Arpit Kapoor

PhD Candidate | Machine Learning | Hydrology
Phone: +61 412 982 005 | Email: kapoor.arpit97@gmail.com
www.arpit-kapoor.com | GitHub (arpit-kapoor) | Google Scholar | LinkedIn

Research Summary

PhD candidate in Mathematics & Statistics at UNSW, specialising in AI-driven environmental process modelling. My research develops hybrid and Bayesian deep learning approaches that integrate physical models with data-driven methods for complex climate and hydrological systems. With a strong record in top-tier ML and environmental modelling journals, I aim to extend my work through cross-disciplinary collaborations with leading researchers in AI and environmental science. My goal is to co-design innovative, interpretable AI methods that address pressing environmental challenges while advancing the state of the art in scientific machine learning.

Education

Doctor of Philosophy (PhD), Mathematics and Statistics Aug 2022 – Feb 2026 (Expected)
University of New South Wales, Sydney
Thesis: *Enhancing hydrological process modelling with scientific deep learning*
Grants: *UNSW University International Postgraduate Award (UIPA)*

Bachelor of Technology (B.Tech) Jul 2015 – May 2019
SRM Institute of Science and Technology, India
Major: *Computer Science & Engineering* **(GPA: 9.01/10.0)**
Final Project: *Deep reinforcement learning methods for locomotion of humanoid robots*

Professional Experience

Bureau of Meteorology, Australia Feb 2023 – Apr 2025, Part-Time
Research Support Scientist

- Implement a multivariate bias correction method (MRNBC) for climate projections to be used by the **Australian Climate Service**.
- Engineered a scalable Python interface to modernise legacy FORTRAN-based software.
- Integrated a **Dask-based** distributed computing approach for improved efficiency.
- Reduced simulation runtime by **80%** on the National Computing Infrastructure (NCI)
- Co-authored an **abstract at EGU 2025** on the analysis of climate data bias-adjustment techniques

Quince, India Mar 2022 – Aug 2022, Full-Time
Data Scientist

- Applied self-supervised learning techniques to model and predict customer churn, which informed strategies leading to a **15% lift in user engagement**.

- Developed a gradient boosting-based forecasting model to **predict and optimize logistics costs**, identifying potential shipping cost reductions of 10%
- Collaborated with operations teams to ensure the successful alignment and integration of model outputs with business processes

3Qi Labs, India Nov 2019 – Nov 2021, Full-Time
Data Scientist

- Designed and deployed an LSTM autoencoder for **anomaly detection** in data pipelines, boosting detection accuracy by 40%.
- Integrated automated ML pipelines for data quality monitoring, saving 50+ analyst hours/month.
- Automated ML workflows on the DataBricks platform using PySpark and related technologies.
- Introduced and deployed **MLOps infrastructure (MLflow & Docker),** improving model lifecycle management.

Bomotix, India Jan 2019 – Nov 2019, Full-Time
Machine Learning Engineer

- Developed a real-time computer vision pipeline for efficient tracking of player performance in sports
- Implemented deep learning based solutions for various computer vision problems like player **detection and tracking** (using DeepSORT), **3D player pose estimation** (using DeepHAR)
- Improved model efficiency using **mixed-precision and GPU optimisations**, leading to a 15% reduction in model training costs
- Cleaned, processed, and curated video data for model training and validation
- Maintained **CI/CD pipelines** for deep learning model development and deployment

The University of Sydney Jan 2018 – Sep 2018, Part-Time
Machine Learning Research Intern

- Research on **Bayesian methods** for uncertainty quantification in neural networks
- Developed Python software for a parallel **Markov chain Monte-Carlo (MCMC)** based approach for **Bayesian transfer learning in neural networks**
- Contributed to research on the Bayesian inversion problem for geoscientific models using **parallel-tempering MCMC methods**
- Co-authored publications on Bayesian machine learning at top-tier ML journals

---

Publications

2023-2025

Kapoor, A., Chandra, R., QDeepGR4J: A quantile-based ensemble of hybrid deep learning rainfall-runoff models catering to extreme values and uncertainties, Journal of Hydrology **(under review)**

Takbash, A., Irving, D., Peter, J., Dao, T. L., **Kapoor, A.,** Gammon, A., Dowdy, A., Black, M., Bende-Michl, U., Jakob, D., and Grose, M.: A Comprehensive Assessment of Climate Data Bias-Adjustment Techniques Over Australia, EGU General Assembly 2025, Vienna, Austria, 27 Apr–2 May 2025, EGU25-252, https://doi.org/10.5194/egusphere-egu25-252, 2025

Chandra, R., **Kapoor, A.**, Khedkar, S., Ng, J., & Vervoort, R. W. (2024). Ensemble quantile-based deep learning framework for streamflow and flood prediction in Australian catchments. *arXiv preprint arXiv:2407.15882*.

