

Innovating for Tomorrow: The Convergence of SE and Green AI

Luís Cruz
L.Cruz@tudelft.nl
Delft University of Technology
The Netherlands

Xavier Franch Gutierrez
xavier.franch@upc.edu
Universitat Politècnica de Catalunya
Barcelona, Spain

Silverio Martínez-Fernández
silverio.martinez@upc.edu
Universitat Politècnica de Catalunya
Barcelona, Spain

ABSTRACT

The latest advancements in machine learning, specifically in foundation models, are revolutionizing the frontiers of existing software engineering (SE) processes. This is a bi-directional phenomena, where 1) software systems are now challenged to provide AI-enabled features to their users, and 2) AI is used to automate tasks within the software development lifecycle. In an era where sustainability is a pressing societal concern, our community needs to adopt a long-term plan enabling a conscious transformation that aligns with environmental sustainability values. In this paper, we reflect on the impact of adopting environmentally friendly practices to create AI-enabled software systems and make considerations on the environmental impact of using foundation models for software development.

CCS CONCEPTS

- Software and its engineering; • Computing methodologies
→ Artificial intelligence;

KEYWORDS

Green AI, Green Software, Sustainability, Software Engineering

ACM Reference Format:

Luis Cruz, Xavier Franch Gutierrez, and Silverio Martínez-Fernández. 2024. Innovating for Tomorrow: The Convergence of SE and Green AI. In *Proceedings of International Workshop on Software Engineering in 2030 (SE 2030)*. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/nnnnnnnn>.

1 INTRODUCTION

Software-related CO₂ emissions from the ICT sector currently account for 2.1%–3.9% of global emissions [17]. In today's context with the widespread use of AI systems, there have been many calls from industry leaders and AI experts admitting a further increase of these emissions. OpenAI chief executive Sam Altman warned that the next wave of generative AI systems will consume vastly more power than expected [8]. Hugging Face Climate Lead Sasha Luccioni has shown the scalability impact of AI systems' inference: "While inference on a single example requires much less computation than that required to train the same model, inference happens far more frequently than model training — as many as billions of

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SE 2030, November 2024, Porto de Galinhas (Brazil)

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-XXXX-X/18/06

<https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

times a day for a model powering a popular user-facing product such as Google Translate" [27]. Focusing only on AI systems, in a middle-ground scenario, by 2027 AI servers could use between 85 to 134 terawatt hours (Twh) annually [10]. That is similar to the annual electricity usage of countries such as Argentina, the Netherlands and Sweden individually and constitutes approximately 0.5 % of the world's current electricity consumption [15].

The convergence of Software Engineering (SE) and AI is a bi-directional phenomena raising key sustainability concerns. On the one hand, software systems of diverse domains are now challenged to provide AI-enabled features to their users. To reduce the emissions from these AI systems from their development and training, to their usage and inference, and to its retirement, emerging AI software development lifecycles shall incorporate energy-aware capabilities. On the other hand, the potential of using AI and foundation models to automate tasks within the software development lifecycle is very promising, with dedicated events such as AIware¹ and LLM4Code². To fully embrace these technologies, the concerns regarding the emitted emissions from its usage shall be understood.

Therefore, the objective of this position paper is twofold:

- (1) Identifying the trans-disciplinary dimensions and dichotomies in which the research from the SE community shall contribute to build greener AI systems, as well as reasoning on the evolution of SE practices in such dimensions and dichotomies.
- (2) Discussing the environmental sustainability of one application domain of AI systems: generative AI for SE tasks like generation of requirements, architecture, or code in which humans and intelligent agents jointly create software.

2 OVERVIEW OF GREEN AI

We define **Green AI** as a trans-disciplinary field that aims to make AI systems environmentally sustainable. Environmental sustainability of software (including AI systems) refers to engineering systems having minimal impact in our planet throughout its whole lifecycle [3]. We distinguish this from AI for Sustainability or AI for Green, where AI is used to make different domains more environmentally friendly – e.g., using AI to make agriculture more sustainable. We argue that it is important to make a clear separation between Green AI and AI for Green as they require different approaches and different scientific backgrounds.

