Team 5.2. Standardizing Sustainability Indicators For AI Model Cards

Ka Man Ng (Matriculation number: 03742359)
Salma Tarik (Matriculation Number: 03751028)
Olivia Steen (Matriculation Number: 03750638)
Inkeri Halme (Matriculation Number: 14184223)
Mikaela Manolva (Matriculation Number: 03731143)

Introduction
Standardizing Sustainability Indicators For AI Model Cards

As AI models become increasingly sophisticated, it has been observed that large models demanding the greatest computational resources generally demonstrate the most significant improvements in accuracy, underscoring the alarming trend of consuming exceptionally large computational resources that require similarly substantial energy.[1] However, current awareness predominantly focuses on the sustainability metrics of AI models in isolation, omitting substantial energy and resources footprint that occur in multiple phases and dimensions of AI practices.

To enable industry-wide sustainability practices in the AI industry, this report presents the recommended Key Performance Indicators (KPIs) and indicators that should be included on AI Model Cards to comprehensively understand the sustainability impact of AI models. The report takes a holistic approach in building comprehensive sustainability assessments through the complete life cycle of AI, encompassing activities associated with model development, the infrastructural and energy frameworks facilitating AI practices, and ramifications beyond.

The following sections introduce sustainability indicators in four respective critical areas: **i) AI models, ii) data centers, iii) hardware,** and **iv) risks and governance implications connecting social and environmental sustainability.**

i) AI models

Sustainability indicators for AI models: Sustainability assessments should permeate data collection, model training, and deployment stages of AI models.

Data collection method disclosure and CO2 equivalent for computationally intensive datasets: While the energy cost during data collection is often overlooked due to complexities in its measurement, particularly when organizations repurpose existing data, it remains crucial to perform a footprint evaluation at this stage. This can be achieved by: 1) Disclosure of data collection methodologies (i.e. sensor-based, human-based, computationally-driven, hybrid, etc.) and 2) Disclosure of the carbon dioxide equivalent (CO2e) from all locales involved when computationally intensive datasets, such as those used in complex climate modeling, exceed a certain threshold (methodology to be explained below)

CO2 equivalent during AI model development: AI models undergo cycles of training and fine-tuning at this stage. While multiple factors can influence energy requirements — such as model complexity including the number of layers and parameters, model type, epochs, the volume and complexity of data, the application of optimization techniques — an effective indicator during this "model development" stage would consolidate the total energy use in order to streamline sustainability reporting.

---

[1] Strubell, E., Ganesh, A. and McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. arXiv (Cornell University). doi:https://doi.org/10.48550/arxiv.1906.02243.

Based on scientific literature, a method would be to aggregate the total power consumption of all computational hardware (GPUs and CPUs, etc.) in kilowatt-hours (kWh), and factor in the power efficiency index of the local data center — a component frequently omitted in carbon footprint trackers. The resulting metric is then converted to a carbon footprint using the local emission intensity of the country based on its energy mix.[2] [3] Considering that modern companies operate data centers in multiple countries, the total footprint across all data centers, discounting carbon offset credits, is proposed as the standard measure to prevent selective reporting.

CO2 equivalent during AI model deployment: In the post-training phase, energy consumption can grow exponentially during the AI product's deployment phase, especially if the product reaches billions of users.[4] Consequently, the carbon footprint during this phase should be quantified using the same methodology applied during the training stage.

ii) Data centers

Data centers are a critical component of an AI model's lifecycle. This section focuses on six key areas of impact regarding data centers: location, water usage, CO2e emissions, energy mix, biodiversity impacts, and energy use and waste heat. These indicators provide valuable insights for policymakers to assess the environmental footprint of AI models and make informed decisions.

Data Center Location: The geographic locations of data centers used to run the model should be disclosed on AI model cards. This information helps policymakers understand the unique effects data centers have on local communities and can contextualize needed regulations. It is important to include both data centers actively used to run the model during the reporting period and those locations used for prior period activities, such as training. Otherwise, companies may run the intensive training activities on cheap but high-impact data centers, while switching their model to cleaner centers after deployment. By capturing all locations used throughout the model's life, policymakers can evaluate the broader impact of AI models and ensure compliance with regional sustainability regulations.

Water Usage: Water usage is a critical aspect of the environmental impact of data centers. Prior to construction, an assessment of the impacts of water usage on the community should be conducted. The assessment should evaluate and disclose any negative impacts on the communities water resources, as well as future-looking risk assessments. This will help prevent the construction of water-hungry centers in water-stressed communities. This assessment should be updated during operation to monitor changes over time. Furthermore, the water usage during operations should be mapped and disclosed on AI model cards. Water usage metrics include water withdrawal, consumption, discharge and water usage efficiency.

