Sam Considine

Stratford, London | sambconsidine@gmail.com | +44 7808813731 | LinkedIn | GitHub

Profile

Machine learning engineer/scientist with 5 years experience designing, training and deploying neural network models and software. Comfortable reading and implementing bespoke models from papers and writing both research and production code. Passionate about solving challenging real-world problems using technology. Strong background in PyTorch, computer vision, neural search, NLP and language models.

Skills

Python, PyTorch, Hugging Face (Transformers, Datasets, PEFT, Accelerate, Candle), NLP, PyTorch Geometric, Data Analysis tools (Pandas, Matplotlib, Seaborn, Polars), Big Data (PySpark, AWS Glue, AWS EMR, AWS Athena, Hadoop, HDFS), databases (SQL, PostgreSQL, MongoDB) NumPy, Terraform, Cython, C, C++, Rust, Git.

Education

University of Cambridge, MPhil Advanced Computer Science *Oct 2021 - Jul 2022*

- **Grade:** Distinction — **Thesis:** Neural Algorithmic Reasoning for Pseudotime Trajectory Inference.
- Awarded best Master's thesis of the year by the faculty committee.
- St Edmund's Award for Academic Excellence.

University of York, BSc Mathematics *Sept 2013 - July 2017*

- **Grade:** First-Class Honours — **Dissertation:** Computer Vision Based Feature Extraction for Assessing Cell Heterogeneity.
- Winner of the Santander Summer Accelerator Business competition and grant for an idea for an e-commerce platform.

Experience

Interim CTO, Noggin HQ *September 2023 – Present*

- Invented and built our main credit scoring product in PyTorch: a bidirectional encoder transformer model that uses personal bank transaction records to create neural representations of borrower behaviour.
- Sourced datasets totalling 7 billion bank transaction records to train the model.
- Communication with investors, as well as full evaluation of the business feasibility and tradeoffs of various technical approaches with respect to company objectives.
- Built the company data ETL pipeline to train neural network models at scale using AWS Glue, EMR, Athena and Sagemaker, as well as custom infrastructure on EC2 to perform distributed training.
- Employed 'infrastructure as code'MLOps, using Terraform to deploy an API that could access our credit model through a Sagemaker endpoint.

Machine Learning Engineer, Loci *November 2022 – September 2023*

- Was essential in building our main product - a multimodal search and recommendation engine for game assets. This secured our first customers, including one of the world's biggest gaming companies and was a crucial component in gaining £4m in seed funding as a team of 4.

- Increased the efficiency of our existing zero-shot game asset classification system by 3 orders of magnitude while improving code simplicity by creating a custom vector store + caching system.
- Fine-tuned and deployed a large multimodal language/image transformer (OpenCLIP-2B) on 10 million novel views of 3D assets.
- Mentored a team of 6 Masters students in a collaboration with Imperial College to build a system for 3D texture synthesis from a text prompt using stable diffusion, UNet and some custom rendering.

Research Assistant, University of Cambridge *July 2022 – October 2022*

- Hired to continue research on my thesis and to build a platform to be used by clinicians to use machine learning systems to assist with diagnosis using Deep Learning.
- Worked with some of the top researchers in explainable AI, developing a UI to bridge the knowledge gap between doctors and data scientists. Allowing doctors to use private, explainable AI methods with their patients.

Senior Data Scientist, Arca-Blanca *May 2021 – October 2021*

- Worked for a Machine Learning and Data Science consulting firm, consulting with executives to and using Deep Learning to create predictive models throughout their enterprise.
- Lead various projects, collaborating with executive at large companies.
- Lead data scientist of a project for one of the UK's largest software companies, creating a model of customer churn and customer segmentation analysis. The segmentation is still used as a core part of the marketing team strategy.

Machine Learning Engineer, After the Off *September 2017 - January 2021*

- Led developer on our main horse racing product, using machine learning on live data feeds to generate in-play odds for horse races. My work helped establish us as a leading provider of in-play horse racing odds, gaining multi-million pound contracts with leading bookmakers.
- Designed and led a project using CNN-based computer vision to derive horse positions on a race track in real time using only a live video feed of the race. Built the whole pipeline starting from the video feed.
- Used techniques in scientific computing such as Monte Carlo alongside Deep Learning to drive the simulation.

Research Assistant, University of York *July 2017 - September 2017*

- Hired to continue work on my dissertation, building computer vision based feature extraction to aid in cancer research.

Self-Employed, Uptodata *June 2012 – December 2016*

- Started a successful marketing company as a teenager, providing marketing data to wholesalers.
- Made contacts in the giftware industry after finding a catalog of wholesalers from an event organised by my father, used this to build a database of giftware retailers in the UK, which allowed me to sell marketing information to wholesalers.
- Paid my way through university, allowing me to do without a maintenance loan or other job.