Given the complexity of AI systems, Green AI needs to be tackled from different angles, having each of these angles an important contribution to the overall carbon emissions of the systems. Data, AI models, and the code of the software systems are the foundations of AI systems [40]. As such, to build and maintain sustainable AI systems, we shall focus in each of these three aspects. As depicted

¹<https://2024.aiwareconf.org/> (visited on April 5, 2024)

²<https://llm4code.github.io/> (visited on April 5, 2024)

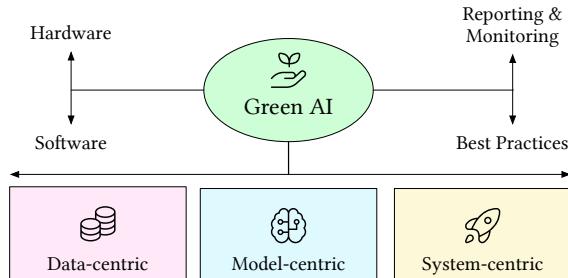


Figure 1: Overview of Green AI

in Figure 1, we divide Green AI across three major dimensions: data-centric, model-centric, and system-centric. These dimensions are pinpointed below in Sections 2.1 to 2.3. The figure also intersects Green AI with two dichotomies: hardware vs. software and reporting & monitoring vs. best practices.

Hardware ↔ Software. With the current trend of specializing hardware for particular AI tasks, practitioners can no longer be agnostic of the hardware they use to run their software. The choice of hardware is essential to ensure that a particular AI pipeline runs properly with no issues downstream. Hence, practitioners now face the challenge of having to be hardware experts. Moreover, previous research demonstrates how different hardware configurations can significantly impact the energy consumption of training deep-learning models [11]. However, these differences are not obvious and require meticulous trade-off analysis, which is sub-optimal in real-world scenarios. Hardware design for AI poses more concerns. Since AI chips are specialized to particular AI tasks, updates to AI models might lead to different hardware requirements. This can significantly reduce the lifetime of hardware, posing challenges to the emerging “Right to repair” legislation efforts being discussed by the EU, the US, the United Nations and so on. Not surprising, the waste created by disposing hardware devices – coined as e-waste – is already a pressing environmental problem, as reported by the technology activist Murzyn³.

Reporting & Monitoring ↔ Best Practices. It is important that software systems are designed in a way that it is possible to report and monitor sustainability indicators [24]). Practitioners need to provide accurate information about the carbon footprint of their experiments and models. This can be done e.g. via meta-data in repositories such as Hugging Face or using more elaborated and structured templates such as the concept of ID-card [2]. On the other side of the spectrum, we have best practices [21]. They provide practitioners with strategies that can be adopted to reduce the environmental footprint of their software.

These two poles complement each other. Effective best practices can only be established when reliable measurement mechanisms are in place. Monitoring and reporting not only incentivize practitioners to adopt these best practices but also provide essential feedback on their implementation and effectiveness.

³Murzyn is on a mission to report the hazards from handling electronic waste, a responsibility that has been delegated to Global South countries: <https://murzyn.asia/en/mission-00-24-42/> (visited on April 5, 2024)

2.1 Data-centric Green AI

Data-centric Green AI revolves around preparing data in a way that is expected to reduce the overall energy footprint of AI systems without hindering their performance. Previous work has addressed this from multiple ways. Reducing data dimensionality, using feature selection or stratified random sampling already leads to linear gains in energy consumption [48].

Further strategies are being explored. Active learning & coresets extraction have promising potential by identifying the most valuable examples that contribute to the learning process and excluding redundant or non-informative ones with confidence [39].

Knowledge transfer/sharing consists of using knowledge rather than solely relying on data to train a model. For instance, existing pre-trained models can be used to train new models. Preliminary research showcases improvements in energy consumption by a factor of 15 with this technique [51].

Dataset distillation or dataset condensation involves synthesizing a smaller dataset derived from the original dataset, aiming to train a model on this reduced set while achieving test accuracy comparable to that of a model trained on the original dataset. This approach differs from coresset extraction by focusing on synthesizing informative samples rather than selecting existing ones. Data distillation has been used to reduce the size of the MNIST image set [26] into just 10 synthetic distilled images (one per class) and achieve close to original performance [52].

Curriculum learning consists of presenting training examples to the model in a specific order, starting with simpler examples and gradually increasing the difficulty of the examples as training progresses. This strategy provides the model with a structured learning experience that mimics how humans learn and requires less iterations to converge, with time reductions of 70% [33].

2.2 Model-centric Green AI

Model-centric Green AI means developing experimental research to improve the AI model performance and energy efficiency. The target is to build and optimize AI models that can achieve similar outcomes while requiring fewer resources.

