Jordan A. **Welsman**

Data Science & Analytics Graduate | Research Affiliate

linkedin.com/in/jordanwelsman github.com/jordanwelsman
jordan.welsman@outlook.com welsman@lbl.gov
+44 7449 469738
London, United Kingdom

I specialise in deep learning, neural network design, data analysis, and **Python** software development. As a recent Data Science and Analytics graduate, I am looking for **Machine Learning engineering** roles where I can experiment and deploy **AI** and **ML** models to solve real-world problems. Currently serving as a Laboratory Affiliate at Berkeley Lab, I focus on research applications of **computer vision** and authoring **scientific research papers**. Beyond academia, I've authored numerous open-source **data science toolkits** and Python packages which have amassed more than 25,000 downloads. My expertise covers deep neural network development, testing, and benchmarking using both **PyTorch** and **TensorFlow** libraries. I've also contributed to ML applications across **data science**, computer science, climate modelling, and microscopy. I'm often acknowledged for my solution-driven research, ability to quickly adapt and learn new technologies, and insatiable hunger for knowledge.

Education

May 2024 **B.Sc. (Hons) in Data Science and Analytics**
Bournemouth University, Poole, Dorset, United Kingdom
First Class - Accredited by the British Computer Society

Qualifications

Machine Learning Experienced in ML model development, optimisation, validation, and deployment using TensorFlow, Keras, Pytorch and Scikit-Learn.

Data Mining Knowledgeable of the CRISP-DM methodology and well-practised in cleaning, analysing, and interpreting data when using it to gain business insight.

Software Development Proficient in toolkit, application, and web development with a philosophy of test-driven development, object-oriented programming, and continuous integration (CI) testing.

Professional Experience

Aug. 2023 – Sept. 2022 | **Machine Learning Research Assistant**
Lawrence Berkeley National Laboratory
Berkeley, California, United States of America

I assisted with achieving research objectives by offering novel solutions and implementing them by developing software applications. I worked across two scientific projects where my responsibilities were :

- Making valuable contributions towards an experimental climate model acceleration project with existing machine learning knowledge and model architecture tools :
  - Developed an intermediary software package (nexport) that bridges the gap between Python and Fortran, allowing the transfusion of weights and biases of *PyTorch* models. *nexport* has been downloaded ∼3,000 times.
  - Explained my existing expertise in backpropagation, forward propagation, stochastic gradient descent, and neural activation functions in a machine learning context to assist in the development of *inference-engine*, a deep learning framework written entirely in Fortran.
  - Implemented hand-initialised AND, NOT, OR, and XOR logic gates as *PyTorch* model test cases and exported their parameters using *nexport* to test the backpropagation functionality of *inference-engine*.
- Exploring methods of improving the efficacy of a scientific material tagging and storage system using deep learning and classification :
  - Developed an open-source software package that enables rapid augmentation of image datasets using a customisable pipeline and built-in augmentation functionality to support augmentation scheme benchmarking.
  - Increased the speed at which variability can be induced in high-resolution image datasets via thread concurrency from 25min/10,000 images to 30sec/10,000 images, increasing data augmentation processing by ∼50×.
  - Trained over 250 various convolutional *TensorFlow* neural networks on image datasets generated with *augmenting* on NERSC GPU nodes over 100 epochs each, totalling over 2,500 hours of parallelised training.
  - Improved classification accuracy of microscopy image resolution classification performance by up to 25%, resulting in validation accuracies exceeding 92%.
  - Compared synthetic metadata generation of popular large language models such as OpenAI's GPT-4, Meta's LLaMA-2, and Google's Bard, and cutting-edge open-source models such as Alpaca, Vicuna, and Alpaca.

Training | Optimisation | Software Engineering | Data Structures | Python | PyTorch | TensorFlow | Numpy | ML | DNNs | CNNs | LLMs

Research Projects

ScienceSearch Mar. - Aug. 2023

github.com/JordanWelsman/augmenting ScienceSearch at LBNL

"ScienceSearch is a framework that will use machine learning to auto-generate metadata to enable searchable data at supercomputing facilities." My role in ScienceSearch is performance analysis of data augmentation techniques, benchmarking of convolutional networks, and development of *augmenting*, a high-performance image augmentation toolkit.

Electron Microscopy | Image Augmentation | Network Training | Model Evaluation | Metadata Generation | Python

ML Climate Model Acceleration Sept. 2022 - Mar. 2023

github.com/NCAR/icar github.com/BerkeleyLab/inference-engine

"ICAR is a simplified atmospheric model designed primarily for climate downscaling, atmospheric sensitivity tests." My role in this project was developing an intermediary software package that supports the transfer of PyTorch model parameters in human-readable file formats to be consumed when instantiating deep neural networks in other frameworks and languages.

Climate Modelling | Network Architecture | Model Parameters | I/O | Import & Export | Python | JSON

</> Software Projects

augmenting 2023 - Present

github.com/JordanWelsman/augmenting augmenting on PyPI

augmenting is an all-in-one data augmentation framework which allows users to synthesise image datasets using a parallelised augmentation pipeline. Users can select integrated augmentation methods and create a customised pipeline which handles dataset importing, concurrent label expansion, image augmentation, and dataset exporting in one method call. It has been used in applications which require the expansion of large image datasets using various augmentation techniques, resulting in improved classification performance in scientific applications.

