# Sustainable Model Selection Is a Design Problem: Supporting Exploration and Comparison in ML Practice

EYA BEN CHAABEN, LISN, Université Paris-Saclay, CNRS, Inria, France
JANIN KOCH, UMR 9189 CRIStAL, Univ. Lille, Inria, CNRS, Centrale Lille, France

CCS Concepts: • **Social and professional topics** → **Sustainability**; • **Human-centered computing** → *HCI theory, concepts and models.*

Additional Key Words and Phrases: Sustainable HCI, Eco-feedback, Machine Learning Practice, Model Selection,

**ACM Reference Format:**
Eya Ben chaaben and Janin Koch. 2026. Sustainable Model Selection Is a Design Problem: Supporting Exploration and Comparison in ML Practice. In *Proceedings of Workshop paper submitted to 'HCI-TERRA: HCI Towards Environmentally Responsible AI Workshop' at the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26).* ACM, New York, NY, USA, 5 pages.

## 1 Introduction

Increased energy, water, and material resource use are some of the environmental costs associated with the fast growth in generative AI and machine learning. Even while most people now agree that these effects exist, the majority of mitigation initiatives still concentrate on technological improvements to hardware and software. This framing is insufficient according to recent demands in HCI, and addressing AI's environmental effect asks for socio-technical approaches that include developers, academics, and impacted communities in addition to solutionist narratives. Especially, everyday development decisions, such as model selection for a particular use case, have a substantial impact on AI's environmental footprint. Developers frequently pick between systems with significantly different energy consumption, carbon emissions, and resource needs, but sustainability is rarely a key consideration in these decisions. This is more than just a missing metrics issue from an HCI standpoint. Interfaces, defaults, heuristics, organizational guidelines, and views of cutting-edge practice all influence decisions. Therefore, it is important to consider both the technical and social circumstances in which model selection occurs in order to support environmentally responsible AI.

We argue that HCI can contribute to sustainable AI by reshaping these decision environments. Rather than optimizing models directly, my work focuses on how HCI can support developers in exploring alternatives, interpreting trade-offs, and reflecting on environmental impacts through comparative and context-aware representations. To showcase potential directions, we outline some empirical studies and design explorations examining ML workflows, sustainability awareness, comparative eco-feedback for model trade-off representations.

Authors' Contact Information: Eya Ben chaaben, eya.ben-chaaben@inria.fr, LISN, Université Paris-Saclay, CNRS, Inria , Paris, France; Janin Koch, UMR 9189 CRIStAL, Univ. Lille, Inria, CNRS, Centrale Lille , Lille, France, janin.koch@inria.fr.

## 2 Related Work

Model selection is a complex, multi-dimensional challenge that requires balancing technical performance, developer constraints, and increasingly, sustainability considerations [2]. While automated machine learning (AutoML) and interactive interfaces have improved accessibility and efficiency, they often overlook nuanced practitioner needs such as interpretability, computational resource limits, and environmental impact. Much work on sustainable and responsible ML has focused on efficiency and measurement, including quantifying the energy cost of large models [18], proposing Green AI principles [16], developing sustainability indicators for early-stage evaluation [15], and exploring transfer learning approaches [23]. Tools such as CodeCarbon [7], Tracarbon [20], and Green Algorithms [10] help estimate energy use and carbon emissions, while interactive frameworks like Symphony support practitioners in identifying issues in ML workflows [1].

However, despite progress in efficiency and transparency, current approaches provide limited guidance for selecting less resource-intensive alternatives during development, reinforcing a dominant "bigger is better" culture in AI [21]. Industrial systems such as Facebook's FBLearner [6] and Google's TFX [11] support scalable ML pipelines, but largely prioritize performance and deployment. In contrast, HCI research increasingly calls for sufficiency, emphasizing whether a model is appropriate and necessary for a given task rather than simply more powerful [8, 19]. Together, these perspectives highlight the need for rethinking tools and practices that not only optimize models or expose impacts, but also help practitioners identify relevant alternatives.