A notable example is the BLOOM model, whose carbon footprint have been studied during the training and inference stages to compete with other Large Language Models (LLMs) [28]. A more recent example are small language models (SLMs), like “Phi” that has shown performance comparable to models 5x larger [22]. Indeed, there is evidence that there is not always a trade-off between green and performance metrics. A mining repository study on 1.417 models from Hugging Face could not find a correlation between model performance and carbon emissions of ML models [5]. This shows the potential for lightweight AI models architectures [6, 30].

Another model-centric strategy is the compression of existing models (aka AI model optimization): structured and unstructured pruning, quantization and binarization, efficient training and inference of models under different resource constraints. In this regard, Pytorch and TensorFlow are already offering AI models optimization libraries [37, 47].

2.3 System-centric Green AI

System-centric green AI refers to the decisions we can take regarding the software architecture and serving of ML systems to make them environmentally sustainable. The ecological footprint of an AI model extends beyond its algorithmic design to encompass the entire system infrastructure that supports it. For instance, suboptimal hardware choices can significantly inflate energy consumption [11], while implementing batched inference can enhance energy efficiency [54]. It is essential to recognize the continuum from cloud to edge computing in considering energy footprints [12]. This continuum encompasses end-user devices (edge), cloud servers, and the network infrastructure that connects them. Each component plays a role in determining the overall environmental impact of AI systems, highlighting the need for holistic considerations in system design and deployment for sustainability.

We anticipate that sustainability challenges will initially be tackled with a system-centric approach, given the overlap between AI systems and traditional software systems. However, we argue that emerging challenges lie in the data-centric and model-centric domains of AI, where SE plays a critical role [41, 42]. The significance of data quality has already been acknowledged for its impact on the reliability, robustness, efficiency, and trustworthiness of modern software systems [32]. From a model-centric perspective, there is an unexplored potential to employ SE to challenge the status quo. For example, adopting energy-conscious practices for model adaptation [35], hyperparameter tuning [55], and so on.

3 REFLECTIONS ON THE FUTURE OF SE

When reflecting about the future of SE, the emergence of Green AI introduces several challenges that warrant reflection. In this section, we delve into these potential challenges, providing *context* and outlining their *implications for SE*.

1. Consideration of the business case. *Context.* Not all problems are alike. Diverse business cases require different positioning with respect to environmental sustainability. An image processing model for cancer diagnose shall prioritize precision no matter what is the carbon footprint required for training. On the contrary, a film affinity recommender system will hardly be considered life-critical, therefore little gains of accuracy at the cost of dramatic increments of energy consumption is not justified; instead, digital sobriety⁴ should start to be the norm, and not the exception. *Implications for SE.* Business case elaboration and analysis lies at the heart of the requirements engineering area [16]. Its application to AI systems, often denoted by RE4AI, is currently a mainstream in SE, yielding to a number of actionable findings, e.g., definition of new scopes for requirements [44] or new quality requirement types [20]. However, as Habibullah et al. remark, there is a research gap in establishing trade-offs among quality requirements [20]. Understanding and specifying these trade-offs would help to translate the business case into quantifiable requirements, associated metrics, and optimization functions allowing for a proper validation of the AI system.

2. Consolidation of fundamental concepts. *Context.* Design of AI systems is still a young research area. As a consequence, fundamental concepts are still not fully understood and lead to

⁴<https://www.vox.com/climate/2024/3/28/24111721/ai-uses-a-lot-of-energy-experts-expect-it-to-double-in-just-a-few-years> (visited on April 5, 2024)

inconsistent or even wrong terminology. For instance, we may find papers using the terms ‘energy efficiency’ with the meaning of ‘energy consumption’, or defining metrics with an incorrect measurement unit. This situation hampers communication and interdisciplinary collaboration [31]. *Implications for SE.* Although there are several excellent works that focus on establishing a consolidated terminology for environmental sustainability [18], the truth is that they have not made their way in the SE community. Observing what happens in other areas, a standard playing the same role as for instance ISO/IEC 25010 in the field of software quality [1] could be a significant asset towards this consolidation. Eventually, such a standard could be the upcoming ISO/IEC 20226 on AI environmental sustainability⁵.

