Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

Towards Sustainability Model Cards

Gwendal Jouneaux[1] and Jordi Cabot[1,2]

gwendal.jouneaux@list.lu jordi.cabot@list.lu

[1] Luxembourg Institute of Science and Technology
[2] University of Luxembourg

2nd Workshop on Green-Aware Artificial Intelligence
co-located with ECAI 2025

Bologna – 26/10/2025

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

Disclaimer : Models

Model	AI model	Quality model
Abstraction of some aspect of a system created to serve particular purposes *e.g.,*mechanical analysis	Machine learning model performing predictions/inferences	Model representing the dimensions and metrics enabling quality assessment

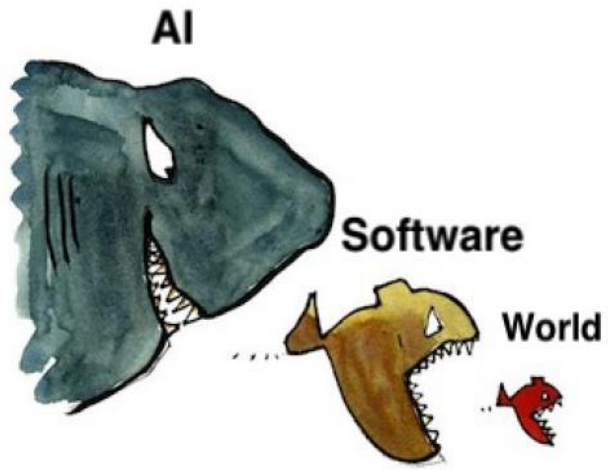
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

Sustainability in ICT is a critical topic

21% of the world electricity demand expectation in 2030

Analysis of AI models carbon footprint [1]

- Carbon footprint of AI models should be reduced
- Model reporting effort needs to be increased
- Prioritize computational efficiency



"Software is eating the world"
Andreessen Horowitz, HP (2011)

"Software is eating the world, but AI is going to eat software"
Jensen Huang, Nvidia CEO (2017)

[1] Energy and policy considerations for deep learning in NLP, Strubel *et al.*

Model Cards for Model Reporting

"Model cards are short documents accompanying trained machine learning models that provide benchmarked evaluation in a variety of conditions, […] that are relevant to the intended application domains"

Limitations:

- Do not address sustainability aspects per se

Model Cards Dimensions

- AI Model Details
- Intended Use
- Factors
- Metrics
- Evaluation Data
- Training Data
- Quantitative Analyses
- Ethical Considerations
- Caveats and Recommendations

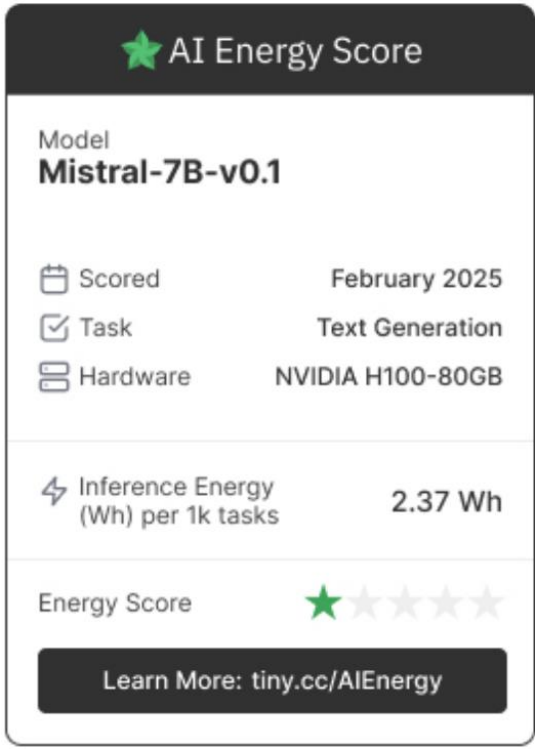
AI Energy Score

For each evaluated AI model reports:

