

# Sam Armstrong

GitHub: [github.com/Sam-Armstrong](https://github.com/Sam-Armstrong)  
LI: [linkedin.com/in/sam-a-a5557112a](https://linkedin.com/in/sam-a-a5557112a)

21 Spindle Lane, Shirley  
B90 1RP  
United Kingdom  
+447413778815  
[samuel.armstrong2014@gmail.com](mailto:samuel.armstrong2014@gmail.com)

## Personal Profile

I am a highly motivated, capable, and experienced ML & software engineer with a passion for deep learning and its applications across NLP, Computer Vision, and many other domains. I have an extensive background in both neural networks and robust software development, which are skills that I would love to advance in a new context.

## Education

2018-2021

University of Sheffield

BSc

Artificial Intelligence & Computer Science

First Class Honours (~ equiv. to 4.0 GPA)

## Skills

Python | Java | PyTorch | Pandas | C++ |  
Scrum and Kanban | Git | Docker |  
Kubernetes | Hadoop | AWS | Database  
Concepts | Data Structures and  
Algorithms | Testing | TensorFlow | SQL

## Relevant Experience

2022

### University High Achiever Profile

My final research project was remastered and submitted as a co-authored paper for publication to Multi-Agent-Based Simulation XXIII (part of the series Lecture Notes in Computer Science), due to be released in 2023. As a result, I was invited to provide a profile for the University of Sheffield website:

[www.sheffield.ac.uk/dcs/  
undergraduate/profiles/samuel](http://www.sheffield.ac.uk/dcs/undergraduate/profiles/samuel)

2021-2022 NEW ZEALAND/UK

### Presented My Academic Research

- Workshop on Multi-Agent Based Simulation (MABS) 2022
- Sheffield SIAM-IMA Applied Mathematics and Statistics Conference 2021

## Employment

JAN 2022-PRESENT

### Machine Learning Engineer - Intradiem

Worked in the AI/ML team developing scalable data pipelines to allow effective data ingestion and transformation for use in machine learning applications.

- Led development of a scalable solution for Parquet file IO that circumvented the shortcomings of Spark by using lower level libraries that made the solution scalable from local machines to deployment on AWS Elastic Kubernetes Service. This project was mission-critical for two teams and relied upon my co-ordination of team members across two time zones (London and Atlanta).
- Orchestrated the creation of a SpringBoot/Temporal Java microservice for transforming data from an S3 datalake into bucketed time-series data to be used for machine learning applications. This involved effective and performant algorithm design, due to the required solution scalability. To ensure quality at pace, a Kanban methodology was employed with pair-programming.
- Engineered a Restful microservice in Python using Flask for interacting with a MariaDB database to keep track of customer/product relationships.
- Formulated machine learning solution proposals based upon real business situations that had been identified. As an example, I developed a proposal for a transformer-based prediction model to capture the long-term seasonality trends of time-series data.
- Leader of the company AI Journal Club - a group of twelve employees where we present and discuss AI/ML topics and papers.

## Personal Projects

### 'Almanac' Football Predictor

Python

An encoder-only transformer model for predicting outcome probabilities of football matches based upon the statistics from each team's previous matches. The dataset for this model is produced from over 25,000 matches that have been scraped from the web and transformed into time-series.

*PyTorch, Pandas, Data Ingestion & Transformation*

### 'Pulsar' Deep Learning Framework

Python

Framework created from scratch in Python for designing and training custom deep neural nets with convolutional, dense, etc., layers; optimised via SGD or Adam; with choices of non-linearities, loss functions, etc. I'm working towards adding a more generalised computational graph, and additional layers such as attention, batch/layer norm, and residual connections.

*Python, Numpy, Deep Learning Concepts*

### tinyGPT / AutoPylot

Python

A generatively pretrained transformer that has been developed from scratch in PyTorch, including the self-attention mechanism. It is currently pretrained on WikiText (tinyGPT), then finetuned for generating Python code using a custom data that has been scraped from GitHub (AutoPylot).

*PyTorch, NLP*