3. Monitoring sustainability. *Context.* Collecting energy data is not a trivial task. Even after collecting it, practitioners have to process and analyze energy data and tracing it back to the AI pipelines, to hotspots and make necessary adjustments if possible. To make matters worse, improving sustainability is often a trade-off problem: we want to reduce energy consumption without hindering other requirements of the system. For example, improving energy efficiency at the cost of privacy is likely unacceptable, depending on the use case. On another note, the impacts of AI systems on the environment can hardly be simplified by looking at energy consumption alone. For the same energy consumption of executing a AI software, the carbon footprint can vary depending on the time, region, and the type electricity used to power the servers at that particular. Besides, carbon footprint, we also have the problematic of water footprint – i.e., the water being evaporated due to the software execution; mostly due to cooling down servers. Moreover, embodied carbon footprint also plays an important role. GPU usage needs to be optimized to make sure hardware is not stalling idle most of the time during their lifetime [53]. *Implications to SE.* There is an open challenge of developing user-friendly energy monitoring tools that seamlessly integrate with various environments, including Edge AI devices, and virtual ones like Docker containers. These tools should not only enable detailed analysis but also offer a comprehensive overview of the energy consumption associated with software products. Additionally, they need to accommodate diverse resource metrics and provide actionable insights to facilitate environmentally-conscious decision-making throughout the development and maintenance phases of these systems [18].

4. Clarification of roles' involvement. *Context.* The emergence of AI demands knowledge and skills that go beyond those traditionally required in classical SE. The consideration of new roles and the interactions among them have been subject of investigation. Specially, the fit of data scientists in the SE team has been thoroughly investigated [23, 50]. In the context of Green AI, knowledge on environmental sustainability is a must for the optimal consideration of carbon footprint in system design. However, experts on sustainability are hardly involved in the development of AI systems *Implications for SE.* Same as sometimes software engineers criticize data scientists for building code not adhering to SE principles, software engineers may be criticized by addressing environmental sustainability without having profound expertise in the team. The role of environmental sustainability expert should be

⁵<https://www.iec.ch/blog/importance-sustainable-ai> (visited on April 5, 2024)

explicitly recognized in the Green AI development team, same as data scientist or domain expert are. This role can be played either by real sustainability professionals, or by software engineers who have acquired along time the necessary knowledge and skills.

5. Changes in the ML lifecycle. *Context.* There has been ongoing efforts to reshape the development lifecycle of modern software systems to cover the inclusion of AI-enabled features. However, along this push to redefine the development lifecycle of AI-enabled software, environmental sustainability is often overlooked [19, 25, 29, 38]. *Implications.* It is quintessential to adding sustainability concerns across the whole lifecycle, from the very early stages until the retirement of software and/or respective AI models. There is an unexplored potential to employ SE to challenge the status quo. For example, model retraining is still the default approach for model adaptation, leaving out other strategies that may be more cost-effective and sustainable [34–36]. Moreover, adhering to the three R’s principle—reduce, reuse, and recycle (in that specific order)—is essential in fostering sustainability practices within AI systems [4]. We need frameworks that help practitioners 1) opt for competitive alternatives to AI when available (reduce), 2) consume existing models instead of training their own (reuse), 3) monitor model degradation and consider model adaptation strategies (recycle) before disposing a model and training a new one [35].

6. Quest for open science. *Context.* Nowadays, the SE community is fully aware of the importance of open science as a basic principle supporting transparency and replication [43]. Initiatives at all levels (journals, conferences, national research programs, ...) push the general adoption of open science by researchers. In the AI field, it has become customary to give access to (preprint of) papers, data, software and models in public repositories such as arxiv, Zenodo, GitHub and Hugging Face. However, there is a lack of clear guidelines on what type of information to store in these repositories. For instance, in the case of Green AI, Castaño et al. report that carbon emission related information is only marginally reported in the most popular model repository, Hugging Face [7]. This means that in spite of community willingness to go open, it is still difficult to replicate experiments or to conduct long-term cohort studies. *Implications for SE.* The research community shall produce consolidated and agreed guidelines to allow software engineers be systematic and rigorous in the documentation of the sustainability dimension of their experiments. Two different actions along this direction are needed: (i) determine what is the information related to environmental sustainability to be described, (ii) creating domain-specific description sheets that can be used to consolidate such information. An example in this direction is the ID-card proposed by Abualhaija et al. in the NLP domain [2]. Another example focusing on environmental sustainability is GAISSALabel, which gathers a repository of training and inference emissions footprints of ML models [13]. Alternatively, the information can be coded as meta-data in the repositories, although usually this will result in less comprehensive descriptions.

7. The role of Education *Context.* Equally important is the imperative to educate the next generation of AI developers in sustainable AI literacy. Green SE is still a niche topic [9], with very few available materials to support educators. Moreover, the rapidly evolving landscape of AI technologies presents challenges in adapting AI and Computer Science curricula to incorporate Green AI

practices. In fact, many Green AI practices are very recent and experimental [21]. *Implications for SE.* The education of SE must explicitly incorporate topics on responsible AI and, more specifically, the environmental considerations of AI systems. Educational materials should be openly accessible to empower aspiring software engineers to prioritize sustainability within their organizations.

