Alex McKinney

✉ alex.f.mckinney@gmail.com | ⌗ vvvm23 | 🖥 afmck.in | G Scholar

---

Experience

ML Research Engineer | *Unitary*   September 2023 -- Present
- ML Engineer in the research team at a content moderation startup, focusing on large multimodal understanding models.

AI Engineer | *Graphcore, Bristol HQ*   September 2022 -- September 2023
- AI 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 interactive Python notebooks for **Dolly 2.0** – an instruction fine-tuned LLM – and **OpenAssistant** – a chat-based AI assistant.
- Now working on text-to-image models, porting **Stable Diffusion** fine-tuning and **Dreambooth** to 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 and supervised by Dr. Benjamin Cauchi.
- 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.
- Carried out an individual project into using LSTM neural networks for computer network intrusion detection.

Highlighted Projects

PyTorch Paper Implementations | *Python, PyTorch, Generative Modelling, Natural Language Processing.*
- *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]**
- **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]**

MeZO JAX: Zeroth order fine-tuning in JAX. | *Python, JAX*
- JAX library implementing memory efficient zeroth order LLM fine-tuning by transforming arbitrary JAX functions. Allows for a **12x reduction in memory usage** compared to full fine-tuning.

ptpt: PyTorch Personal Trainer | *Python, PyTorch*

- Personal framework built around PyTorch for rapid experimentation and iteration. Focused on abstracting implementation details and engineering tricks whilst retaining flexibility.
- Features automatic mixed-precision, compute device management, arbitrary callback functions, and Weights and Biases integration that can be applied with minimal code change.

Stable Diffusion Inpainting x Segmentation | *Python, PyTorch, Huggingface, Gradio*

- Web demo in **Gradio** that combines **Stable Diffusion** inpainting with a segmentation model.
- Segmentation model produces per-class masks which can be individually controlled for fast and controllable mask generation.

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:*	Python (6 years).
	Experience with:	Rust, C/C++, JavaScript, LaTeX.
Libraries and Frameworks	*Proficient in:*	PyTorch (4 years), NumPy (6 years), Huggingface (2 years).
	Experience with:	JAX, Flax, Gradio, TensorFlow, Matplotlib, Scikit-learn, Pandas, W&B.
Software		Git, GitHub, Bash, Zsh, Linux, Slurm, Vim, Jupyter.