Edwin Chacko

🗘 /EdwinChacko75 🖵 edwinchacko.netlify.app/ 🛅 /in/edwin-chacko 💌 <u>edwin.chacko@mail.utoronto.ca</u> 🛽 <u>647-540-5048</u>

EDUCATION

University of Toronto

BASc. in Engineering Science, Machine Intelligence Option.

Relevant Coursework: Machine Learning, Data structures, Algorithms, Matrix Algebra and Optimization, Computational Linguistics, Natural Language Computing, Computer Architecture, Probability & Statistics

Technical Skills

Languages: Python, C/C++, CUDA, SQL, JS, Assembly, Verilog, MATLAB, React, Django Libraries: TensorFlow, PyTorch, scikit-learn, keras, NLTK, NumPy, pandas, Matplotlib, Hugging Face Transformers Developer Tools: Git, LLMs, Docker, VS Code, Visual Studio, Linux (Ubuntu), Shell/Bash, HDF5

Experience

Machine Learning Researcher

McMaster University - ChemAI Lab

- Founded and leading the Spectro project at McMaster University, exploring novel applications of AI in chemistry
- Designed a multi-modal molecule prediction model, **Spectro**, achieving an **accuracy of 93%** (see projects)
- Worked closely with domain experts to integrate chemistry-specific insights in model development and validation
- Co-authored a paper accepted to AI4Mat-NeurIPS and will present findings at the 2024 NeurIPS conference

Calibrations Engineering Intern

VACS Calibrations

- Calibrated electronic and mechanical equipment, following the IEE and ISO17025 standards
- Performed statistical analysis (e.g., uncertainty propagation, regression), validating calibration accuracy within a 95% confidence interval, and improving equipment reliability by 10%.

Projects

Spectro | TensorFlow, PyTorch, Hugging Face Transformers, Linux

- Led development of a multimodal molecule elucidation model using IR and NMR data
- Fine-tuned a CNN-based vision model for functional group prediction from images, achieving an **f1 score of 91%**
- Designed and tuned a **RNN** with **LSTM** decoder, achieving 93% accuracy in molecule prediction
- Utilized LLaMA 3 and GPT-2 to generate embeddings from NMR text data, enabling multimodal integration
- Implemented a complete ML pipeline in TensorFlow, incorporating data augmentation, oversampling, cross-validation, custom learning rate scheduling, and distributed training

Advanced NLP and Computational Linguistics Models | PyTorch, NLTk, Transformers Sep. 2024 - Present

- Implemented sentiment analysis models using Llama 3.1 and feature-based classifiers and built transformer models for machine translation, evaluated using BLEU scores
- Developed algorithms using WordNet, NLTK, and BERT to resolve word ambiguity, with word2vec embeddings
- Conducted causal tracing experiments on GPT-2 to analyze information flow, investigating the impact of different model sizes and prompts on language model behaviour

Chess NNUE (Efficiently Updatable Neural Network) | PyTorch, C++, SQL, HDF5May 2024 – Present

- Developing chess static evaluation with NNUE in PyTorch, currently achieving 80% accuracy
- Augmented and preprocessed 83 million data points, exploring utilizing a custom C dataloader
- Bridging the int_8 quantized variant into my C++ chess engine, reducing inference time by 70%

Chess Engine | C++, CUDA, Docker, Postman, JavaScript

- C++ chess engine (rated 1800) with hardware optimizations like BitBoards and compile-time optimizations
- Implemented multithreaded Negamax search with alpha-beta pruning, reducing search time by 45%
- Integrated Zobrist hashing, a transposition table, and quiescence search to boost performance by 90%
- Utilized custom **CUDA** kernels for move generation and evaluations, significantly accelerating parallelizable tasks to achieve around 35 million nodes per second in performance test

PUBLICATIONS

[1] Chacko, Sondhi, et al. A Multi-modal Approach for Molecule Elucidation Using IR and NMR Data. AI4Mat-NeurIPS 2024. December 2024.

May 2024 – Present

Jun. 2023 – Apr. 2024

Toronto, ON

May 2023 – Sep. 2023

Sep. 2022 – May 2027 Toronto, Canada

May 2024 – Present

Hamilton, ON