Module: Machine Translation Quality Assessment
Lecturer: João Lucas Cavalheiro Camargo
Instructions - Lab 4 - Ethics in Evaluation

- This is a summary of a paper about sustainability in NLP.
- Your task is to read the summarised version of the paper and identify which aspects belong to the Triple Bottom Line (TBL): People, Planet and Performance. If you do not identify all aspects in your paper, leave it blank.
- Based on the findings from this paper, provide ideas to how that could be implemented in Machine Translation Quality Assessment.
- And remember: Actively do good! After this class, I suggest you read the full paper and reflect about your practice.

Title: *The Ecological Footprint of Neural Machine Translation Systems*
Reference: Shterionov, D., & Vanmassenhove, E. (2021). The ecological footprint of neural machine translation systems. Tilburg University. Retrieved from https://arxiv.org/pdf/2202.02170.

Framework for Measuring Ecological Impact in Neural Machine Translation (NMT)
This study explores the environmental footprint of neural machine translation (NMT) models, particularly focusing on energy consumption and carbon emissions resulting from model training and translation. The authors analyze two widely used NMT architectures, Long Short-Term Memory (LSTM) and Transformer, examining their performance on different types of GPUs (consumer-grade and workstation-grade) to assess the ecological impact.

The paper highlights that NMT systems, while achieving high performance, are energy-intensive, especially when leveraging GPUs, which draw significantly more power than CPUs. To quantify this impact, the authors conducted experiments across two language pairs (English-French and English-Spanish) and measured energy usage, carbon emissions, and performance across models and GPU types.

The framework emphasizes calculating energy consumption and carbon emissions during both training and inference phases. For more accurate estimations, emissions are based on specific energy intensities of the electricity grids in two locations, Ireland and the Netherlands, allowing a nuanced analysis of the environmental costs based on regional energy profiles.

An important contribution of the framework is its emphasis on model efficiency for sustainability. Results showed that Transformer models, while more power-intensive during translation, are more efficient at training compared to LSTM models, indicating trade-offs based on application needs. The authors also recommend quantization as an optimization technique, which allows models to run on lower-power CPUs without significant loss in translation quality, offering an eco-friendly alternative for inference.

The study proposes future work focused on developing “greener” AI practices within NMT, such as model reusability, selective data usage, and optimizations of hyperparameters to reduce the overall carbon footprint. This framework is part of a growing call within the NLP community to mitigate the environmental impact of machine learning technologies, underscoring the need for sustainable practices in AI research and application.

Module: Machine Translation Quality Assessment
Lecturer: João Lucas Cavalheiro Camargo
Instructions - Lab 4 - Ethics in Evaluation

From this paper, what could fall under the category “People”?

From this paper, what could fall under the category “Planet”?

From this paper, what could fall under the category “Performance”?

How could we adapt this to Machine Translation Quality Assessment?