Dec. 2021), e6815. https://doi.org/10.1002/cpe.6815

[44] Paula Fraga-Lamas, Sérgio Ivan Lopes, and Tiago M. Fernández-Caramés. 2021. Green IoT and Edge AI as Key Technological Enablers for a Sustainable Digital Transition towards a Smart Circular Economy: An Industry 5.0 Use Case. *Sensors* 21, 17 (Aug. 2021), 5745. https://doi.org/10.3390/s21175745

[45] Eva Garcia-Martin, Niklas Lavesson, and Håkan Grahn. 2017. Identification of Energy Hotspots: A Case Study of the Very Fast Decision Tree. In *Green, Pervasive, and Cloud Computing*. Springer, Cham, Switzerland, 267–281. https://doi.org/10.1007/978-3-319-57186-7_21

[46] Eva García-Martín, Niklas Lavesson, Håkan Grahn, Emiliano Casalicchio, and Veselka Boeva. 2021. Energy-aware very fast decision tree. *Int. J. Data Sci. Anal.* 11, 2 (March 2021), 105–126. https://doi.org/10.1007/s41060-021-00246-4

[47] Ángel M. García-Vico and Francisco Herrera. 2021. A Preliminary Analysis on Software Frameworks for the Development of Spiking Neural Networks. In *Hybrid Artificial Intelligent Systems*. Springer, Cham, Switzerland, 564–575. https://doi.org/10.1007/978-3-030-86271-8_47

[48] Eva García-Martín, Crefeda Faviola Rodrigues, Graham Riley, and Håkan Grahn. 2019. Estimation of energy consumption in machine learning. *J. Parallel and Distrib. Comput.* 134 (2019), 75–88.

[49] Kent Gauen, Rohit Rangan, Anup Mohan, Yung-Hsiang Lu, Wei Liu, and Alexander C. Berg. 2017. Low-power image recognition challenge. In *2017 22nd Asia and South Pacific Design Automation Conference (ASP-DAC)*. 99–104. https://doi.org/10.1109/ASPDAC.2017.7858303

[50] Stefanos Georgiou, Maria Kechagia, Tushar Sharma, Federica Sarro, and Ying Zou. 2022. Green AI: do deep learning frameworks have different costs? In *ICSE '22: Proceedings of the 44th International Conference on Software Engineering*. Association for Computing Machinery, New York, NY, USA, 1082–1094. https://doi.org/10.1145/3510003.3510221

[51] Santosh Gondi and Vineel Pratap. 2021. Performance and Efficiency Evaluation of ASR Inference on the Edge. *Sustainability* 13, 22 (Nov. 2021), 12392. https://doi.org/10.3390/su132212392

[52] Achim Guldner, Sandro Kreten, and Stefan Naumann. 2021. Exploration and systematic assessment of the resource efficiency of Machine Learning.. In *GI-Jahrestagung*. 287–299.

[53] Abhishek Gupta, Camylle Lanteigne, and Sara Kingsley. 2020. SECure: A Social and Environmental Certificate for AI Systems. *ICML 2020 Challenges in Deploying and monitoring Machine Learning Systems Workshop* (June 2020). https://doi.org/10.48550/arXiv.2006.06217 arXiv:2006.06217

[54] María Gutiérrez, Ma Ángeles Moraga, and Félix García. 2022. Analysing the energy impact of different optimisations for machine learning models. In *2022 International Conference on ICT for Sustainability (ICT4S)*. IEEE, 46–52.