Kapoor, A., Pathiraja, S., Marshall, L., & Chandra, R. (2023). DeepGR4J: A deep learning hybridization approach for conceptual rainfall-runoff modelling. *Environmental Modelling & Software*, 169, 105831.

Kapoor, A., Negi, A., Marshall, L., & Chandra, R. (2023). Cyclone trajectory and intensity prediction with uncertainty quantification using variational recurrent neural networks. *Environmental Modelling & Software*, 162, 105654.

2020 - 2022

Kapoor, A., Nukala, E., & Chandra, R. (2022). Bayesian neuroevolution using distributed swarm optimization and tempered MCMC. *Applied Soft Computing*, 129, 109528.

Chandra, R., Azam, D., **Kapoor, A.**, & Müller, R. D. (2020). Surrogate-assisted Bayesian inversion for landscape and basin evolution models. *Geoscientific Model Development*, 13(7), 2959-2979.

Chandra, R., Jain, K., **Kapoor, A.**, & Aman, A. (2020). Surrogate-assisted parallel tempering for Bayesian neural learning. *Engineering Applications of Artificial Intelligence*, 94, 103700.

Chandra, R., & **Kapoor, A.** (2020). Bayesian neural multi-source transfer learning. *Neurocomputing*, 378, 54-64.

2017 - 2019

Sripada, A., Asokan, H., Warrier, A., **Kapoor, A.**, Gaur, H., Patel, R., & Sridhar, R. (2018). Teleoperation of a humanoid robot with motion imitation and legged locomotion. In *2018 3rd International Conference on Advanced Robotics and Mechatronics (ICARM)* (pp. 375-379). IEEE.

Sripada, A., Warrier, A., **Kapoor, A.**, Gaur, H., & Hemalatha, B. (2017). Dynamic lateral balance of humanoid robots on unstable surfaces. In *2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT)* (pp. 1-6). IEEE.

Grants & Awards

- PhD Scholarship, Australian Research Council Training Centre in Data Analytics for Resources and Environment (DARE) Centre
- PhD Scholarship, UNSW University International Postgraduate Award (UIPA)
- Finalist at the IEEE/RSJ IROS 2017 Humanoid Application Challenge (Canada)
- Secured 1 Gold, 2 Silver, 1 Bronze medal at Robogames 2017 (USA)

Academic Activities

- Selected Participant, Information Resilience PhD School, University of Queensland (2024)
- Challenge Facilitator, Data Study Group, Alan Turing Institute (UK, 2024)
- **Co-organised and hosted** the UNSW Transitional AI research group **seminar series** (2023-2024)
- **Tutored for postgraduate machine learning courses** at UNSW (MATH3856, MATH5836 and COMP9417)
- Affiliate student researcher at the UNSW Data Science Hub (now part of **UNSW AI Institute**)
- Team Leader - SRM Team Humanoid, student-led humanoid robotics research (2018-2019)

Technical Skills

- **Programming & Scientific Computing:** Python (NumPy, pandas, scikit-learn, PyTorch, JAX/Flax), C/C++, SQL; distributed computing with Dask, PySpark, MPI on HPC systems (NCI, Katana).
- **Machine Learning & AI:** Deep learning (CNNs, LSTMs, transformers, autoencoders), uncertainty quantification (Bayesian deep learning, variational inference, MCMC, ensembles), anomaly detection, generative models, reinforcement learning.

- **Scientific Machine Learning:** Hybrid modelling (physics–ML integration), operator learning (FNO, GINO), surrogate modelling
- **Domain Applications:** Hydrological and climate modelling (rainfall–runoff, flood forecasting, cyclone prediction), geoscientific inversion and parameter estimation, environmental process emulation.
- **High-Performance & Cloud Computing:** Parallel/distributed training pipelines, containerisation (Docker, Singularity), workflow orchestration (Airflow), cloud platforms (AWS, GCP, Azure).
- **Data Management & Visualisation:** Large-scale environmental datasets (NetCDF, xarray), Snowflake, Elasticsearch; data cleaning/validation; visualisation with Matplotlib, Plotly, and scientific dashboards.