

Arpit Kapoor

PhD Candidate | Data Scientist

Phone: +61 412 982 005 | Email: kapoor.arpit97@gmail.com

www.arpit-kapoor.com | [GitHub \(arpit-kapoor\)](https://github.com/arpit-kapoor) | [Google Scholar](https://scholar.google.com/citations?user=...) | [LinkedIn](https://www.linkedin.com/in/arpitkapoor)

Professional Summary

Data Scientist (5+ yrs) and PhD candidate translating research into production systems. Strong command over machine learning tools for computer vision, language modelling, time-series, uncertainty-aware forecasting, large-scale data processing and deployment over cloud. Demonstrated record of improving customer retention (~15%), and partnering with cross-functional teams to land models in products.

Technical Skills

- **Programming & Data:** Python, SQL, PySpark, Dask, Databricks
- **Machine Learning & AI:** scikit-learn, PyTorch, TensorFlow, XGBoost, CatBoost; computer vision (CNNs, AutoEncoders, Diffusion models); NLP (transformers, text classification, topic modelling, embeddings, summarisation); time-series forecasting (ARIMA, Prophet, quantile models)
- **Generative AI & LLMs:** LLM fine-tuning, retrieval-augmented generation (RAG), reinforcement learning with human feedback (RLHF), embeddings & semantic search, and prompt-engineering
- **Statistics & Causal Inference:** Hypothesis testing, experiment design, A/B testing, uplift modelling, causal inference, uncertainty quantification
- **Cloud & MLOps:** AWS (SageMaker, Bedrock, Lambda), GCP (Vertex AI, BigQuery), Azure (AI Studio); MLflow, Docker, Git, CI/CD (GitHub Actions), Airflow, Databricks Jobs, TorchServe, Kubernetes
- **Data Engineering & Visualisation:** Data warehousing (Snowflake, Elasticsearch), large-scale data pipelines, data validation (Great Expectations), dashboards with Matplotlib, Plotly, Tableau, Power BI

Professional Experience

Bureau of Meteorology, Australia

Feb 2023 – Apr 2025, Part-Time

Research Support Scientist

- Implement multivariate bias correction for climate models for the **Australian Climate Service**.
- Developed a scalable **Python interface** for legacy FORTRAN-based models
- Integrated a **Dask-based** distributed computing approach for improved efficiency.
- Reduced simulation runtime by **80%** on the National Computing Infrastructure (NCI)

Quince, India

Mar 2022 – Aug 2022, Full-Time

Data Scientist

- Developed churn propensity models that **lifted repeat user engagement by ~15%**; partnered with CRM to target
- Led the development of a logistics cost-forecasting model using gradient boosting with readily available and engineered features
- Contributed an estimate of **10% shipping cost reduction**.
- Drove stakeholder alignment: defined success metrics, dashboards, and experiment readouts to translate model performance into business outcomes.

3Qi Labs, India

Nov 2019 – Nov 2021, Full-Time

Data Scientist

- Designed and deployed ML-based **anomaly detection** for 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)** for better 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
- Developed, trained and maintained various deep learning based computer vision models.
- 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

Education

Doctor of Philosophy (PhD), Mathematics and Statistics

Aug 2022 – Feb 2026 (Expected)

University of New South Wales, Sydney

Thesis: **Enhancing process-based hydrological modelling with deep learning**

Bachelor of Technology (B.Tech)

Jul 2015 – May 2019

SRM Institute of Science and Technology, India

Major in **Computer Science & Engineering**

Extra Curricular: Team Leader of SRM Team Humanoid (student-led humanoid robotics team)

Awards & Achievements

- Selected Participant, **Information Resilience PhD School**, University of Queensland, Brisbane (2024)
 - **Challenge Facilitator**, Data Study Group - Theyr Challenge, **Alan Turing Institute** (UK, 2024)
 - PhD Scholar, Australian Research Council Training Centre in Data Analytics for Resources and Environment (DARE) Centre & UNSW **University International Postgraduate Award (UIPA) Grant**
 - Finalist, IEEE/RSJ IROS 2017 Humanoid Application Challenge (Canada)
 - **Team Lead, SRM Team Humanoid** (Robogames 2017 – 1 Gold, 2 Silver, 1 Bronze)
 - Published in several tier one machine learning and engineering journals - more details on personal website and Google Scholar profile
-