Samuel Oliveira

+44 7873937455 | samuelccoliveira@gmail.com | Personal Website

Education

University College London — London, UK
MSc Machine Learning — *Sept 2023 – Sept 2024*

- Predicted grade: 79% (First Class Honours).
- Courses: Probabilistic & Unsupervised Learning (81%) (Gatsby Unit PhD Module), Reinforcement Learning (87%) (taught by Google DeepMind), Statistical NLP, Applied Deep Learning (80%).
- Master's Thesis on using Diffusion Models with Transformers for Inverse Reinforcement Learning with Ilija Bogunovic. Currently training diffusion models in a multi-GPU setting on UCL's HPC cluster.

Imperial College London — London, UK
MEng Computational Bioengineering — *Sept 2019 – June 2023*

- Final Classification: 78.45 % (First Class Honours). Dean's List (top 10% of cohort).
- Relevant courses:Artificial Intelligence, Optimization (90%), Mathematical Modeling (89%), Software Engineering.

Experience

Research Assistant in Statistical Machine Learning — Oct 2022 – Sept 2023
Imperial College London — *London, UK*

- Led the implementation of a Sequential Monte-Carlo framework to predict eczema severity in under 1 second.
- Organised and taught tutorials on several ML topics, including SMC, to the research group.
- Writing it as a paper "EczemaPF: an efficient algorithm for predicting the evolution of eczema severity".

Software Engineering Intern — June 2022 – Aug 2022
Goldman Sachs — *Birmingham, UK*

- Implemented from scratch a fully functional penetration testing framework (Python, AWS S3, AWS CloudWatch) for a new web app hosted on AWS.
- Led the integration of the penetration testing framework into QA testing.
- Contributed 10K+ lines of code to an established codebase via Git.
- Presented virtually to 150+ colleagues across the firm.

Undergraduate Research Assistant in Computer Vision — April 2022 – June 2022
Imperial College London — *London, UK*

- Implemented generative models to change the skin colour of medical image data while preserving eczema symptoms.
- Developed a data processing pipeline and integrated it into an existing computer vision pipeline.
- Investigated ways to reduce the impact of inter-study image variability on eczema predictive performance.

Projects

LLM and Autoencoder based Clinical Decision Support System | *Python, PyTorch, Git*

- Developed a multi-modal pipeline aiming to output diagnostic reports for chest X-rays.
- Built a LLM-based approach with Retrieval-Augmented Generation (RAG) for trustworthy diagnostic report output.

Impact of the pre-training data distribution on the performance of MAEs | *Python, PyTorch, SQL, Linux*

- Pre-trained and finetuned Masked Autoencoders for different pre-training data distributions.
- Built a pipeline to pre-train and finetune the models on the UCL HPC system (Linux-based).

Remote Patiet Monitoring | *Java, PostgreSQL, MATLAB, Git*

- Created a Java desktop application to monitor patients' vital signs across a hospital.
- Led the backend and database (PostgreSQL) team. Created the data processing pipeline.

Technical Skills

Languages: Python, C++, SQL (PostgreSQL), R, MATLAB, Terraform
Developer Tools: Git, Docker, Singularity, AWS, CUDA, Linux-based systems
Libraries: PyTorch, NumPy, Pandas, Scikit-learn, Matplotlib, Boto3