Model selection is also shaped by practical constraints and developer practices. Developers typically prioritize accuracy and loss [9, 14, 26], yet feasibility depends on hardware availability, preferred frameworks, runtime limits, explainability requirements, legal compliance, and trust considerations [5, 12, 13, 17, 22, 25]. Poor model choices can lead to wasteful training and long-term maintenance challenges [4, 24], while data quality remains a limiting factor regardless of architecture [24]. At the same time, the rapid growth of model repositories (Hugging Face alone hosts over 1.7 million models) creates an overwhelming landscape of alternatives [3]. To support navigation of this space, researchers have started developing interactive systems to demonstrate how interface design can lower barriers to ML use and help users articulate requirements, yet they remain largely focused on performance-oriented trade-offs. Factors such as necessary model size, energy use, long-term computational cost, and environmental impact are rarely integrated into actionable recommendations, leaving practitioners without adequate support for incorporating sustainability into model selection decisions.

## 3 Understanding ML Workflows and Sustainability Awareness

ML model selection is not a purely technical optimization process, but a situated practice shaped by developers' familiarity, available tools, organizational pressures, and incomplete information about sustainability impacts. Through interviews with 13 ML developers [2], you find that participants primarily select models based on accuracy, interpretability, and prior experience, often defaulting to large or familiar architectures while rarely evaluating energy or infrastructure impacts. Sustainability awareness exists, especially among senior practitioners, but it seldom translates into everyday practice; younger developers in particular reported little exposure to sustainability during their education. Model choices are influenced by accessible literature, model repositories, AutoML systems, and increasingly LLMs, reinforcing a trend toward complex multi-purpose models even when simpler alternatives might suffice. Developers also face fragmented guidance: performance benchmarks are available, but sustainability data are largely missing or inconsistent, making it difficult to reason about trade-offs. As a result, model selection is driven more by availability and perceived norms than by structured reflection on task requirements or long-term environmental consequences.

Importantly, our work highlighted that sustainability emerges as a socio-technical issue spanning education, tooling, organizational expectations, and broader cultural narratives about AI. Participants described tensions between performance, interpretability, and sustainability, with accuracy typically prioritized unless environmental costs could be shown not to affect outcomes. Few considered social or economic sustainability, focusing mainly on energy consumption, and even this was hard to assess due to hidden infrastructure impacts such as GPU production, water use, and data center operations. We argue that addressing these challenges requires more than better metrics: it calls for changes in ML education to support critical reflection on model choice, standardized reporting of model consumption, and HCI-led tools that help developers explore alternatives and understand trade-offs in context. Our implications emphasize that sustainable ML depends on collective practices, including shared benchmarks, transparency, and organizational support, rather than individual goodwill alone. By positioning model selection as a key intervention point, our work reframes sustainability as embedded in everyday development workflows, and calls on both ML and HCI communities to redesign socio-technical systems so that choosing smaller, more appropriate models becomes visible, supported, and legitimate in practice

### 3.1 Supporting Sustainable Model Exploration

Building on our earlier findings that ML model selection is shaped by familiarity, performance norms, and limited visibility into sustainability impacts, we frame sustainable model exploration as a socio-technical design challenge rather than a purely technical optimization problem. Our recent work explores interactive model selection interfaces designed to help developers articulate project goals and constraints in order to surface model alternatives that are both suitable and more sustainable. Rather than ranking models primarily by popularity or benchmark performance, the system integrates task guidance, hardware constraints, and estimated energy consumption directly into the selection process. This approach reflects the observation that developers rarely exclude sustainability intentionally; instead, it remains structurally underrepresented in existing tools, which prioritize accuracy and familiarity while leaving environmental considerations implicit.

Findings from the first comparative studies with professional ML developers show that guided exploration can support three interrelated processes: expanding awareness of novel and task-appropriate models, building trust in recommendations through transparency and project alignment, and prompting reflection on practitioners' roles in sustainability. We saw that surfacing energy implications encouraged practitioners to reflect on their own responsibility in model choice, highlighting that sustainability is not only a matter of tooling but also of professional practice. These early results position sustainable model exploration as a design space that can be explored by the HCI community for more actively supporting responsible AI.

