Salva Rühling Cachay

salvaruehling@gmail.com | salvarc.github.io | Google Scholar | LinkedIn | GitHub

I aim to develop and use machine learning (ML) methods for positive real-world impact in areas like *weather forecasting, climate modeling, and sustainability.* On the ML side, I am particularly interested in *self-supervised learning, high-dimensional forecasting, and generative modeling.*

Education

University of California, San Diego — La Jolla, USA
PhD in Computer Science; Advisor: Prof. Rose Yu — *Sep. 2022 - present*
Selected coursework: *Recommender Systems; Data Systems for ML; Deep Generative Models; Unsupervised Learning*

Technical University of Darmstadt — Darmstadt, Germany
B.Sc. in Computer Science; **With Honors** (GPA = 1.24/1.0, lower is better) — *Sep. 2018 – May 2022*

Peer-Reviewed Publications (Conferences and Journals)

S. Rühling Cachay, B. Zhao, H. Joren, R. Yu. "DYffusion: A Dynamics-informed Diffusion Model for Spatiotemporal Forecasting". *NeurIPS 2023, [Blog post]*

S. Rühling Cachay*, V. Ramesh*, J. Cole, H. Barker, D. Rolnick. "ClimART: A Benchmark Dataset for Emulating Atmospheric Radiative Transfer in Weather and Climate Models". *NeurIPS Track on Datasets, 2021*

S. Rühling Cachay, B. Boecking, A. Dubrawski. "End-to-End Weak Supervision". *NeurIPS, 2021*

Pre-prints and Workshop Papers (Selected)

S. Rühling Cachay, A. Fender Bucker, W. Potosnak, E. Pokropek, E. Erickson, S. Bire, S. Osei, B. Lütjens. "The World as a Graph: Improved El Niño Forecasting with Graph Neural Networks". *preprint*

S. Rühling Cachay*, V. Ramesh*, J. Cole, H. Barker, D. Rolnick. "ClimART: A Benchmark Dataset for Emulating Atmospheric Radiative Transfer in Weather and Climate Models". *NeurIPS Tackling Climate Change with Machine Learning, 2021* (Spotlight), *and Helmholtz-Zentrum Hereon, Data Science Symposium* (Spotlight)

Research Experience

Allen Institute for AI (AI2), *Climate Modeling Research Intern* — Summer 2023
- Achieved competitive or better stability, probabilistic skill in terms of CRPS and forecast reliability, and realistic weather variability than relevant baselines for data-driven climate simulation of up to 10 years.

UC San Diego, *Research Assistant* — since Fall 2022
- Working on AI for Science, Generative Modeling, and Probabilistic spatiotemporal forecasting
- Proposed a novel dynamics-informed diffusion model for probabilistic spatiotemporal forecasting (NeurIPS 2023).

Palo Alto Research Center (PARC), *Research Intern and Visiting Researcher* — Summer 2022
- Worked on the AIBEDO project with Dr. Kalai Ramea at the intersection of climate modeling and ML.
- Applied a Fourier Neural Operator (FNO)-based neural architecture to successfully emulate climate variability as a response to cloud property forcings.

Mila - Quebec AI Institute, *Research Intern* — March 2021 – June 2022
- Worked with Prof. David Rolnick on improving and speeding-up climate models via ML parameterizations. Joint work with Environment and Climate Change Canada.
- Created ClimART: A large-scale benchmark dataset for emulating physics models of atmospheric radiation, and proposed new models such as graph networks that outperform prior baselines (NeurIPS 2021).
- Stay was extended to write my bachelor thesis at Mila.

Carnegie Mellon University, *Research Intern* — June 2020 – March 2021
- Worked at the Auton Lab — initially started as a Robotics Institute Summer Scholar (RISS).
- Researched the effect of modeling and misspecifying dependencies in weak supervision.
- Developed WeaSEL: A novel, neural core framework for multi-source weak supervision (NeurIPS 2021).
- Open-sourced a Pytorch Lightning+Hydra-based framework (> 100 GitHub stars).

Technical University of Darmstadt, *Undergraduate Researcher* Apr. 2020 – June 2020

- Worked with Prof. Gurevych at the UKP lab on NLP for the case law of the European Court of Human Rights.
- Scraped, parsed and structured as XML files the whole court's database (>160k case law documents).
- Built ML algorithms (Transformers and a SVM) to predict the judgement given the facts section.

Projects

Graph Neural Networks (GNN) for Improved El Niño Forecasts Sep. 2020 – March 2021

- Competed with the international, diverse team I formed at ProjectX, a ML for climate change research competition hosted by University of Toronto AI.
- *Led the research agenda* and the effort to, successfully, receive a Microsoft AI for Earth grant *(Showcased profile)* .
- Developed a GNN to better forecast El Niño/ENSO, with enhanced interpretability.
- Our model outperforms state-of-the-art methods for up to six months forecasts & learns meaningful patterns.