---

[2] Budennyy, S.A., Lazarev, V.D., Zakharenko, N.N., Korovin, A.N., Plosskaya, O.A., Dimitrov, D.V., Akhripkin, V.S., Pavlov, I.V., Oseledets, I.V., Barsola, I.S., Egorov, I.V., Kosterina, A.A. and Zhukov, L.E. (2022). eco2AI: Carbon Emissions Tracking of Machine Learning Models as the First Step Towards Sustainable AI. Doklady Mathematics, [online] 106(S1), pp.S118–S128. doi:https://doi.org/10.1134/s1064562422060230

[3] Strubell, E., Ganesh, A. and McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. arXiv (Cornell University). doi:https://doi.org/10.48550/arxiv.1906.02243.

[4] Kaack, L.H., Donti, P.L., Strubell, E., Kamiya, G., Creutzig, F. and Rolnick, D. (2022). Aligning artificial intelligence with climate change mitigation. Nature Climate Change, 12(6), pp.518–527. doi:https://doi.org/10.1038/s41558-022-01377-7.

Furthermore, to encourage sustainable practices, the percentage of input water from recycled sources should be noted. This indicator highlights efforts and commitments to resource efficiency.

CO2e Emissions: Emissions from data centers should be disclosed in two forms: total emissions from day-to-day operations and emissions resulting from the construction of the centers. By quantifying and disclosing these emissions on AI model cards, policymakers can gain insights into the carbon footprint associated with running AI models. This information is crucial for assessing the environmental impact of AI models and developing strategies to mitigate emissions.

Energy Mix: To evaluate the transition towards renewable energy sources, companies should disclose their overall data center energy mix across various locations. This information helps policymakers understand the composition of energy sources and assess alignment with emissions goals. It also contextualizes the emissions and energy usage indicators by understanding from which source the energy is derived. By tracking and reporting the energy mix, stakeholders and policymakers can monitor progress towards renewable energy adoption goals.

Biodiversity Impacts: AI Model Cards should include an initial biodiversity assessment conducted prior to data center construction, as well as any updates made during operation. This assessment helps policymakers understand the impact on biodiversity and monitor progress towards biodiversity goals. By considering biodiversity impacts, the sustainability of AI models can be assessed more comprehensively. This indicator highlights the importance of preserving ecosystems when deciding where to construct and operate data centers.

Energy Use & Waste Heat: Energy usage during the reporting period should be disclosed on AI Model Cards. This goes hand-in-hand with the emissions and energy mix data. It includes metrics such as total consumption and power usage effectiveness.

Given that data centers generate significant amounts of waste heat, disclosing the cooling methods employed and any innovative ways waste heat is reused can incentivize greener practices. This information showcases efforts to reduce energy waste and highlights initiatives that benefit operations or the community through heat reuse. By considering energy use and waste heat, policymakers can encourage sustainable energy management and foster the development of energy-efficient AI models.

iii) Hardware

Measuring the hardware in data centers is crucial for several reasons. It helps in ensuring optimal performance, resource allocation, energy efficiency, and overall cost-effectiveness. Measuring the sustainability of hardware in data centers is becoming increasingly important due to the growing focus on environmental concerns and the need for sustainable practices in the industry. A deeper and specific look about key performance indicators have been done in the appendix.

Sustainability aspect of hardware can be divided based on the life cycle. That increases the simplicity of the future AI model card and ensures that every aspect of the supply chain has been taken into account. This study focuses on four main states: extraction, manufacturing, distribution and end-of-life -state.

The main focus in extraction is to focus on the quality of raw material, how extraction affects the environment, for example, measuring by Living Planet Index. Measuring sustainability helps identify opportunities for resource conservation, such as using recycled or low-impact materials, reducing water consumption, and measuring air, water and soil quality.
It is critical to take into account the whole supply chain of hardware and how sustainability it can be. Usage of recycled materials in manufacturing states can lead to better results by identifying the exact results of the actions. Studies show that recycled materials can be implemented for distribution. Allocating the correct figures can lead to better actions and helps to better sustainability performance.

