Edwin Chacko

 /EdwinChacko75   edwinchacko.netlify.app/   /in/edwin-chacko   edwin.chacko@mail.utoronto.ca   647-540-5048

EDUCATION

University of Toronto   Sep. 2022 – May 2027
BASc. in Engineering Science, Machine Intelligence Option.   *Toronto, Canada*
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   May 2024 – Present
McMaster University - ChemAI Lab   *Hamilton, ON*
- **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).
- Collaborated with Dr. Kylie Luksa and other domain experts to inform model development and behaviour.
- **Co-authored** a paper accepted to **AI4Mat-NeurIPS** and will present findings at the 2024 NeurIPS conference.

Calibrations Engineering Intern   May 2023 – Sep. 2023
VACS Calibrations   *Toronto, ON*
- Calibrated electronic and mechanical equipment, following the **IEE and ISO17025** standards.
- Performed **statistical analysis**, including standard deviation, uncertainty propagation, and regression,
  to validate calibration accuracy and assess reliability.

PROJECTS

Spectro | *TensorFlow, PyTorch, Hugging Face Transformers, Linux*   May 2024 – Present
- **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**.

NLP and Computational Linguistics Course Projects | *PyTorch, NLTk, Transformers*   Sep. 2024 - Present
- Applied **LLMs** (BERT, LLaMA) to tasks like text classification, sentiment analysis, and dependency parsing.
- Used transformer-based embeddings for sentiment analysis in the WSJ dataset, improving accuracy by 20%.
- Applied **tokenization, vectorization**, and sequence modeling techniques to real world scenarios.

Chess NNUE (Efficiently Updatable Neural Network) | *PyTorch, C++, SQL, HDF5*   May 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*   Jun. 2023 – Apr. 2024
- 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 improve search accuracy and
  eliminate redundant calculations.
- 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.