XIANGXIN (SEAN) ZHAO

Email: xiangxinzhao801@gmail.com | LinkedIn: linkedin.com/in/seanzhao801

EDUCATION

IMPERIAL COLLEGE LONDON — **LONDON, UK**

MSc Environmental Data Science and Machine Learning — *Sep 2024 – Sep 2025*

- **Core Modules:** Data Science and Machine Learning, Deep Learning, Big Data Analytics, Numerical Programming in Python, Advanced Programming, Inversion and Optimization, Computational Mathematics.

UNIVERSITY COLLEGE LONDON — **LONDON, UK**

BSc Mathematics (2:1 Honors) — *Sep 2021 – Jun 2024*

- **Core Modules:** Combinatorial Optimization, Graph Theory, Game Theory, Real & Complex Analysis, Decision & Risk

Work EXPERIENCE

AI Product Manager Intern — **BEIJING, CHINA**

MetaDream (Bondee) — *Jun 2024 – Sep 2024*

- Led a cross-functional team (engineers from Stanford, UIUC and NYU) to develop an AI-driven text editor with 10+ generative AI features across web and mobile platforms, including text generation from pictures, AI archive, etc.
- Enabled real-time multimedia synchronization and coherent context generation using advanced LLM pipelines and agentic workflow.
- Streamlined Agile workflows with 1-week Scrum sprints, enhancing feature delivery efficiency.

Founder & Product Manager — **GUANGZHOU, CHINA**

YoDreams Tech — *May 2023 – Aug 2024*

- Designed an AI agent and a Transformer-based admission prediction model trained on 30,000+ academic profiles; automated school selection strategy, reducing manual workload by 70%.
- Built "AI Novelist", an interactive novel generator web app, obtained funding from Amazon China (AI Supernova) and ZhenFund.
- Developed "AI Fortune Teller", an LLM-based lifestyle app, now in pre-launch testing with confirmed first-round backing.

RESEARCH PROJECTS

Copula-VAE Hybrid Synthesis for Small-Sample Geoscience Data - MSc Graduation Project

Imperial College London — *May 2025 – Sep 2025*

- Developed a hybrid copula–variational autoencoder framework enabling transferable data synthesis for small-sample rock-blasting datasets, addressing data scarcity in geoscience modelling.
- Consolidated eight heterogeneous datasets into a unified 12-feature schema (n=262) with zero missingness on core variables, then assessed cross-domain generalization with train-only preprocessing, a fixed XGBoost, and multi-seed aggregation.
- Introduced a two-stage statistical–deep learning pipeline: copula-based dependency modelling (PIT + Kendall τ with shrinkage, AIC-per-observation-guided family selection) followed by TVAE-GMM synthesis with β-annealing for posterior stability and GMM priors for multi-regime heterogeneity.
- Achieved $R^2 \approx 0.74$, RMSE ≈ 1.70, MMD ≈ 0.16, FAED ≈ 0.44, meeting fidelity, utility, and robustness benchmarks for synthetic data generation; produced fully reproducible Jupyter notebooks and a methodological roadmap for small-sample synthesis in the physical sciences.

Storm Prediction using Deep Learning

Imperial College London — *Jan 2025*

- Applied 3D-CNN, U-Net, and Transformer models on multi-source meteorological datasets for spatiotemporal storm forecasting.
- Ranked 2nd individually; contributed core model architecture and data pipeline development for final project (top 4 out of cohort).

Medical Imaging Reconstruction with DDRM

Imperial College London — *Dec 2024*

- Developed a Diffusion Denoising Restoration Model (DDRM), leveraging synthetic data generation techniques to reconstruct missing regions in MRI brain scans, effectively mitigating data scarcity in clinical imaging and improving clinical diagnostic data quality.
- Achieved A*(80%+) for outstanding technical performance and comprehensive analysis.

SKILLS

Python (pandas, NumPy, scikit-learn), LangGraph, MySQL, C++, PyTorch, TensorFlow, Git, Tableau (basic), Excel, Figma, Data Analytics, Statistical Inference, Predictive Modeling, Scrum/Agile