The industry generates a substantial amount of electronic waste (e-waste) due to the constant upgrading and replacement of hardware. Measuring sustainability encourages data centers to consider the entire lifecycle of their hardware, including proper disposal and recycling practices, to minimize the environmental impact of e-waste. By allocating the whole process of e-waste aims to get information on how sustainable the solutions can be. By measuring the sustainability of hardware in data centers, organizations can make informed decisions to reduce their environmental impact, improve resource efficiency, comply with regulations, and align with sustainable business practices. This not only benefits the environment but also enhances their reputation, reduces operational costs, and supports long-term sustainability goals.

iv) Risk and Governance

As a final dimension of our model card, the scope of Risk and Governance is also included to highlight two main indirect impact indicators: Development Pre-Assessment, as well as Compliance and Governance. The goal is to take into consideration aspects of social sustainability.

Development Pre-Assessment addresses the preliminary decision-making process of the AI model. This indicator mainly serves to provide justification for applying artificial intelligence in consideration of other available approaches, such as human capacity, less intensive computational methods, and so on. Valid reasoning behind the development of the model should be presented, proving the necessary application of artificial intelligence within this system. Decisions disclosure and justification can be properly validated by conducting one of the methodologies suggested in our model card, such as Scenario Analysis, Delphi Method, or Normative Scenario Analysis[5] .

Furthermore, Development Pre-Assessment includes a Future Impact Evaluation, which provides insight into future repercussions of the AI model based on various factors. This

[5] Fauré, E., Arushanyan, Y., Ekener, E., Miliutenko, S., & Finnveden, G. (2017). Methods for assessing future scenarios from a sustainability perspective. *European Journal of Futures Research*, *5*(1), 1-20. https://doi.org/10.1007/s40309-017-0121-9

indicator acts as a safeguard in regard to not only general concerns but also tecl-specific ones. For instance, the assessment ranges from broader considerations, such as determining the social and ecological risks related to the application of the model, to individual observations, such as the risks and benefits associated with API usage, or the justification behind closed/open API. This indicator can be quantified using the recommended methods, more particularly Material Assessment, Normative Scenario Development, and Scenario Analysis. The goal of a Future Impact Assessment is to potentially guide users, stakeholder, policy makers, and other actors to make more informed decisions regarding potential risks and needed mitigation efforts throughout the lifecycle of the AI model [6].

Next, the **Compliance and Governance indicator** first emphasizes the disclosure and transparency of the AI model in relation to any information regarding updates of the system. As displayed on the basic information section of the model card template, an ''Update'' field must be filled with the time of the last update of the AI Model. This disclosure and transparency specification should comply with the standard of The General Data Protection Regulation (GDPR).
Finally, the labor risk association part of the Compliance and Governance indicator poses the question: ''Can your company confirm its adherence to the standards set forth by the International Labour Organization (ILO) (or any other comparable regulations, guidelines, frameworks) regarding labor laws, including those related to automation, AI, and labor protection?''

Conclusion and Learnings:

Over the course of the project week, Team 5.2 worked together in order to draft a comphrehensive policy brief with recommendations of environmental indicators for standardized AI model cards.

The discussions were fruitful and engaging, as each member came from different academic and cultural backgrounds. Different insights and perspectives were offered, which only enriched the collaborative efforts.

In terms of challenges, the limited timeframe of the project posed certain difficulties in terms of research and the execution of the assignment. However, the team was able to overcome this challenge by delegating specific tasks, as well as setting personal deadlines. Furthermore, the variety of viewpoints at times made decision-making challenging. The team rose above this obstacle through clear communication, responsive listening, and constructive criticism. In the end, the members were able to come up with a policy brief that was benefited with well-rounded recommendations by the variety of educational and cultural backgrounds.

---

[6] Stahl, B. C., Brooks, L., Hatzakis, T., Santiago, N., & Wright, D. (2023). Exploring ethics and human rights in artificial intelligence – A Delphi study. Technological Forecasting and Social Change, 191, 122502. https://doi.org/10.1016/j.techfore.2023.122502

The assignment required an interdisciplinary approach, as it combined different expert fields such as policy analysis, environmental science, and technology. The team effectively addressed the sustainability challenges within the project by delving into each respective field, then combining them in order to craft cohesive and applicable model card suggestions.

In regards to technicalities, the main challenge of this project has been tracking the environmental footprint of the indicators. It can be tricky due to the interconnections between footprints on a local AI model level and footprints on a physical infrastructure level. Through extensive and divided research, the team learned to integrate both, as well as to extend it to the national level, e.g. factoring the energy mix of the specific country where the data centers are.

Finally, the group can agree that, although the project week was challenging in terms of time constraints and the variety of viewpoints, it was also extremely rewarding. It served as a learning experience not only in terms of project-related research, but also teamwork. Through communication, time-management, teamwork, and extensive research, Group 5.2's shared passion for the development of sustainable AI practices delivered high-quality results.