8. Construction of theories. *Context.* There is a large corpus of research works that report experiments of any kind with an environmental sustainability dimension. In fact, the original set of 98 studies considered by Verdecchia et al. in their 2023’s paper [49] is falling short today after only a couple of years. While every individual experiment may be valuable per se, their results still remain mainly de-aggregated and are difficult to be integrated into a holistic body of knowledge, i.e., a theory [45]. *Implications for SE.* Software engineers shall gradually adopt what Stol and Fitzgerald call ‘a theory-focused research approach’ [46]. Building theories is not only important because of the outcome, but also a theory construction process brings the need of critical thinking to researchers and make them aware of the need of better reporting their experiments as to allow building theories on solid basis. A critical point is the consideration of context [14], determined by the independent and confounding variables of the experiment. Modeling context in the theory may facilitate, for instance, integrating the source of energy in the calculation of carbon emissions.

4 GREEN FOUNDATION MODELS FOR SE.

The usage of large language code models is a powerful tool that is becoming more and more popular. It has been reported recently by Microsoft that, amongst GitHub Copilot users, 40% of the code they commit is ‘AI-generated and unmodified’⁶. Foundation models open the stage for developing software systems without developers creating a single line of code. Whole software systems can soon be generated from natural language prompts.

A wide adoption of foundation models in coding activities poses an important question on the impact it has on the environmental sustainability of SE. On one hand, there is a clear reduction of number of developers that can lead to smaller carbon footprints inherited from having teams working together. On the other hand, there is a concerning energy overhead from continuously prompting code generation models to converge to a version of the software that can go to production. Furthermore, it is not clear whether generated code will abide by energy efficiency coding practices. State of the art code generation models are purely statistical: we argue that the most frequent coding patterns are not necessarily the most energy efficient.

Foundation models ought to provide transparency in terms of sustainability indicators at different stages of the software lifecycle. Such indicators should paint a clear picture on the ecological footprint of training those models, using them, but also using their outputs. Strategies should be adopted to make sure these models are more likely to learn energy-efficient coding practices than regular ones.

⁶Scott Guthrie, Executive of the Cloud and AI group at Microsoft, as of April 5, 2024: <https://www.microsoft.com/en-us/Investor/events/FY-2023/Morgan-Stanley-TMT-Conference>