[55] Başak Güler and Aylin Yener. 2021. Energy-Harvesting Distributed Machine Learning. In *2021 IEEE International Symposium on Information Theory (ISIT)*. 320–325. https://doi.org/10.1109/ISIT45174.2021.9518045

[56] Raluca Maria Hampau, Maurits Kaptein, Robin van Emden, Thomas Rost, and Ivano Malavolta. 2022. An Empirical Study on the Performance and Energy Consumption of AI Containerization Strategies for Computer-Vision Tasks on the Edge. In *Proceedings of the International Conference on Evaluation and Assessment in Software Engineering 2022* (Gothenburg, Sweden) (*EASE '22*). Association for Computing Machinery, New York, NY, USA, 50–59. https://doi.org/10.1145/3530019.3530025

[57] Walid A. Hanafy, Tergel Molom-Ochir, and Rohan Shenoy. 2021. Design Considerations for Energy-efficient Inference on Edge Devices. In *e-Energy '21: Proceedings of the Twelfth ACM International Conference on Future Energy Systems*. Association for Computing Machinery, New York, NY, USA, 302–308. https://doi.org/10.1145/3447555.3465326

[58] Soheil Hashemi, Nicholas Anthony, Hokchhay Tann, R Iris Bahar, and Sherief Reda. 2017. Understanding the impact of precision quantization on the accuracy and energy of neural networks. In *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2017*. IEEE, 1474–1479.

[59] Peter Henderson, Jieru Hu, Joshua Romoff, Emma Brunskill, Dan Jurafsky, and Joelle Pineau. 2020. Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning. *Journal of Machine Learning Research* 21, 248 (2020), 1–43. http://jmlr.org/papers/v21/20-312.html

[60] Miro Hodak and Ajay Dholakia. 2021. Recent Efficiency Gains in Deep Learning: Performance, Power, and Sustainability. In *2021 IEEE International Conference on Big Data (Big Data)*. IEEE, 2040–2045.

[61] Hamzaoui Ikhlasse, Duthil Benjamin, Courboulay Vincent, and Medromi Hicham. 2022. Recent implications towards sustainable and energy efficient AI and big data implementations in cloud-fog systems: A newsworthy inquiry. *Journal of King Saud University - Computer and Information Sciences* 34, 10, Part A (Nov. 2022), 8867–8887. https://doi.org/10.1016/j.jksuci.2021.11.002

[62] Nitthilan Kanappan Jayakodi, Syrine Belakaria, Aryan Deshwal, and Janardhan Rao Doppa. 2020. Design and optimization of energy-accuracy tradeoff networks for mobile platforms via pretrained deep models. *ACM Transactions on Embedded Computing Systems (TECS)* 19, 1 (2020), 1–24.

[63] Wandri Jooste, Rejwanul Haque, and Andy Way. 2022. Knowledge Distillation: A Method for Making Neural Machine Translation More Efficient. *Information* 13, 2 (Feb. 2022), 88. https://doi.org/10.3390/info13020088

[64] Sorin Liviu Jurj, Flavius Oprițoiu, and Mircea Vladutiu. 2020. Environmentally-friendly metrics for evaluating the performance of deep learning models and systems. In *International Conference on Neural Information Processing*. Springer, 232–244.

[65] Petra Jääskeläinen, Daniel Pargman, and André Holzapfel. 2022. On the environmental sustainability of AI art(s). In *Eighth Workshop on Computing within Limits*. https://doi.org/10.21428/bf6fb269.c46375fa

[66] Lynn H. Kaack, Priya L. Donti, Emma Strubell, George Kamiya, Felix Creutzig, and David Rolnick. 2022. Aligning artificial intelligence with climate change mitigation. *Nat. Clim. Change* 12 (June 2022), 518–527. https://doi.org/10.1038/s41558-022-01377-7

[67] Minsu Kim, Walid Saad, Mohammad Mozaffari, and Merouane Debbah. 2022. On the Tradeoff between Energy, Precision, and Accuracy in Federated Quantized Neural Networks. In *ICC 2022 - IEEE International Conference on Communications*. 2194–2199. https://doi.org/10.1109/ICC45855.2022.9838362

[68] Young Geun Kim and Carole-Jean Wu. 2020. Autoscale: Energy efficiency optimization for stochastic edge inference using reinforcement learning. In *2020 53rd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)*. IEEE, 1082–1096.