Image Augmentation | Dataset Expansion | Process Parallelism | High Performance | Flexibility | Label Duplication

nexport 2022 - Present

github.com/JordanWelsman/nexport nexport on PyPI

nexport is an open-source neural network interface package which exports network weights and biases to human-readable formats. It is designed to interface with PyTorch and exports neural networks and their architecture metadata to JSON. nexport has been used in research projects which required parameter transfusion of PyTorch models to models defined in other ML frameworks and programming languages.

Deep Neural Networks | Model Architecture | Trainable Parameters | Neural Connections | Network Metadata

¶ Publications

› Poster : Enhancing Electron Microscopy Image Classification Using Data Augmentation : DOI : 10.13140/RG.2.2.19365.68323
Enhancing Electron Microscopy Image Classification Using Data Augmentation Berkeley Lab Summer Program

Skills

Machine Learning	**Supervised**, Regression, Decision Trees, **Unsupervised**, Ensemble, Random Forests
Deep Learning	**CNNs**, DNNs, LLMs, Transformers, **Model Evaluation**, Performance Benchmarking
Data Analytics	**Data Visualisation**, Matplotlib, Seaborn, RStudio, Tableau, PowerBI, Graphvis
Programming Languages	**Python**, R, SQL, BASH, PHP, JavaScript, C, C++, Java, Kotlin
ML Frameworks	**PyTorch**, TensorFlow, Keras, **NumPy**, pandas, SciPy, scikit-learn, Pillow
Python Libraries	**multiprocessing**, tkinter, argparse, threading, functools, shutils, logging
Development & Collaboration	**Git**, VSCode, GitHub, PyPI, PyTest, LaTeX, Jira, Docker, Airtable
Development Practices	**Agile**, Test-driven, CI Testing, **Documentation**, Conventional Commits, Markdown
Web Development	**PHP**, HTML, CSS, Flask, Oracle SQL, MySQL, SQL Server, MAMP, XAMPP
HPC Technologies	**NERSC Systems**, Cray, iPython, Jupyter, SLURM, SSH, GPU Parallelisation
Operating Systems	**MacOS**, Linux, Windows, Raspberry Pi OS

Research Areas

Artificial Intelligence | Deep Learning | Data Science | Data Visualisation | Model Design | Software Development | Performance Evaluation | Model Benchmarking | Large Language Models | Data Structures & Algorithms

Synergistic Activities

- Authored and submitted a conference paper with Berkeley Lab on the work I did during my placement.
- Hosted a talk on international work placements at Bournemouth University to over 30 students in January 2024.
- Spoke to over 200 students about my industrial work experience at Berkeley Lab on a panel at Bournemouth University in December 2023.
- Became a Berkeley Lab affiliate to continue my academic efforts towards ScienceSearch in September 2023.
- Presented a research poster at Berkeley Lab's Summer Student Program poster session in August 2023.
- Attended the Fortran Standards Committee winter meetings at Berkeley Lab in February 2023.
- Attended the International Conference for High-Performance Computing, Networking, Storage, and Analysis (SC22) in November 2022.

Voluntary Experience

Present – **Sept. 2023**
Laboratory Affiliate
Lawrence Berkeley National Laboratory
Remote : Berkeley, California, United States of America

My responsibility as an LBNL lab affiliate involves :

- Continuing my work on the ScienceSearch and climate model acceleration research projects :
  - Working on label synthesis and synonym generation for automatic tagging of scientific text with large language models.
  - Developing and improving relevant software packages for improvements in scientific procedures.
  - Supporting the current and future use of my software packages in scientific projects.
  - Submitting software disclosures per lab policy to publicly release future open-source software.

Research | Science | Remote | Scientific Documentation | LaTeX

Jun. 2022 – **Nov. 2020**
Student Representative → Computing & Informatics Departmental Representative
Students' Union at Bournemouth University
Poole, Dorset, United Kingdom

The Students' Union at Bournemouth University (SUBU) is a non-profit organisation focused on improving BU students' experiences throughout their time at university. Liaising between students and departmental representatives as a student representative, and student representatives and faculty representatives as a departmental representative, my responsibilities included :

- Liaising between university staff, SUBU officers, and student representatives as the departmental representative for BU's Computing & Informatics department :
  - Positively impacted the university experiences of ∼750 computing students by recommending improvements to university policy and department teaching practices, leaving an improved experience for future students.
  - Escalated critical feedback from student representatives to a departmental level, reducing the response time from faculty by an average of 2 weeks, expediting the process of policy and content improvement by ∼100%.
  - Assisted in improving the attendance system used in online lectures and seminars during the COVID-19 pandemic by providing feedback from the students that used it, resulting in a ∼25% increase in use.
- Providing a feedback pipeline from Data Science and Analytics students to the departmental representatives as a student representative :
  - Implemented polls in a non-official computing communication channel with over 200 members, expediting the process of providing concise opinions and feedback on computing department policies.
  - Collaborated with ∼25 students on my course to elevate feedback and problems with course content and advocated for changes to unclear assignment briefs and coursework guidelines.
  - Built and ran a gaming server for 2 years during the COVID-19 pandemic with the goal of stimulating conversation and encouraging community among students.

Representation | Quality | Improvement | Communication | Equity | Cooperation

Languages

English ●●●●●
Spanish ●●○○○
French ●○○○○
ASL ●○○○○

Interests

- Formula 1
- Weightlifting
- Keyboard Building
- Home Automation

References

References available on request