### 3.2 Comparative Eco-Feedback for ML Developers

Eco-feedback has traditionally focused on presenting users with information about their own environmental impact, such as energy consumption or carbon emissions, often through dashboards or single-option indicators. While such feedback can raise awareness, prior work shows that information alone rarely leads to sustained behavioral change. Looking at supporting decision support for ML developers highlighted that a critical moment instead lies in how people weigh alternatives. By reviewing a large corpus of academic and industrial tools, we found that most eco-feedback systems emphasize isolated numerical metrics, with limited support for comparing options or situating actions within broader contexts. Comparative eco-feedback reframes sustainability as a decision-making process: designs explicitly contrast alternatives (e.g., routes, products, devices) or benchmark behavior against references such as past performance or peers. These comparisons make trade-offs tangible and can activate social and motivational mechanisms, yet current practice remains fragmented and dominated by quantitative displays.

Translating these insights to ML development suggests that sustainability support should move beyond reporting energy metrics toward comparative design interventions embedded in model selection workflows. Previous studies show that developers already reason in terms of trade-offs, e.g. between accuracy, interpretability, feasibility, yet sustainability is rarely integrated into these comparisons [2]. Exploring more comparative eco-feedback offers a way to surface environmental impact alongside performance, enabling practitioners to evaluate alternative models in relation to task requirements, infrastructure limits, and long-term costs. Rather than prescribing smaller models, this approach frames sustainability as an explicit dimension of everyday technical decision-making. Combined with guided exploration and task-aware recommendations, comparative eco-feedback is another promising way to support developers in moving beyond familiar or popular architectures and toward context-appropriate alternatives.

## 4 Future Directions

Looking ahead, HCI has a critical role to play in integrating sustainability into ML development workflows. Rather than presenting isolated metrics, we should drive to design tools that support meaningful comparison across performance, resource use, and environmental impact. Making impacts visible during model selection encourages reflection and facilitates the exploration of more efficient alternatives, transforming sustainability from an abstract concern to a practical design consideration.

Equally important is to broaden this support beyond individual developers to collective practices. Shared baselines, historical project comparisons, and team-level indicators can help organizations identify patterns of compute usage and incremental improvement over time. Comparative eco-feedback can also help teams align technical decisions with larger environmental goals, resulting in a better understanding and accountability. Ultimately, we argue for treating ML model selection as a socio-technical activity shaped by tools, norms, and values. HCI research is well-positioned to create interfaces that highlight trade-offs, support learning over time, and foster cultures of responsible AI development, transforming sustainability from a supplementary constraint to a core component of everyday ML practice.

## Acknowledgments

This work is supported by the European Union's Horizon Europe research and innovation programme (HORIZON-CL4-2021-HUMAN-01) under grant agreement No 101070408, project SustainML (Application Aware, Life-Cycle Oriented Model-Hardware Co-Design Framework for Sustainable, Energy Efficient ML Systems).

## References

[1] Alex Bäuerle, Ángel Alexander Cabrera, Fred Hohman, Megan Maher, David Koski, Xavier Suau, Titus Barik, and Dominik Moritz. 2022. Symphony: Composing interactive interfaces for machine learning. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–14.

[2] Eya Ben chaaben, Janin Koch, and Wendy E. Mackay. 2025. "Should I choose a smaller model?": Understanding ML Model Selection and Its Impact on Sustainability. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1008, 13 pages. doi:10.1145/3706598.3713240

[3] Rishi Bommasani. 2021. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258* (2021).

[4] Jie Ding, Vahid Tarokh, and Yuhong Yang. 2018. Model selection techniques: An overview. *IEEE Signal Processing Magazine* 35, 6 (2018), 16–34.

[5] Jaimie Drozdal, Justin Weisz, Dakuo Wang, Gaurav Dass, Bingsheng Yao, Changruo Zhao, Michael Muller, Lin Ju, and Hui Su. 2020. Trust in AutoML: exploring information needs for establishing trust in automated machine learning systems. In *Proceedings of the 25th international conference on*

*intelligent user interfaces*. 297–307.

[6] J. Dunn. 2016. *Introducing FBLearner Flow: Facebook's AI backbone*. Accessed: 2016.

[7] Wilson Friedler. 2021. Track and reduce CO2 emissions from your computing. https://codecarbon.io/.

[8] Lon Åke Erni Johannes Hansson, Teresa Cerratto Pargman, and Daniel Sapiens Pargman. 2021. A Decade of Sustainable HCI: Connecting SHCI to the Sustainable Development Goals. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 300, 19 pages. doi:10.1145/3411764.3445069

[9] Zezhen He and Yaron Shaposhnik. 2023. Visualizing the Implicit Model Selection Tradeoff. *Journal of Artificial Intelligence Research* 76 (2023), 829–881.