REFERENCES

- [1] ISO/IEC 25010:2023. 2023. Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE). <https://www.iso.org/standard/78176.html>
- [2] Sallam Abualhaija, Fatma Basak Aydemir, Fabiano Dalpiaz, Davide Dell'Anna, Alessio Ferrari, Xavier Franch, and Davide Fucci. 2023. Replication and Verifiability in Requirements Engineering: the NLP for RE Case. *Accepted in TOSEM* (2023). <https://doi.org/10.48550/ARXIV.2304.10265> arXiv:2304.10265
- [3] Christoph Becker, Ruzanna Chitchyan, Leticia Duboc, Steve Easterbrook, Birgit Penzenstadler, Norbert Seyff, and Colin C Venters. 2015. Sustainability design and software: The karlskrona manifesto. In *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*, Vol. 2. IEEE, 467–476.
- [4] Sadia Sultana Bristy, Tooba Azam, Mukta Marion Islam, Raisur Rahman, Ahmed Wasif Reza, and Mohammad Shamsul Arefin. 2023. Green Cloud Computing: A Sustainable Energy-Efficiency Approach for Business Rapidity and the Environment. In *International Conference on Intelligent Computing & Optimization*. Springer, 312–327.
- [5] Joel Castaño, Silverio Martínez-Fernández, Xavier Franch, and Justus Bogner. 2023. Exploring the Carbon Footprint of Hugging Face's ML Models: A Repository Mining Study. In *ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, ESEM 2023, New Orleans, LA, USA, October 26–27, 2023*. IEEE, 1–12. <https://doi.org/10.1109/ESEM56168.2023.10304801>
- [6] Roger Creus Castanyer, Silverio Martínez-Fernández, and Xavier Franch. 2024. Which design decisions in AI-enabled mobile applications contribute to greener AI? *Empirical Software Engineering* 29, 1 (2024), 2.
- [7] Joel Castaño, Silverio Martínez-Fernández, Xavier Franch, and Justus Bogner. 2023. Exploring the Carbon Footprint of Hugging Face's ML Models: A Repository Mining Study. In *2023 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*. 1–12. <https://doi.org/10.1109/ESEM56168.2023.10304801>
- [8] Kate Crawford. 2024. Generative AI is guzzling water and energy. Available online at: <https://www.nature.com/articles/d41586-024-00478-x>, last accessed on 05.04.2024.
- [9] Luís Cruz. 2022. The Five Dimensions of Sustainable Software Engineering and How Education Can Help! <http://luiscruz.github.io/2022/01/01/sustainable-se-intro.html>. <https://doi.org/10.6084/m9.figshare.25036772> Blog post..
- [10] Alex de Vries. 2023. The growing energy footprint of artificial intelligence. *Joule* 7, 10 (2023), 2191–2194.
- [11] Santiago del Rey, Silverio Martínez-Fernández, Luís Cruz, and Xavier Franch. 2023. Do DL models and training environments have an impact on energy consumption?. In *2023 49th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*. IEEE, 150–158.
- [12] Santiago del Rey, Silverio Martínez-Fernández, and Xavier Franch. 2023. A review on green deployment for Edge AI-Abstract. (2023).
- [13] Pau Duran, Joel Castaño, Cristina Gómez, and Silverio Martínez-Fernández. 2024. GAISSALabel: A tool for energy labeling of ML models. *arXiv preprint arXiv:2401.17150* (2024).
- [14] Tore Dybå, Dag I.K. Sjøberg, and Daniela S. Cruzes. 2012. What works for whom, where, when, and why? on the role of context in empirical software engineering. In *Proceedings of the ACM-IEEE International Symposium on Empirical Software Engineering and Measurement (Lund, Sweden) (ESEM '12)*. Association for Computing Machinery, New York, NY, USA, 19–28. <https://doi.org/10.1145/2372251.2372256>
- [15] Delger Erdenesaa. 2023. A.I. Could Soon Need as Much Electricity as an Entire Country. Available online at: <https://www.nytimes.com/2023/10/10/climate/ai-could-soon-need-as-much-electricity-as-an-entire-country.html>, last accessed on 05.04.2024.
- [16] Xavier Franch, Andreas Jedlitschka, and Silverio Martínez-Fernández. 2023. A requirements engineering perspective to AI-based systems development: A vision paper. In *International Working Conference on Requirements Engineering: Foundation for Software Quality*. Springer, 223–232.
- [17] Charlotte Freitag, Mike Berners-Lee, Kelly Widdicks, Bran Knowles, Gordon S Blair, and Adrian Friday. 2021. The real climate and transformative impact of ICT: A critique of estimates, trends, and regulations. *Patterns* 2, 9 (2021).
- [18] Achim Guldner, Rabea Bender, Coral Calero, Giovanni S Fernando, Markus Funke, Jens Gröger, Lorenz M Hilti, Julian Hörschemeyer, Geerd-Dietger Hoffmann, Dennis Junger, et al. 2024. Development and evaluation of a reference measurement model for assessing the resource and energy efficiency of software products and components—Green Software Measurement Model (GSMM). *Future Generation Computer Systems* (2024).
- [19] Mark Haakman, Luis Cruz, Hennie Huijgens, and Arie Van Deursen. 2021. AI lifecycle models need to be revised: An exploratory study in fintech. *Empirical Software Engineering* 26, 5 (2021), 95.
- [20] Khan Mohammad Habibullah, Gregory Gay, and Jennifer Horkoff. 2023. Non-functional requirements for machine learning: understanding current use and challenges among practitioners. *Requir. Eng.* 28, 2 (2023), 283–316. <https://doi.org/10.1007/S00766-022-00395-3>
- [21] Heli Järvenpää, Patricia Lago, Justus Bogner, Grace Lewis, Henry Muccini, and İpek Ozkaya. 2023. A Synthesis of Green Architectural Tactics for ML-Enabled Systems. *arXiv preprint arXiv:2312.09610* (2023).
