

Calum Hand PhD.

calumhand201@gmail.com | (+44) 07765771639 | <https://www.linkedin.com/in/calum-hand/> | Bristol, UK

PROFESSIONAL SUMMARY

A highly motivated machine learning engineer with 3+ years experience in developing, and productionising end-to-end AutoML pipelines in drug discovery with a further 4+ years in chemistry based AI research. A subject matter expert across a range of technologies and their intersection with AI; always looking forwards to expand my knowledge to learn about new fields and help them succeed.

SKILLS

- **Technologies:** Python, SQL, Metaflow, AWS, Git, Bash, SQLAlchemy, Poetry, Anaconda
- **Libraries:** Pytest, Scikit-learn, Pytorch, Optuna, DuckDB, Pyarrow, Pandas, Numpy, Numba, Cuda, MLFlow, GPFlow, RDKit
- **DevOps:** Docker, CodeBuild, CodeDeploy, Github Actions

EXPERIENCE

Machine Learning Engineer

Glamorous AI : An X-CHEM Subsidiary

Remote, UK

May 2022 - Current

Machine Learning Engineer Intern

Glamorous AI : An X-CHEM Subsidiary

Remote, UK

Oct 2021 - April 2022

Worked on end-to-end training, deployment, and scaling of a range of AI models and other data-driven technologies to cloud environments for the "ArtemisAI" SaaS platform. Combined AI and domain experience to act as the intersection between AI and chemistry teams.

- **AI Research & Model Deployment:**
 - Deployed numerous AI models to production environments including: LM models (ChemBERTA), GNNs, NNs, and Random Forests (AWS, Docker, Python, Pytorch, Scikit-Learn)
 - Implemented and trained state-of-the-art graph convolution networks for multi-input binding affinity prediction models (Pytorch, AWS Sagemaker, Docker, EC2, Python)
 - Shortlisted and deployed several 3D CNN models for docking pose prediction to accelerate molecular docking simulations while minimising compute costs
 - Integrated hyperparameter optimisation into production training pipelines, resulting in 10-15% model performance increase (Python, Docker, AWS, Optuna)
- **Domain Applicability Research:**
 - Led the research, development and optimisation of several unsupervised approaches to identify when the domain of the trained model doesn't match the domain of the prediction dataset (Python, Scikit-learn, Poetry, Pytest, Numpy, Optuna)
 - Research was conducted on internal chemical datasets resulting in correctly identifying 66% of benchmark AI model prediction errors
- **Production Datastore:**
 - Independently planned and implemented the entire private data lakehouse to reduce computation times during AI training and inference by taking user input and transforming into persisted features for ML models (Python, DuckDB, SQL, Pyarrow, SQLAlchemy, S3, RDS)
 - Designed the datastore to be a user friendly client library and tool, ensuring full adoption across research and engineering teams
 - Reduced prediction pipeline run times by half
- **Docking Simulation Production Pipeline:**
 - Lead the design and productionization of a molecular docking simulation pipeline to pre-filter large molecular databases for high throughput drug candidate screenings (Python, Docker, Metaflow, Argo, Poetry, Pytest, SQLAlchemy)
 - Completed pipeline could process hundreds of thousands of molecules in under 2 hours while minimising costs through considered compute instance collection
 - Designed P.O.C integrating with active learning screenings of large molecular datasets producing promising results
- **Similarity Screening Production Pipeline:**
 - Designed and Implemented a large scale pipeline using EKS cluster for a molecular fingerprint similarity calculation achieving 2.5 billion similarity calculations in 10 minutes (Python, Docker, Metaflow, Pyarrow, Poetry, Pytest)
- **Internal Tool Development:**
 - Visualisation of chemical library overlap
 - Unified data retrieval pipelines from multiple public data sources
 - Consolidated and a wide collection of metrics for molecular similarity

- (Python, Docker, github actions, Conda, Scikit-learn, Pandas, Dask, Numpy, Rdkit, Requests)

◦ **Other Responsibilities:**

- Routinely presenting to stakeholders and clients as a subject matter expert
- Acting as scrum master in fortnightly agile development sprints
- Following TDD, striving for 90% test coverage in all projects worked on
- Instigating and organising a weekly meetup for the researchers to present problems and findings
- Maintaining AI prediction pipelines, internal data lakehouse, and associated APIs
- Conducted market analysis to identifying three further market areas for potential expansion with findings presented to stakeholders

IKEEP Collaborator

Co-Forest Ltd

Bath, UK

Feb 2021

- Worked as part of three person intrapreneurial knowledge exchange program between local business and PhD researchers
- Project involved creating a plan to establish sense of community within Co-forest public and industrial clients
- The outcomes of this collaboration included the identification of company goals in order to create content for specific customers and the drafting of relevant promotional materials.

Account Team Member

RIMES Technologies (Financial Data Management)

London, UK

Aug 2017 - Aug 2018

- Was responsible for maintaining clients data feeds, including the investigation and analysis of any data discrepancies and errors raised in within RIMES in house coding language
- Rapidly promoted within six months to work with RIMES largest client, acting as subject matter expert and advising the client on the development of their new data feeds

Climbing Tower Staff & Counsellor

YMCA Camp Silver Beach

Virginia, USA

May 2015 - Aug 2015

- Worked as staff for the rock climbing tower and counsellor to cabin of 13 campers
- Expected to be autonomous in our day to day work caring for over 120 campers a week

EDUCATION

PhD in Chemical Engineering

University of Bath

Bath, UK

Oct 2019 - March 2023

Developed methodologies for Bayesian optimisation based active learning for reduced cost high throughput screenings

MRes Sustainable Chemical Technologies

University of Bath

Bath, UK

Sept 2018 - Sept 2019

Researched Active learning in material screenings, and novel vaccination preparations to mitigate cold chain risks

MRes Green Chemistry

Imperial College London

London, UK

Sept 2016 - Sept 2017

Developed novel synthetic routes to sustainable furan monomer units for organic based solar cells

BSc Chemistry with Medicinal Chemistry

University of St. Andrews

St. Andrews, UK

Sept 2012 - June 2016

Developed a novel synthetic route valorising ligning as a biorenewable source of aromatic carbon

PUBLICATIONS

- “Scalable Drug Property Prediction via Automated Machine Learning” : *J. Chem. Inf. Model.* : Pending
- “Multi-Class Synthetic Accessibility Prediction” : *J. Chem. Inf. Model.* : Pending
- “Autonomous Exploration and Identification of High Performing Adsorbents using Active Learning” : *ChemRxiv* : 10.26434/ChemRxiv.14555706.V1
- “Bayesian Screening: Multi-test Bayesian Optimization Applied to in silico Material Screening” : *arXiv* : 2009.05418
- “The synthesis of Hibbert ketone structures in technical lignins” : *Org. Biomol. Chem.*, 2016, 14, 10023 - 10030