- Inference energy consumption
- Inference task benchmarked
- Hardware used

Limitations:

- Carbon footprint will depend on the server location
- Training costs are not evaluated/reported



Objectives of Sustainability Model Cards

- Provide an AI sustainability quality model to enable model comparison
  - Define a set of dimensions and metrics
- Provide tool support for automatic processing
  - Define a notation to encode Sustainability Model Cards
  - Implement a parser instantiating a model of the Sustainability Model Card

Sustainability Model Cards Dimensions

Sustainability Model Card

Metadata	Training	Inference *(for each task)*	Platform
– Name	– Training Duration	– Inference Type	– Hardware Details
– Version	– Energy Consumption	– Energy Consumption	– Platform Provider
– Type	– Carbon Emissions	– Carbon Emissions	– Platform Region
– Provider	– Water Consumption	– Water Consumption	– Carbon Offset Credit
– License	– *Platform*	– *Platform*	– Energy Sources

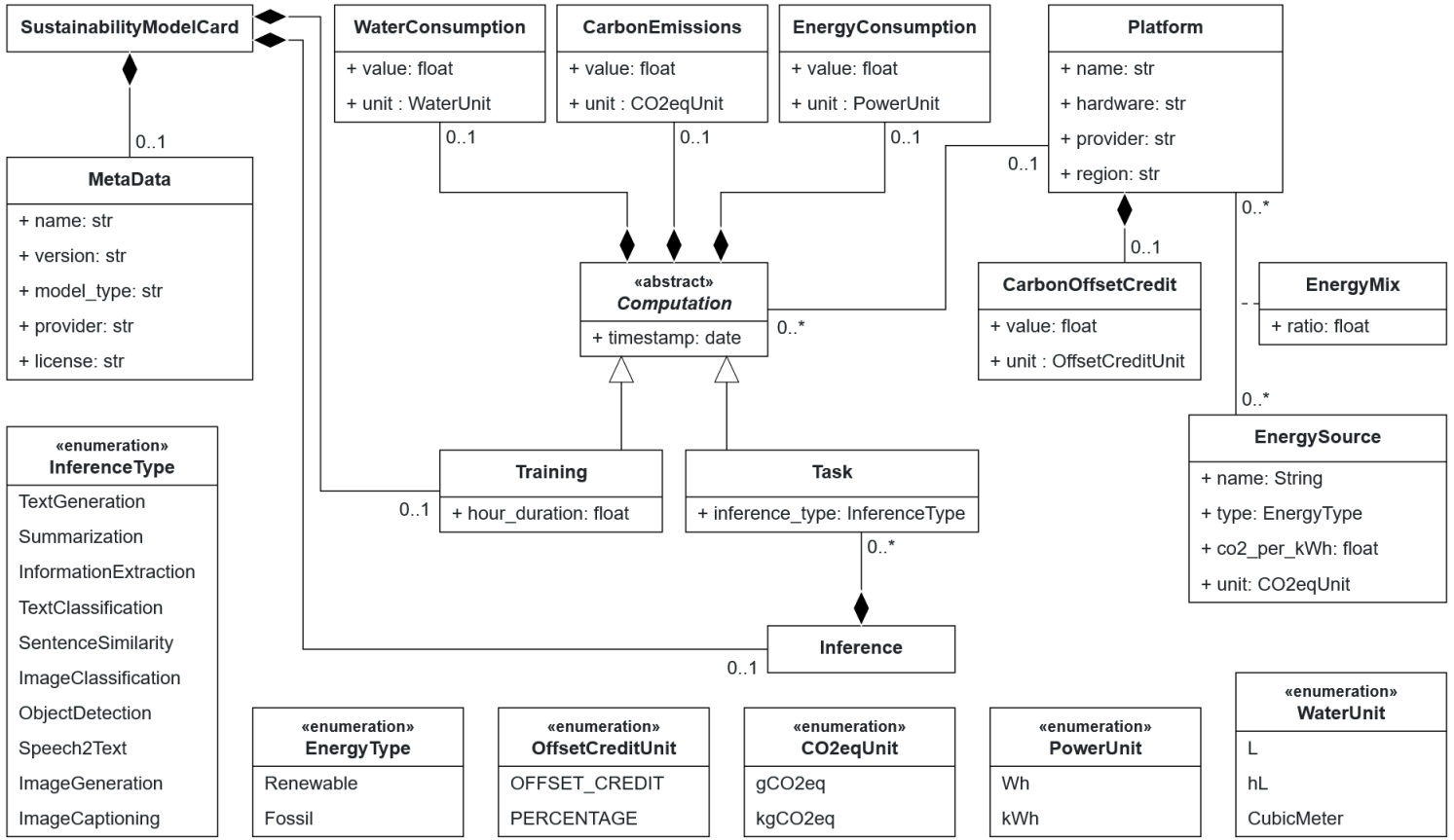
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

