Robert Adriel Mostoghiu Paun

Data Scientist

+61 422 847 973 | robertadriel@gmail.com | linkedin.com/in/robert-mostoghiu-paun/ | github.com/robmost

Executive Summary

I build large-scale predictive models and automated ETL pipelines, with 8+ years of experience working across terabyte-scale datasets in distributed HPC environments. Currently at the ARC Dark Matter Centre of Excellence, I lead international projects optimising high-fidelity simulations, achieving a **40% accuracy increase** in HPC environments and a **98% reduction in engineering cycle times** through AI orchestration. My core stack is Python and C/C++, with hands-on experience in scikit-learn and PyTorch across statistical inference and large-scale numerical analysis.

Work Experience

Postdoctoral Research Associate February 2022 – Present

ARC Centre for Dark Matter and Particle Physics, Swinburne University *Melbourne, VIC, Australia*

- **Model Development & Validation**: Engineered and integrated a novel computational module (GIZMO-PBHEF) into a legacy codebase to model complex system dynamics in distributed High Performance Computing (HPC) environments. Developed and validated novel simulation techniques that achieved a **40% improvement in model accuracy** by identifying and correcting systemic errors within terabyte-scale predictive models.
- **Data Pipeline Engineering**: Architected **end-to-end Extract, Transform, Load (ETL) workflows** in Python and SQL to transform terabyte-scale raw simulation data into structured, high-fidelity formats for distributed HPC processing.
- **Anomaly Detection**: Applied A/B testing, feature engineering, statistical thresholds, and population-level diagnostics to identify anomalies and **distinguish genuine signals from computational noise**.
- **Applied Machine Learning**: Followed up the anomaly work with "Spurious Halo Classifier", a project replacing a hand-tuned threshold with scikit-learn and PyTorch models trained on independently derived labels, structured on a **medallion data architecture** with MLflow experiment tracking, model interpretability (SHAP) and Continuous Integration (CI).
- **AI Orchestration**: Built the "SAGE Tree Converter", an LLM-powered CLI framework using an adapter design pattern and persistent knowledge base to automate complex hierarchical data conversion, cutting a 1-3 month manual engineering cycle to a single day (**98% reduction in cycle time**).
- **Platform Administration**: Built and administered a custom collaboration platform for **a 100-member international research consortium** (The Three Hundred), replacing a commercial Slack setup and **avoiding ongoing licensing costs**.
- **Governance & Operations**: Appointed to the Swinburne Time Allocation Committee (STACK) to **assess and rank competing project proposals** against technical and scientific merit criteria to allocate access to scarce, high-value observatory time across an international research consortium. Managed **access and compliance for cross-institutional research partnerships**, acting as the primary point of contact between visiting researchers and internal technical teams to ensure observatory access policies are met.
- **Stakeholder Engagement & Leadership**: Supervised **3 undergraduate** and **2 postgraduate** researchers. Delivered weekly progress reports on technical challenges and outputs to a cross-disciplinary team spanning astrophysics, experimental physics, and engineering.

PhD Researcher September 2016 – November 2020

Universidad Autónoma de Madrid *Madrid, Spain*

- **Data Integration & Standardisation**: Built Python workflows (NumPy/Pandas) for data ingestion and R notebooks for statistical analysis, standardising **terabyte-scale** datasets across multiple international HPC facilities, enabling large-scale comparative analysis for "The Three Hundred" consortium.
- **Statistical Benchmarking**: Ran rigorous statistical benchmarking of multiple modelling architectures to isolate numerical artefacts from physical signals, establishing **reproducible evaluation criteria** across model comparisons.
- **Technical Communication**: Translated complex model outputs for both technical and non-technical audiences across **6 international conference**s and **9 peer-reviewed publications**.

Education

Universidad Autonóma de Madrid Madrid, Spain
Doctor of Philosophy in Computational Astrophysics *September 2016 – November 2020*

- **Awards:** Graduated with Summa Cum Laude (2020). Awarded a Doctorate Honourable Mention (2022).

Universidad Autónoma de Madrid Madrid, Spain
Master of Science in Cosmology and Particle Physics *September 2015 – June 2016*

- **Awards:** General enrolment funding and academic stipend by the Spanish Ministry of Education

Universidad Autónoma de Madrid Madrid, Spain
Bachelor's Degree in Physics *September 2011 – June 2015*

- **Awards:** General enrolment funding and academic stipend by the Spanish Ministry of Education

Professional Development

Commercialisation Training Program for Medical Researchers Melbourne, Australia
AAMRI VIC *September 2025 – December 2025*

- Selected for a 20-person cohort to develop commercialisation pathways for cross-sector digital health projects. Completed the program with an investor pitch night, presenting technical research as a commercial proposition to an industry panel.

Research Translation & Design Thinking Geneva, Switzerland
IdeaSquare, CERN *July 2025*

- Completed an intensive program in research translation and design thinking, covering commercial literacy, visual storytelling, and communicating technical work to non-specialist audiences.

Skills

Programming & Databases: Python (NumPy, SciPy, Pandas, Polars, Seaborn, Matplotlib), Jupyter, SQL (PostgreSQL, mySQL), R, C/C++.
Data Science & Modelling: Numerical & Predictive Modelling, Machine Learning (scikit-learn, PyTorch, joblib), Generative AI, AI Orchestration (CLIs), Bayesian Inference, CI/CD & MLOps (MLflow), Large-Scale Data Analysis, Data Visualisation
Developer Tools: High-Performance Computing (HPC), Cloud Services (Azure Databricks), Dashboards (Power BI, Excel), Git, Docker, Linux Shell Script, SLURM, LaTeX
Languages: English (fluent), Spanish (native), Romanian (native)