Vaishnavi Kannan

vk2580@columbia.edu (929)4938653 linkedin.com/in/vaishnavikannan42

github.com/vaishnavikannan42

SUMMARY

MS Data Science candidate at Columbia (Dec 2026) focused on climate and weather ML, with one thread across all of it: turning climate data into models and tools people can actually trust and use. On the modeling side, subseasonal-to-seasonal monsoon forecasting, probing NASA's Prithvi-WxC weather foundation model (NeurIPS 2026 workshop paper), and building an ML emulator with a NASA GISS group. On the data side, getting messy climate data into usable, trustworthy form, currently validating an AI-rebuilt FEMA climate-risk model with a climate-data startup. In short: builds the models, and does the data work that makes them trustworthy.

EDUCATION

Columbia University, *MS in Data Science* Aug 2025 – Present | New York, USA

- **Relevant Coursework**:Machine Learning, Computer Vision, Probability & Statistics, Linear Algebra, Algorithms, statistical inference and modeling

Kumaraguru College of Technology, *BTech in Information Technology* Jun 2020 – May 2024 | Coimbatore, India

TECHNICAL SKILLS

- **Weather-Climate ML:** weather-climate foundation models (NASA Prithvi-WxC), latent extraction & adaptation, subseasonal-to-seasonal (S2S) forecasting, teleconnection-conditioned modeling (MJO, MISO, ENSO, IOD), U-Net / CNN, FiLM conditioning, ML emulation
- **Climate Data:** xarray, netCDF, reanalysis & gridded climate data, regridding, geospatial pipelines, data validation & lineage, reproducible research pipelines
- **Evaluation:** probabilistic forecast skill (RPSS), held-out (leave-one-year-out) validation, ablation studies, calibration, data-leakage detection
- **Programming & Infra:** Python (NumPy, pandas), PyTorch, Transformers, SQL, Git; FastAPI, Docker, AWS

RESEARCH

Research Assistant — S2S Monsoon Forecasting, *Advised by Prof. Andrew Robertson (Senior Research Scientist, IRI for Climate and Society)* Jan 2026 – Present | New York,USA

- Built FiLM-conditioned U-Net models for Week 3–4 Indian monsoon rainfall forecasting, conditioned on MJO, MISO, ENSO, and IOD; across an ablation of 11+ variants, every model beat the extended-logistic-regression baseline.
- Discovered a data-alignment leak (conditioning indices matched to the target window, not the forecast issue date), built a per-sample diagnostic exposing it (~47% of apparent skill; RPSS 0.047 → 0.025), and identified an operationally valid fix using issue-time predicted indices.
- Replaced bootstrapping with deterministic leave-3-years-out cross-validation and evaluated skill per grid cell with spatial skill maps, documenting negative results without overclaiming.

Research Assistant — Foundation-Model Latents (Prithvi-WxC) Jan 2026 – Present | New York,USA
Columbia Climate School

Built a latent-extraction workflow for NASA's Prithvi-WxC weather-climate foundation model, pulling daily encoder/decoder representations over South Asia and the Indian Ocean, and trained LSTM/Transformer models to reconstruct subseasonal climate-variability indices (MISO, MJO, IOD) under leave-one-year-out evaluation, accepted at the TCCML workshop, NeurIPS 2026.

Climate-Risk Model Reconstruction, *with Fulton Ring* New York, USA

Making preserved public climate-risk data usable and trustworthy: validating an agent-built reconstruction of FEMA's National Risk Index Future Risk model (recomputed with newer data) against FEMA's methodology, checking data lineage and spot-checking county-level results, and co-authoring a case study on how projected risk shifts.

Research Assistant - ML Climate-Model Emulator New York, USA
with Prof. Anastasia Romanou's group (NASA Goddard Institute for Space Studies)

Developing ML emulators for components of the GISS climate model to stand in for expensive physics, accelerating simulation while preserving fidelity, and preparing the training and evaluation data the emulator learns from.

PROFESSIONAL EXPERIENCE

Holland America Line, *Data Science Intern — Onboard Revenue* Jun 2026 – Aug 2026 | Miami, Florida

- Built and shipped a voyage-level forecasting system (Python, Snowflake) at 3–4% WAPE against closed actuals, replacing an incumbent formula with an explainable glass-box model stakeholders could interrogate.
- Engineered a day-proration allocation layer to disaggregate coarse general-ledger data to the grain forecasting required, the pipeline step that made voyage-level modeling possible.

People+ai (part time), *AI/ML Engineer* Aug 2024 – Mar 2025 | Bangalore, India

- Built NLP extraction and knowledge-graph pipelines over regulatory documents and a RAG chatbot returning cited answers, structuring messy, unstructured data for retrieval.