A DSL to support Sustainability Model Cards

- What is a Domain-Specific Language ?
  - Software languages specially designed to model systems for a certain domain
- Why use Domain-Specific Languages ?
  - Allows automatic processing of the specified model
- How Domain-Specific Languages are defined ?
  - Abstract syntax (defines concepts of the language and their structure)
  - Concrete syntax (defines the notation of the language, *e.g.* textual, graphical)

Abstraction of the Domain Specific Language (Abstract Syntax)

- Reification of the Quality model
- Abstraction of the computations
- Typing of the metrics (Units)
- Inference type from HuggingFace AI energy score



Notation of the Domain Specific Language (Concrete Syntax)

YAML-based syntax

- To integrate with existing model card DSLs

Writing rules

- Class instances represented as mapping
- Attributes are part of instances mapping
- Compositions defined through nesting
- Multiplicity higher than one as sequences
- Associations are cross-referenced

```yaml
sustainability_model_card:
  meta_data:
    name: GPT-3 175B
    model_type: LLM
    provider: OpenAI
  platforms:
    - platform:
        name: Infrastructure
        hardware: Multiple V100
        provider: Microsoft Azure
        region: US
        carbon_offset_credit:
          value: 100.0
          unit: PERCENTAGE
        energy_mix:
          - energy_mix:
              ratio: 100.0
              energy_source: Azure US
  energy_sources:
    - energy_source:
        name: Azure US
        type: Fossil
        co2_per_kWh: 0.3496
        unit: kgCO2eq
  training:
    platform: Infrastructure
    carbon_emissions:
      value: 449935.2
      unit: kgCO2eq
    energy_consumption:
      value: 1287000
      unit: kWh
    water_consumption:
      value: 5439
      unit: CubicMeter
    timestamp: 2025-01-02T09:00:00
  inference:
    - task:
        inference_type: TextGeneration
        platform: Infrastructure
        carbon_emissions:
          value: 1.398
          unit: gCO2eq
        energy_consumption:
          value: 4
          unit: Wh
        water_consumption:
          value: 0.017
          unit: L
        timestamp: 2025-01-21T09:00:00
```

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

Tool support and DSL implementation

Python-based implementation:

- Metamodel implemented as a set of python class
- Two-step parser
  1. YAML parsing to Python dict/list
  2. Type-checking and model instantiation

The tool is available in open-source on GitHub

Implementation's GitHub repository

Future Work (I)

Extending the coverage and granularity

- More granular description of the training phase
  - Pre-training and fine-tuning part of the training
  - The hyperparameters values
  - Dataset used

Graphical notation

- Form-based
- Conversational-based

Future Work (II)

Tighter integration with Model Cards

- *E.g.,* Hugging Face model cards

Analyzing impact on model users

- User study on how users decide on a model

Application on different scenarios

- Automatic model selection
- Optimize model deployment
- Runtime monitoring to enforce sustainability-aware Service Level Agreements