Internship Opportunity in Causal ML for Climate Modelling

Deadline to apply: August 16th 2024, 11.59pm AoE
Preferred start date: February 1st 2025 (to be discussed)
Submit your application here

Summary:

Six to eight months, full-time internship at Mila in applied machine learning research focusing on emulating climate models.

Role Responsibilities:

- Collaborate on improving existing causal ML models for climate model emulation, and evaluate their performance on simulated and real data.
- Contribute to code development and deployment.
- Participate in analyzing the models' outputs on various emissions scenarios.
- Help identify new climate science applications for the developed models.
- Help writing scientific content, such as submissions to academic journals or conferences.

Logistics:

- Duration: 6-8 months, paid full-time.
- Start date: Preferred February 1st 2025 (to be discussed)
- Location: Preferred in-person at Mila in Montreal, Canada. Partial virtual participation may be possible.
- Required minimal expertise: Graduate studies in computer science, climate science, or other related disciplines

Candidate Profile:

- Good level of expertise in Python, with experience using Pytorch.
- Good knowledge and understanding of statistics, data analysis and Machine Learning.
- Good knowledge and understanding of causal ML.
- Basic knowledge of climate science.
- Additional assets:
    - Experience with building and training large deep learning models.
    - Experience with building and deploying an end-to-end ML pipeline.
    - Experience using ML models for climate science applications.
    - Experience with version control / git / writing code in a team

Please **go to the form*** to submit your cv and short responses:

- A link to your cv online, or upload pdf.
- Provide your response to the following 4 questions (total maximum 1 page):
    - Can you describe your level of experience in ML / statistics / data analysis? What classes have you taken and what did you learn from them? Do you have any experience in a ML research environment?
    - Can you describe a data analysis / ML project that you conducted? What was the challenge, and how did your solution address it? How did you evaluate the statistics/ML methods that you used? Please focus on your contributions to the project.
    - Can you describe a project/application you've completed using Python? What were the computational challenges and how did you address them? If you have some code from a project that you are able to share, please feel free to include a GitHub link or similar.
    - Can you describe your level of experience in climate science? Did you take any related classes, did you conduct any related project, or did you have any experience in a climate science research environment?

*When submitting your materials online, it won't work from an incognito mode browser.

Thanks! Prospective interviewees will be contacted by the 21st of August the latest. If you do not hear back, thank you for your interest… Unfortunately we are not able to respond to all applicants.

About the project:

Climate models are crucial for simulating future climate scenarios. Physics-based models provide a comprehensive understanding of climate dynamics but are computationally expensive, making it hard to compare and evaluate multiple policy scenarios. ML presents an opportunity to approximate climate projections more efficiently by learning from existing big data archives. However, ML climate emulators often lack interpretability and do not robustly reconstruct causal relationships between climatic conditions and factors such as greenhouse gas emissions, hindering the analysis of various policy-informing scenarios. As causal discovery and causal representation learning has emerged as a valuable tool to improve our understanding of physical systems across various fields, including Earth sciences, our goal is to develop novel causal discovery methods for climate emulation, in order to improve robustness, generalizability and interpretability of efficient ML models. We aim to use these methods to build a novel climate emulator that would thus be more interpretable, better adapt to various unseen emissions scenarios and potentially be used for attribution of extreme climate conditions to human activity, possibly becoming an essential tool for policy makers and advocates for climate justice.