Skills

Programming Languages: Python, Java *(proficient)*, MATLAB, C, C++, CUDA *(familiar)*
Languages: Spanish and German *(native)*, English *(fluent*, TOEFL iBT: 112/120), French *(advanced)*, Portuguese *(beginner)*
Libraries & Tools: PyTorch (+Lightning), NumPy, Numba, Xarray, Hydra, Git, Github Actions, AWS, Azure

Professional Service & Volunteering

Reviewing at various conferences, *Reviewer*

- International Conference on Machine Learning (ICML); 2024
- International Conference on Learning Representations (ICLR); 2024
- Advances in Neural Information Processing Systems (NeurIPS); 2023
- Fragile Earth: AI for Climate Mitigation, Adaptation, and Environmental Justice workshop at ACM KDD; 2022

16th Graduate Climate Conference, *Workshop Organizer* Oct. 2022

- Organized an ML for climate workshop (as one of 6, out of 30, proposals). Notebook tutorial link.

Jacobs Undergraduate Mentoring Program (JUMP), *Graduate Mentor* since Oct. 2022

TU Darmstadt, *Teaching Assistant in Maths I for CS (linear algebra and discrete maths)* 2019 – 2020

Vrindhavan Kindergarten, *International Youth Volunteer* Aug. 2017 – Aug. 2018

- Worked mainly with 4 - 6-year-olds, e.g. in sports, language & crafts activities.

Awards & Honors

NeurIPS Scholar Award 2023

Jane Street Graduate Research Fellowship, Honorable Mention – One of 39 (> 600 applicants) 2023

Jacobs School of Engineering Fellowship – Awarded to 5 students in my department 2022

Sponsored NASA Summer School on Satellites & Climate Models – One of 22 participants (> 175 applicants) 2022

Microsoft AI for Earth Grantee – Project leader *(Showcased profile and interview)*. 2020

DAAD RISE scholarship – cancelled due to Covid-19 2020

Germany Scholarship – awarded to 1% of students in Germany 2019 & 2020

Invited Talks

Tübingen University, ML in Climate Science group – DYffusion: A Dynamics-informed Diffusion Model Jan. 24

Zalando GNN reading group – GNNs for Long-Range Forecasting Aug. 22

ICAI congress of IEEE UPC, Lima, Peru – Climate Change and Machine Learning: An Overview Jul. 22

NEC Labs Europe – Climate Change and Machine Learning: An Overview Apr. 22

UC Berkeley AI+Climate Change reading group – ClimART benchmark dataset Jan. 22

McGill University, RLL Lab – ClimART benchmark dataset Nov. 21

NEC Labs Europe – End-to-End Weak Supervision Nov. 21

IBM Research, Future of Climate Group – GNNs for Long-Range Forecasting Aug. 21

Imperial College London, Data Science Institute – GNNs for Long-Range Forecasting *(video)* Mar. 21

Presentations

16th Graduate Climate Conference, Pack Forest, WA – Emulating Atmospheric Radiative Transfer with ML *(oral)*	Oct. 22
Helmholtz-Zentrum Hereon, Data Science Symposium – ClimART benchmark dataset *(contributed talk)*	Jun. 22
NeurIPS Climate Change+ML – ClimART benchmark dataset *(spotlight) (video)*	Dec. 21
ICLR WeaSuL – Dependency Structure Misspecification in Multi-Source Weak Supervision Models *(contributed talk) (video)*	Apr. 21
NeurIPS LatinX in AI Workshop – Model Misspecification in Multiple Weak Supervision *(oral) (video)*	Dec. 20

Pre-prints and Workshop Papers (Complete)

S. Rühling Cachay, A. Fender Bucker, W. Potosnak, E. Pokropek, E. Erickson, S. Bire, S. Osei, B. Lütjens. "The World as a Graph: Improved El Niño Forecasting with Graph Neural Networks". *preprint*

S. Rühling Cachay, Peetak Mitra, Haruki Hirasawa, Sookyung Kim, Subhashis Hazarika, Dipti Hingmire, Phil Rasch, Hansi Singh, Kalai Ramea. "ClimFormer – a spherical Transformer model for long-term climate projections". *NeurIPS Machine Learning and the Physical Sciences workshop, 2022*

S. Rühling Cachay*, V. Ramesh*, J. Cole, H. Barker, D. Rolnick. "ClimART: A Benchmark Dataset for Emulating Atmospheric Radiative Transfer in Weather and Climate Models". *NeurIPS Tackling Climate Change with Machine Learning, 2021* (Spotlight)*, and Helmholtz-Zentrum Hereon, Data Science Symposium* (Contributed talk)

S. Rühling Cachay, B. Boecking, A. Dubrawski. "Dependency Structure Misspecification in Multi-Source Weak Supervision Models". *ICLR Workshop on Weakly Supervised Learning, 2021 (*Contributed talk*)*

S. Rühling Cachay, A. Fender Bucker, W. Potosnak, E. Pokropek, E. Erickson, S. Osei, B. Lütjens. "Graph Deep Learning for Long-Range Forecasting". *European Geosciences Union (EGU) General Assembly, 2021*

S. Rühling Cachay, A. Fender Bucker, W. Potosnak, E. Pokropek, E. Erickson, S. Osei, B. Lütjens. " Graph Neural Networks for Improved El Niño Forecasting". *NeurIPS Tackling Climate Change with Machine Learning, 2020*

S. Rühling Cachay, B. Boecking, A. Dubrawski. "Model Misspecification in Multiple Weak Supervision". *NeurIPS LatinX in AI workshop, 2020 (*Oral*)*