

# Alex McKinney

✉ [alex.f.mckinney@gmail.com](mailto:alex.f.mckinney@gmail.com) |  [vvwm23](https://github.com/vvwm23) |  [afmck.in](https://www.linkedin.com/in/afmck) |  [Scholar](https://orcid.org/0000-0001-9485-4141)

---

## Experience

**Machine Learning Research Engineer** | *Unitary, United Kingdom* September 2023 – November 2023

- Machine Learning Research Engineer in the research team at a content moderation startup, focusing on large multimodal understanding models.

**Artificial Intelligence Engineer** | *Graphcore, Bristol HQ* September 2022 – September 2023

- Artificial Intelligence Engineer in the large models team at an AI accelerator startup.
- Ported a **176 billion** parameter large language model (**Bloom-176B**) to IPU, utilising tensor parallelism and phased execution across 16 accelerators.
- Developed Jupyter notebooks for **Dolly 2.0** – an instruction fine-tuned LLM – and **OpenAssistant** – a chat-based AI assistant. Also developed **Stable Diffusion** and **Dreambooth** fine-tuning for IPU.

**Teaching Assistant** | *Durham University, United Kingdom* September 2020 – March 2022

- Taught introductory Python programming and propositional logic to first-year students at a top-10 UK university.
- Involved remote and in-person teaching, presenting content, creating class notes, and answering questions from students with varied technical backgrounds.

**Research Intern** | *OFFIS – Institut für Informatik, Oldenburg, Germany* June – September 2021

- Research Intern as part of the DAAD RISE Germany research exchange scheme.
- Self-proposed project using **contrastive predictive coding** for the unsupervised representation learning of binaural audio to improve non-intrusive speech intelligibility prediction systems. Our measure highly correlated with the ground truth (>90%) and surpassed all baselines.
- Accepted at IEEE Signal Processing Letters.**

**Cyber Security Intern** | *Her Majesty's Government, United Kingdom* July – September 2019

- Completed cyber security training courses on offensive and defensive tactics.
- Involved a self-proposed project to train LSTM networks for computer network intrusion detection.

---

## Highlighted Projects

**Pytorch Projects** | *Python, PyTorch, Generative Modelling, NLP, Diffusion Models, RL, PEFT.*

- Many open-source projects reimplementing developments in AI research, with a focus on modularity, cleanliness, and educational value. Below are some highlighted projects:
- VQ-VAE-2** implementation that supports an arbitrary number of vector quantization codebooks, evaluated on FFHQ-1024 image reconstructions. [\[Github\]](#)
- Step-unrolled denoising autoencoders (**SUNDAE**) for non-autoregressive, character-level text generation. Improved inference speed via masked sampling. [\[Github\]](#)
- Personal framework around Pytorch** for supporting research experiments. Includes features such as automatic mixed-precision, device management, and Weights and Biases integration. [\[Github\]](#)
- Stable Diffusion x Segmentation model demo** in **Gradio** for the fast generation of inpainting masks based on detected objects in the scene. [\[Huggingface Spaces\]](#)
- ALBERT** (A Lite BERT) with efficient attention finetuned for multi-label sentiment analysis on the JIGSAW Toxicity Classification Dataset using Huggingface datasets. [\[Github\]](#)
- Implementations of DQN variants and **Rainbow DQN** in the Atari Learning Environment. [\[Github\]](#)
- Currently integrating a new **parameter efficient fine-tuning method VeRA** into Huggingface's **peft** library. [\[Github\]](#)

**JAX Projects** | *Python, JAX, Flax, Equinox, Generative Modelling, PEFT, NLP.*

- **TchAIkovsky** – a transformer decoder model for **MIDI generation** trained from scratch on a dataset of piano performances. Implemented using JAX library **Equinox** and trained on **8 TPUs**. [\[Github\]](#)
- **MeZO** – 0th order fine-tuning using function transformations. Allows for fine-tuning arbitrary JAX models with a **12x reduction in memory usage** compared to full fine-tuning. [\[Github\]](#)
- **Llama** – implementation of **Llama** and variants in JAX using the **Flax** neural network library. Currently being integrated into **Huggingface's transformers library**. [\[Github\]](#)
- Led a team during the **Huggingface Diffusers Sprint 2023** into **image generation** using **discrete diffusion models**. Implemented in **Flax** and trained on **4 TPUs** provided by Google Cloud. [\[Github\]](#)

*Miscellaneous Projects*

- **Technical writing** on my blog that cover topics such as JAX deep-dives, productive computing, and interesting use cases for AI. [\[Website\]](#)

---

## Education

**Durham University**

United Kingdom

*MEng. Computer Science*

*October 2018 – June 2022*

- Graduated with a **first class** honours degree with a **79.66%** average.
- Master's thesis on fast image generation using step-unrolled denoising autoencoders, capable of generating **megapixel images in  $\approx 2$  seconds**.
- *Relevant Modules: Deep Learning, Reinforcement Learning, Machine Learning, Advanced Computer Vision, Natural Language Processing, Parallel Scientific Computing I/II, Single Mathematics A.*

---

## Research

- **Alex F. McKinney** and Chris G. Willcocks | Megapixel Image Generation with Step-Unrolled Denoising Autoencoders | 2022 | [\[arXiv\]](#) [\[Github\]](#)
- **Alex F. McKinney** and Benjamin Cauchi | Non-intrusive Speech Intelligibility Prediction from Discrete Latent Representations | 2022 | [\[IEEE Signal Processing Letters\]](#) [\[Github\]](#)

---

## Skills

**Programming Languages**

*Proficient in:*  
*Experience with:*

Python (6 years).  
Rust, C/C++, JavaScript,  $\text{\LaTeX}$ .

**Libraries and Frameworks**

*Proficient in:*  
*Experience with:*

PyTorch (5 years), NumPy (7 years), Huggingface (3 years), JAX (1 year).  
Flax, Equinox, Gradio, TensorFlow, Matplotlib, Scikit-learn, Pandas, W&B.

**Machine Learning**

Distributed Training & Inference, Generative Modelling, Natural Language Processing, Computer Vision, Unsupervised Representation Learning, Audio Processing, Diffusion Models, Multimodal Models, Video Understanding Models, Large Language Models, Parameter Efficient Fine-tuning.

**Software**

Git, GitHub, Bash, Zsh, Linux, MacOS, Slurm, Vim, VSCode, Jupyter.

**Languages**

Native English; Intermediate Reading & Writing Simplified & Traditional Chinese.