- [22] Mojtaba Javaheripour and Sébastien Bubeck. 2023. Phi-2: The surprising power of small language models. Available online at: <https://www.microsoft.com/en-us/research/blog/phi-2-the-surprising-power-of-small-language-models>, last accessed on 05.04.2024.
- [23] Miryung Kim, Thomas Zimmermann, Robert DeLine, and Andrew Begel. 2018. Data Scientists in Software Teams: State of the Art and Challenges. *IEEE Transactions on Software Engineering* 44, 11 (2018), 1024–1038. <https://doi.org/10.1109/TSE.2017.2754374>
- [24] Michael Kläs, Jens Heidrich, Jürgen Münch, and Adam Trendowicz. 2009. CQML scheme: A classification scheme for comprehensive quality model landscapes. In *2009 35th Euromicro Conference on Software Engineering and Advanced Applications*. IEEE, 243–250.
- [25] Dominik Kreuzberger, Niklas Kühl, and Sebastian Hirsch. 2023. Machine learning operations (mlops): Overview, definition, and architecture. *IEEE access* (2023).
- [26] Yann LeCun, Corinna Cortes, and CJ Burges. 2010. MNIST handwritten digit database. *ATT Labs [Online]*. Available: <http://yann.lecun.com/exdb/mnist> 2 (2010).
- [27] Alexandra Sasha Luccioni, Yacine Jernite, and Emma Strubell. 2023. Power hungry processing: Watts driving the cost of ai deployment? *arXiv preprint arXiv:2311.16863* (2023).
- [28] Alexandra Sasha Luccioni, Sylvain Viguerie, and Anne-Laure Ligozat. 2023. Estimating the carbon footprint of bloom, a 176b parameter language model. *Journal of Machine Learning Research* 24, 253 (2023), 1–15.
- [29] Lucy Ellen Lwakatare, Ivica Crnkovic, and Jan Bosch. 2020. DevOps for AI—Challenges in Development of AI-enabled Applications. In *2020 international conference on software, telecommunications and computer networks (SoftCOM)*. IEEE, 1–6.
- [30] Silverio Martínez-Fernández, Xavier Franch, and Francisco Durán. 2023. Towards green AI-based software systems: an architecture-centric approach (GAISSA). In *2023 49th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*. IEEE, 432–439.
- [31] N. Nahar, H. Zhang, G. Lewis, S. Zhou, and C. Kastner. 2023. A Meta-Summary of Challenges in Building Products with ML Components – Collecting Experiences from 4758+ Practitioners. In *2023 IEEE/ACM 2nd International Conference on AI Engineering – Software Engineering for AI (CAIN)*. IEEE Computer Society, Los Alamitos, CA, USA, 171–183. <https://doi.org/10.1109/CAIN58948.2023.00034>
- [32] Phu H Nguyen, Sagar Sen, Beatriz Bretones-Cassoli, Nicolas Jourdan, and Maria Chiara Magnanini. 2023. Software Engineering and AI for Data Quality in Cyber-Physical Systems/Internet of Things-SEA4DQ'22 Report. *ACM SIGSOFT Software Engineering Notes* 48, 1 (2023), 108–111.
- [33] Emmanouil Antonios Platanios, Otilia Stretcu, Graham Neubig, Barnabas Poczos, and Tom M Mitchell. 2019. Competence-based curriculum learning for neural machine translation. *arXiv preprint arXiv:1903.09848* (2019).
- [34] Lorena Poenaru-Olaru, Luis Cruz, Jan S Rellermeyer, and Arie Van Deursen. 2023. Maintaining and monitoring AIOps models against concept drift. In *2023 IEEE/ACM 2nd International Conference on AI Engineering–Software Engineering for AI (CAIN)*. IEEE, 98–99.
- [35] Lorena Poenaru-Olaru, Natalia Karpova, Luis Cruz, Jan Rellermeyer, and Arie van Deursen. [n. d.]. Maintenance Techniques for Anomaly Detection AIOps Solutions. *2024 IEEE/ACM 2nd International Conference on AI Engineering–Software Engineering for AI (CAIN)* ([n. d.]).
- [36] Lorena Poenaru-Olaru, June Sallou, Luis Cruz, Jan S Rellermeyer, and Arie Van Deursen. 2023. Retrain AI Systems Responsibly! Use Sustainable Concept Drift Adaptation Techniques. In *2023 IEEE/ACM 7th International Workshop on Green And Sustainable Software (GREENS)*. IEEE, 17–18.
- [37] PyTorch. 2024. Quantization - PyTorch 2.2. Available online at: <https://pytorch.org/docs/stable/quantization.html>, last accessed on 05.04.2024.
- [38] Dhia Elhaq Rziz, Foyzul Hassan, and Marouane Kessentini. 2022. An empirical study on ML DevOps adoption, trends, efforts, and benefits analysis. *Information and Software Technology* 152 (2022), 107037.
- [39] S. Salehi and A. Schmeink. 5555. Data-Centric Green Artificial Intelligence: A Survey. *IEEE Transactions on Artificial Intelligence* 1, 01 (sep 5555), 1–18. <https://doi.org/10.1109/TAI.2023.3315272>
- [40] Danilo Sato, Arif Wider, and Christoph Windheuser. 2019. Continuous Delivery for Machine Learning. Available online at: <https://martinfowler.com/articles/cd4ml.html>, last accessed on 05.04.2024.
- [41] Arumoy Shome, Luis Cruz, and Arie Van Deursen. 2022. Data smells in public datasets. In *Proceedings of the 1st International Conference on AI Engineering: Software Engineering for AI*. 205–216.
- [42] Arumoy Shome, Luis Cruz, and Arie van Deursen. 2024. Data vs. Model Machine Learning Fairness Testing: An Empirical Study. *DeepTest 2024* (2024).
- [43] Forrest Shull, Jeffrey C. Carver, Sira Vegas, and Natalia Juristo Juzgado. 2008. The role of replications in Empirical Software Engineering. *Empir. Softw. Eng.* 13, 2 (2008), 211–218. <https://doi.org/10.1007/S10664-008-9060-1>
- [44] Julien Siebert, Lisa Joeckel, Jens Heidrich, Koji Nakamichi, Kyoko Ohashi, Isao Namba, Rieko Yamamoto, and Mikio Aoyama. 2020. Towards Guidelines for