[69] Pascal D. König, Stefan Wurster, and Markus B. Siewert. 2022. Consumers are willing to pay a price for explainable, but not for green AI. Evidence from a choice-based conjoint analysis. *Big Data & Society* 9, 1 (Jan. 2022), 20539517211069632. https://doi.org/10.1177/20539517211069632

[70] Adam Krzywaniak, Pawel Czarnul, and Jerzy Proficz. 2022. GPU Power Capping for Energy-Performance Trade-Offs in Training of Deep Convolutional Neural Networks for Image Recognition. In *International Conference on Computational Science*. Springer, 667–681.

[71] Mohit Kumar, Xingzhou Zhang, Liangkai Liu, Yifan Wang, and Weisong Shi. 2020. Energy-Efficient Machine Learning on the Edges. In *2020 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*. 912–921. https://doi.org/10.1109/IPDPSW50202.2020.00153

[72] Jaeha Kung, Duckhwan Kim, and Saibal Mukhopadhyay. 2015. A power-aware digital feedforward neural network platform with backpropagation driven approximate synapses. In *2015 IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED)*. IEEE, 85–90.

[73] Loïc Lannelongue, Jason Grealey, and Michael Inouye. 2021. Green Algorithms: Quantifying the Carbon Footprint of Computation. *Adv. Sci.* 8, 12 (June 2021), 2100707. https://doi.org/10.1002/advs.202100707

[74] Nicola Lenherr, René Pawlitzek, and Bruno Michel. 2021. New universal sustainability metrics to assess edge intelligence. *Sustainable Computing: Informatics and Systems* 31 (Sept. 2021), 100580. https://doi.org/10.1016/j.suscom.2021.100580

[75] Da Li, Xinbo Chen, Michela Becchi, and Ziliang Zong. 2016. Evaluating the Energy Efficiency of Deep Convolutional Neural Networks on CPUs and GPUs. In *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom)*. 477–484. https://doi.org/10.1109/BDCloud-SocialCom-SustainCom.2016.76

[76] Anne-Laure Ligozat, Julien Lefevre, Aurélie Bugeau, and Jacques Combaz. 2022. Unraveling the Hidden Environmental Impacts of AI Solutions for Environment Life Cycle Assessment of AI Solutions. *Sustainability* 14, 9 (April 2022), 5172. https://doi.org/10.3390/su14095172

[77] Liangkai Liu, Jiamin Chen, Marco Brocanelli, and Weisong Shi. 2019. E2M: an energy-efficient middleware for computer vision applications on autonomous mobile robots. In *Proceedings of the 4th ACM/IEEE Symposium on Edge Computing*. 59–73.

[78] Michele Magno, Michael Pritz, Philipp Mayer, and Luca Benini. 2017. Deep-Emote: Towards multi-layer neural networks in a low power wearable multi-sensors bracelet. In *2017 7th IEEE International Workshop on Advances in Sensors and Interfaces (IWASI)*. IEEE, 32–37.

[79] Susmita Dey Manasi, Farhana Sharmin Snigdha, and Sachin S Sapatnekar. 2020. NeuPart: Using analytical models to drive energy-efficient partitioning of CNN computations on cloud-connected mobile clients. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 28, 8 (2020), 1844–1857.

[80] Andrea McIntosh, Safwat Hassan, and Abram Hindle. 2019. What can Android mobile app developers do about the energy consumption of machine learning? *Empir. Software Eng.* 24, 2 (April 2019), 562–601. https://doi.org/10.1007/s10664-018-9629-2

[81] Sachin Mehta, Mohammad Rastegari, Linda Shapiro, and Hannaneh Hajishirzi. 2019. Espnetv2: A light-weight, power efficient, and general purpose convolutional neural network. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 9190–9200.

