MSc. Thesis: Overview and Project Description

Sophia Wilson

June 13, 2024

Formalities

General

Student	Sophia Natasha Wilson (KU-ID: ldr934)
Supervisors	Raghavendra Selvan (DIKU) and Jens Hesselbjerg Christensen (NBI)
Time period	August 20th 2024 - May 19th 2025 (approx end date)
Work load	60 ECTS

Specific information for contract

The following agreements involve only Sophia Wilson and Raghavendra Selvan unless otherwise stated.

Frequency of meetings	Weekly, 30-45 minutes if needed. In person if possible. Bi-monthly meetings with both supervisors. Sophia can join weekly group meetings in both research groups.
Expectations for meetings	A designated time/day, with flexibility as needed. Sophia gives status of what she has done since last meeting, asks questions and Sophia and Raghavendra agree on what Sophia will work on until next meeting.
Expectations of your cooperation	Communication through teams channel. Raghavendra has access to an overleaf document where Sophia writes continuously throughout the project.

Project description

Preliminary title

Quantifying the reduction in carbon footprint of physics-informed machine learning in the pursuit of greener artificial intelligence practices.

Description

This thesis aims to investigate the potential reduction in the carbon footprints of machine learning (ML) models by incorporating physical constraints into their design. Traditionally, ML models have relied on vast amounts of data and model parameters, which contribute significantly to computational time and energy consumption, and subsequently to carbon dioxide (CO2) emissions. However, by adopting a physics-informed approach, it may be possible to achieve substantial improvements in carbon footprint during development and deployment.

In this work we will strive to provide concrete evidence of the CO2 emission advantages of physics-informed ML. By comparing the carbon footprint of traditional ML models with those that incorporate physical constraints, the research will seek to quantify the potential reduction in CO2 emissions achievable through this approach. The study will analyse the trade-offs between additional key metrics such as size of data set, model performance, computational efficiency, and environmental impact.

Ultimately, this thesis strives to contribute to the emerging field of sustainable artificial intelligence by providing insight into how integrating physical constraints can lead to more environmentally friendly ML models. The findings of this research have the potential to inform future developments in ML design and promote the adoption of environmentally conscious practices within the ML community.

Datasets and ML models

We will be using benchmark datasets and existing ML models. We have not decided on specific datasets or ML models yet, but the following considerations should be taken into account; accessibility, ability to incorporate physical constraints, potential of CO2-reduction, run time, and relevance.

We expect to use datasets and ML models from the field of climate research due to its high importance and to highlight the self-contradiction; that people believe ML models have the potential to save the climate crises but are not aware of how much CO2 they emit themselves.

Framework

The project will span over a period of nine months. We will divide the period into four phases with different focuses and sub-goals. I will write continuously to avoid a long writing phase in the end. A draft for what the four phases will look like is seen below

1. Study relevant literature to gain knowledge about; how ML is used in climate research, different ways of implementing physical constraints, how to measure CO2 footprint of ML models, etc. In this phase, I will draft a comprehensive literature review, and decide on a dateset and an existing ML model to work with.

2. Implement physical constraints into a ML model. Conduct a comparison of the ML model with and without physical constraints including the CO2 footprint and other relevant metrics. We will be making use of tools like carbontracker.info to measure the carbon/energy footprint of ML models.

3. The focus in the phase is yet to be decided and depends on the outcome of the second phase. However, it will be some form of extended research. Preliminary ideas include; presenting a novel approach of phase 2, explore what type of physics informed ML has the potential of reducing the CO2 footprint the most, and perform a similar CO2 footprint comparison of existing physics informed ML models to obtain more general results.

4. Wrap up.

Learning objectives

- Gain a deep understanding of physics informed ML and ways of measuring the CO2 footprint of ML models by studying relevant literature.
- Write a literature review.
- Implement physical constraints into an existing ML model.
- Quantify the carbon footprint of the model with and without physical constraints.
- Compare additional metrics e.g. size of dataset, model performance, and computational efficiency to obtain a thorough comparison of the ML models.
- Discuss the future and potential of physics informed ML as a sustainable way of doing AI.