[10] Loïc Lannelongue, Jason Grealey, and Michael Inouye. 2021. Green algorithms: quantifying the carbon footprint of computation. *Advanced science* 8, 12 (2021), 2100707.

[11] Akshay Naresh Modi, Chiu Yuen Koo, Chuan Yu Foo, Clemens Mewald, Denis M Baylor, Eric Breck, Heng-Tze Cheng, Jarek Wilkiewicz, Levent Koc, Lukasz Lew, et al. 2017. Tfx: A tensorflow-based production-scale machine learning platform. *KDD 2017* (2017).

[12] Felix Neutatz, Felix Biessmann, and Ziawasch Abedjan. 2021. Enforcing constraints for machine learning systems via declarative feature selection: an experimental study. In *Proceedings of the 2021 International Conference on Management of Data*. 1345–1358.

[13] Rafiullah Omar. 2024. Energy-Efficient Development of ML-Enabled Systems: A Data-Centric Approach. In *Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering-Software Engineering for AI*. 259–263.

[14] Matthew Parry, A Philip Dawid, and Steffen Lauritzen. 2012. Proper local scoring rules. (2012).

[15] Vasileios Perifanis, Nikolaos Pavlidis, Selim F. Yilmaz, Francesc Wilhelmi, Elia Guerra, Marco Miozzo, Pavlos S. Efraimidis, Paolo Dini, and Remous-Aris Koutsiamanis. 2023. Towards Energy-Aware Federated Traffic Prediction for Cellular Networks. In *2023 Eighth International Conference on Fog and Mobile Edge Computing (FMEC)*. 93–100. doi:10.1109/FMEC59375.2023.10306017

[16] Roy Schwartz, Jesse Dodge, Noah A Smith, and Oren Etzioni. 2020. Green AI. *Commun. ACM* 63, 12 (2020), 54–63.

[17] Tor Sporsem. 2024. Discovering Explainability Requirements in ML-Based Software. In *Proceedings of the 2024 IEEE/ACM 46th International Conference on Software Engineering: Companion Proceedings*. 170–172.

[18] Emma Strubell, Ananya Ganesh, and Andrew McCallum. 2019. Energy and policy considerations for deep learning in NLP. *arXiv preprint arXiv:1906.02243* (2019).

[19] Thomas Thibault, Léa Mosesso, Camille Adam, Aurélien Tabard, Anaëlle Beignon, and Nolwenn Maudet. 2025. Environmental (in)considerations in the Design of Smartphone Settings. In *LIMITS 2025: 11th Workshop on Computing within Limits*. arXiv, Online, France. doi:10.48550/arXiv.2507.19094

[20] Florian Valeye. 2022. Tracarbon — Track your device's carbon footprint. https://medium.com/@florian.valeye/tracarbon-track-your-devices-carbon-footprint-fb051fcc9009/.

[21] Gael Varoquaux, Sasha Luccioni, and Meredith Whittaker. 2025. Hype, Sustainability, and the Price of the Bigger-is-Better Paradigm in AI. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)*. Association for Computing Machinery, New York, NY, USA, 61–75. doi:10.1145/3715275.3732006

[22] 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)*. IEEE, 245–251.

[23] 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)*. 265–266. doi:10.1109/ASEW52652.2021.00058

[24] Jiayi Wang, Chengliang Chai, Nan Tang, Jiabin Liu, and Guoliang Li. 2022. Coresets over multiple tables for feature-rich and data-efficient machine learning. *Proceedings of the VLDB Endowment* 16, 1 (2022), 64–76.

[25] Daniel Karl I Weidele, Justin D Weisz, Erick Oduor, Michael Muller, Josh Andres, Alexander Gray, and Dakuo Wang. 2020. AutoAIViz: opening the blackbox of automated artificial intelligence with conditional parallel coordinates. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*. 308–312.

[26] Yair Weiss and William T Freeman. 2007. What makes a good model of natural images?. In *2007 IEEE conference on computer vision and pattern recognition*. IEEE, 1–8.

Received 12 February 2026