[82] Thaha Mohammed, Aiiad Albeshri, Iyad Katib, and Rashid Mehmood. 2020. UbiPriSEQ—Deep Reinforcement Learning to Manage Privacy, Security, Energy, and QoS in 5G IoT HetNets. *Appl. Sci.* 10, 20 (Oct. 2020), 7120. https://doi.org/10.3390/app10207120

[83] Elena Morotti, Davide Evangelista, and Elena Loli Piccolomini. 2021. A green prospective for learned post-processing in sparse-view tomographic reconstruction. *Journal of Imaging* 7, 8 (2021), 139.

[84] David Patterson, Joseph Gonzalez, Urs Hölzle, Quoc Le, Chen Liang, Lluis-Miquel Munguia, Daniel Rothchild, David R. So, Maud Texier, and Jeff Dean. 2022. The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink. *Computer* 55, 7 (2022), 18–28. https://doi.org/10.1109/MC.2022.3148714

[85] Natasa Perucica and Katarina Andjelkovic. 2022. Is the future of AI sustainable? A case study of the European Union. *Transforming Government: People, Process and Policy* 16, 3 (June 2022), 347–358. https://doi.org/10.1108/TG-06-2021-0106

[86] Supadchaya Puangpontip and Rattikorn Hewett. 2020. Energy Usage of Deep Learning in Smart Cities. In *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE, 1143–1148.

[87] Lucas Høyberg Puvis de Chavannes, Mads Guldborg Kjeldgaard Kongsbak, Timmie Rantzau, and Leon Derczynski. 2021. Hyperparameter Power Impact in Transformer Language Model Training. In *Proceedings of the Second Workshop on Simple and Efficient Natural Language Processing*. Association for Computational Linguistics, Virtual, 96–118. https://doi.org/10.18653/v1/2021.sustainlp-1.12

[88] Crefeda Faviola Rodrigues, Graham Riley, and Mikel Luján. 2018. SyNERGY: An energy measurement and prediction framework for Convolutional Neural Networks on Jetson TX1. In *Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA)*. The Steering Committee of The World Congress in Computer Science, 375–382.

[89] Friederike Rohde, Maike Gossen, Josephin Wagner, and Tilman Santarius. 2021. Sustainability challenges of Artificial Intelligence and Policy Implications. *Ökologisches Wirtschaften-Fachzeitschrift* 36, O1 (2021), 36–40.

[90] Bita Darvish Rouhani, Azalia Mirhoseini, and Farinaz Koushanfar. 2016. De-Light: Adding Energy Dimension To Deep Neural Networks. In *ISLPED '16: Proceedings of the 2016 International Symposium on Low Power Electronics and Design*. Association for Computing Machinery, New York, NY, USA, 112–117. https://doi.org/10.1145/2934583.2934599

[91] Kanokwan Rungsuptaweekoon, Vasaka Visoottiviseth, and Ryousei Takano. 2017. Evaluating the power efficiency of deep learning inference on embedded GPU systems. In *2017 2nd International Conference on Information Technology (INCIT)*. IEEE, 1–5.

[92] Roy Schwartz, Jesse Dodge, Noah A. Smith, and Oren Etzioni. 2020. Green AI. *Commun. ACM* 63, 12 (Nov. 2020), 54–63. https://doi.org/10.1145/3381831

[93] Ilia Shumailov, Yiren Zhao, Daniel Bates, Nicolas Papernot, Robert Mullins, and Ross Anderson. 2021. Sponge examples: Energy-latency attacks on neural networks. In *2021 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, 212–231.

[94] Martino Sorbaro, Qian Liu, Massimo Bortone, and Sadique Sheik. 2020. Optimizing the energy consumption of spiking neural networks for neuromorphic applications. *Frontiers in neuroscience* 14 (2020), 662.

[95] Dimitrios Stamoulis, Ermao Cai, Da-Cheng Juan, and Diana Marculescu. 2018. HyperPower: Power- and memory-constrained hyper-parameter optimization for neural networks. In *2018 Design, Automation, and Test in Europe Conference*. 19–24. https://doi.org/10.23919/DATE.2018.8341973

[96] Dimitrios Stamoulis, Ting-Wu Rudy Chin, Anand Krishnan Prakash, Haocheng Fang, Sribhuvan Sajja, Mitchell Bognar, and Diana Marculescu. 2018. Designing adaptive neural networks for energy-constrained image classification. In *2018 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*. ACM, 1–8.