- Assessing Qualities of Machine Learning Systems. In *Quality of Information and Communications Technology*, Martin Shepperd, Fernando Brito e Abreu, Alberto Rodrigues da Silva, and Ricardo Pérez-Castillo (Eds.). Springer International Publishing, Cham, 17–31.
- [45] Dag I. K. Sjøberg, Tore Dybå, Bente C. D. Anda, and Jo E. Hannay. 2008. *Building Theories in Software Engineering*. Springer London, London, 312–336. https://doi.org/10.1007/978-1-84800-044-5_12
- [46] Klaas-Jan Stol and Brian Fitzgerald. 2013. Uncovering theories in software engineering. In *2013 2nd SEMAT Workshop on a General Theory of Software Engineering (GTSE)*. 5–14. <https://doi.org/10.1109/GTSE.2013.6613863>
- [47] TensorFlow. 2024. TensorFlow Model Optimization. Available online at: https://www.tensorflow.org/model_optimization?hl=en, last accessed on 05.04.2024.
- [48] 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.
- [49] Roberto Verdecchia, June Sallou, and Luis Cruz. 2023. A systematic review of Green AI. *WIREs Data. Mining. Knowl. Discov.* 13, 4 (2023). <https://doi.org/10.1002/WIDM.1507>
- [50] Andreas Vogelsang and Markus Borg. 2019. Requirements Engineering for Machine Learning: Perspectives from Data Scientists. In *2019 IEEE 27th International Requirements Engineering Conference Workshops (REW)*. 245–251.
- <https://doi.org/10.1109/REW.2019.00050>
- [51] Paul Walsh, Jhilam Bera, Vibhu Saujanya Sharma, Vikrant Kaulgud, Raghotham M Rao, and Orlaith Ross. 2021. Sustainable ai in the cloud: Exploring machine learning energy use in the cloud. In *2021 36th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW)*. IEEE, 265–266.
- [52] Tongzhou Wang, Jun-Yan Zhu, Antonio Torralba, and Alexei A Efros. 2018. Dataset distillation. *arXiv preprint arXiv:1811.10959* (2018).
- [53] Carole-Jean Wu, Ramya Raghavendra, Udit Gupta, Bilge Acun, Newsha Ardalani, Kiwan Maeng, Gloria Chang, Fiona Aga Behram, James Huang, Charles Bai, et al. 2021. Sustainable ai: Environmental implications, challenges and opportunities. *arXiv preprint arXiv:2111.00364* (2021).
- [54] Tim Yarally, Luís Cruz, Daniel Feitosa, June Sallou, and Arie van Deursen. 2023. Batching for Green AI-An Exploratory Study on Inference. In *2023 49th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*. IEEE, 112–119.
- [55] Tim Yarally, Luis Cruz, Daniel Feitosa, June Sallou, and Arie Van Deursen. 2023. Uncovering energy-efficient practices in deep learning training: Preliminary steps towards green ai. In *2023 IEEE/ACM 2nd International Conference on AI Engineering–Software Engineering for AI (CAIN)*. IEEE, 25–36.