[97] Emma Strubell, Ananya Ganesh, and Andrew McCallum. 2019. Energy and Policy Considerations for Deep Learning in NLP. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, 3645–3650. https://doi.org/10.18653/v1/P19-1355

[98] Yuyang Sun, Zhixin Ou, Juan Chen, Xinxin Qi, Yifei Guo, Shunzhe Cai, and Xiaoming Yan. 2021. Evaluating Performance, Power and Energy of Deep Neural Networks on CPUs and GPUs. In *National Conference of Theoretical Computer Science*. Springer, 196–221.

[99] Yuxuan Sun, Sheng Zhou, and Deniz Gündüz. 2020. Energy-aware analog aggregation for federated learning with redundant data. In *ICC 2020-2020 IEEE International Conference on Communications (ICC)*. IEEE, 1–7.

[100] Guglielmo Tamburrini. 2022. The AI Carbon Footprint and Responsibilities of AI Scientists. *Philosophies* 7, 1 (Jan. 2022), 4. https://doi.org/10.3390/philosophies7010004

[101] Yudong Tao, Rui Ma, Mei-Ling Shyu, and Shu-Ching Chen. 2020. Challenges in Energy-Efficient Deep Neural Network Training With FPGA. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

[102] Roberto Verdecchia, Luís Cruz, June Sallou, Michelle Lin, James Wickenden, and Estelle Hotellier. 2022. Data-Centric Green AI An Exploratory Empirical Study. In *2022 International Conference on ICT for Sustainability (ICT4S)*. IEEE, 35–45. https://doi.org/10.1109/ICT4S55073.2022.00015

[103] Chengcheng Wan, Muhammad Santriaji, Eri Rogers, Henry Hoffmann, Michael Maire, and Shan Lu. 2020. {ALERT}: Accurate learning for energy and timeliness. In *2020 USENIX Annual Technical Conference (USENIX ATC 20)*. 353–369.

[104] Cong Wang, Bin Hu, and Hongyi Wu. 2022. Energy Minimization for Federated Asynchronous Learning on Battery-Powered Mobile Devices via Application Co-running. In *2022 IEEE 42nd International Conference on Distributed Computing Systems (ICDCS)*. 939–949. https://doi.org/10.1109/ICDCS54860.2022.00095

[105] Qu Wang, Yong Xiao, Huixiang Zhu, Zijian Sun, Yingyu Li, and Xiaohu Ge. 2021. Towards Energy-efficient Federated Edge Intelligence for IoT Networks. In *2021 IEEE 41st International Conference on Distributed Computing Systems Workshops (ICDCSW)*. 55–62. https://doi.org/10.1109/ICDCSW53096.2021.00016

[106] Yu Wang, Rong Ge, and Shuang Qiu. 2020. Energy-Aware DNN Graph Optimization. *Resource-Constrained Machine Learning (ReCoML) Workshop of MLSys 2020 Conference* (May 2020). https://doi.org/10.48550/arXiv.2005.05837 arXiv:2005.05837

[107] Yue Wang, Ziyu Jiang, Xiaohan Chen, Pengfei Xu, Yang Zhao, Yingyan Lin, and Zhangyang Wang. 2019. E2-Train: Training State-of-the-art CNNs with Over 80% Energy Savings. In *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Eds.), Vol. 32. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2019/file/663772ea088360f95bac3dc7ffb841be-Paper.pdf

[108] Simon Wenninger, Can Kaymakci, Christian Wiethe, Jörg Römmelt, Lukas Baur, Björn Häckel, and Alexander Sauer. 2022. How Sustainable is Machine Learning in Energy Applications? – The Sustainable Machine Learning Balance Sheet. https://aisel.aisnet.org/wi2022/sustainable_it/sustainable_it/1

[109] Martina Willenbacher, Torsten Hornauer, and Volker Wohlgemuth. 2021. Rebound Effects in Methods of Artificial Intelligence. In *Advances and New Trends in Environmental Informatics*. Springer, Cham, Switzerland, 73–85. https://doi.org/10.1007/978-3-030-88063-7_5

[110] Carole-Jean Wu, Ramya Raghavendra, Udit Gupta, Bilge Acun, Newsha Ardalani, Kiwan Maeng, Gloria Chang, Fiona Aga, Jinshi Huang, Charles Bai, et al. 2022. Sustainable AI: Environmental implications, challenges and opportunities. *Proceedings of Machine Learning and Systems* 4 (2022), 795–813.

[111] Haichuan Yang, Yuhao Zhu, and Ji Liu. 2019. Energy-constrained compression for deep neural networks via weighted sparse projection and layer input masking. *International Conference on Learning Representations (ICLR)* (2019).

[112] Tien-Ju Yang, Yu-Hsin Chen, and Vivienne Sze. 2017. Designing Energy-Efficient Convolutional Neural Networks Using Energy-Aware Pruning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[113] Xiangyu Yang, Sheng Hua, Yuanming Shi, Hao Wang, Jun Zhang, and Khaled B. Letaief. 2020. Sparse Optimization for Green Edge AI Inference. *Journal of Communications and Information Networks* 5, 1 (2020), 1–15. https://doi.org/10.23919/JCIN.2020.9055106

[114] Zhaohui Yang, Mingzhe Chen, Walid Saad, Choong Seon Hong, and Mohammad Shikh-Bahaei. 2020. Energy efficient federated learning over wireless communication networks. *IEEE Transactions on Wireless Communications* 20, 3 (2020), 1935–1949.

[115] Chunrong Yao, Wantao Liu, Weiqing Tang, Jinrong Guo, Songlin Hu, Yijun Lu, and Wei Jiang. 2021. Evaluating and analyzing the energy efficiency of CNN inference on high-performance GPU. *Concurrency and Computation: Practice and Experience* 33, 6 (2021), e6064.

[116] Barzan A Yosuf, Sanaa H Mohamed, Mohammed M Alenazi, Taisir EH El-Gorashi, and Jaafar MH Elmirghani. 2021. Energy-Efficient AI over a Virtualized Cloud Fog Network. In *Proceedings of the Twelfth ACM International Conference on Future Energy Systems*. 328–334.

[117] Jia-Ruei Yu, Chun-Hsien Chen, Tsung-Wei Huang, Jang-Jih Lu, Chia-Ru Chung, Ting-Wei Lin, Min-Hsien Wu, Yi-Ju Tseng, and Hsin-Yao Wang. 2022. Energy Efficiency of Inference Algorithms for Clinical Laboratory Data Sets: Green Artificial Intelligence Study. *J. Med. Internet Res.* 24, 1 (Jan. 2022), e28036. https://doi.org/10.2196/28036

[118] Boyu Zhang, Azadeh Davoodi, and Yu Hen Hu. 2018. Exploring Energy and Accuracy Tradeoff in Structure Simplification of Trained Deep Neural Networks. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 8, 4 (2018), 836–848. https://doi.org/10.1109/JETCAS.2018.2833383

[119] Xingzhou Zhang, Yifan Wang, and Weisong Shi. 2018. pCAMP: Performance Comparison of Machine Learning Packages on the Edges. In *USENIX workshop*

on hot topics in edge computing (HotEdge 18).

[120] Sha Zhu, Kaoru Ota, and Mianxiong Dong. 2021. Green AI for IIoT: Energy Efficient Intelligent Edge Computing for Industrial Internet of Things. *IEEE Trans. Green Commun. Networking* 6, 1 (Aug. 2021), 79–88. https://doi.org/10.1109/TGCN.2021.3100622