

Weizhi PENG

Ontario, Canada

Phone: +1 (778)-885-8117; Email: jimpeng98@gmail.com Website: pengwz.info/about/

Education

Lakehead University

MSc Eng in Electrical and Computer Engineering

Thunder Bay, ON, Canada

01.2024 - now

King's College London

MSc in Artificial intelligence

London, UK

09.2021 - 01.2023

University of Birmingham

Major in BSc Computer Science

Birmingham, UK

09.2019 - 07.2021

B.Sc. Computer Science with Honors, Class I. GPA: 4.25/4.25

Coursework includes C/C++, Java, Python Programming, Neural Networking, Machine Learning and Intelligent Data Analysis, Advanced Networking, Computer Vision and Imaging, Mathematic Modelling & Decision Making, Software Engineering, Algebra and Geometry, Discrete Mathematics, Probability and Statistics, Data Structures, etc.

Work Experience

Shengqu Game

Machine Learning Engineer Internship

Shanghai, China

Unreal Department - Artificial Intelligence Group

06.2023 - 12.2023

- **Large-Scale LLM Development (PyTorch):** Led the creation of a 20B-parameter language model for real-time code completion, optimizing for low-latency inference and high developer satisfaction.
- **Distributed Training Optimization (Accelerate & DeepSpeed):** Streamlined multi-GPU training pipelines on a 4× A100 Linux cluster, implementing ZeRO optimization and a custom evaluation framework to maintain accuracy. Deployed the model with PageAttention to achieve 3x faster token output, significantly improving the user experience for coding suggestions.
- **Low-Latency Inference (PageAttention):** Integrated PageAttention to achieve 3× faster token generation, significantly improving user responsiveness.
- **IDE Integration (VS Code Extension):** Developed a TypeScript-based extension with GitHub Copilot-style code completion and fine-tuned chat, enabling direct API queries inside the editor.
- **Open-Source Contribution:** Published project as DevAssistant, showcasing AI-assisted coding workflows.
- **Github:** <https://github.com/Miraclove/DevAssistant>

BOC International

Quantitative Machine Learning Engineer Internship

Shanghai, China

Fintech Department

01.2023 - 06.2023

- **Financial Time-Series Modeling (Transformer/GPT/BERT):** Designed and optimized HFT models for market forecasting, adapting NLP architectures to numeric data.
- **Market Data Engineering (Pandas/SQL):** Processed large-scale financial datasets, extracting actionable trading signals and managing feature engineering pipelines.
- **Model Deployment in Production Trading Systems:** Collaborated with software engineers to embed ML models into low-latency execution environments for live trading.
- **Hyperparameter Optimization (Optuna/Grid Search):** Conducted rigorous tuning and backtesting to enhance model reliability under volatile market conditions.

Research Experience

King's College London

Few-shot Learning for Text Generation

London, UK

Researcher

10.2021 - 06.2022

- **Few-Shot Learning with Pre-Trained Models (T5, GPT-2, BART):** Applied state-of-the-art NLP architectures in academic research to enable efficient few-shot learning capabilities.
- **Model Evaluation & Optimization (BLEU-4, ROUGE-4):** Trained, tested, and fine-tuned models using standardized NLP metrics to assess performance and improve output quality.

University of Birmingham

Development of the COVID Corpus Website

Birmingham, UK

Developer

10.2020 - 06.2021

- **NLP API Deployment (Django + React Integration):** Deployed NLP functions on a Django server, delivering robust APIs for seamless retrieval of analysis results. Led the NLP module development for the COVID Corpus website using React, collaborating with the dev team to ensure accurate and effective data visualization.
- **Data Preprocessing & Topic Modelling (NLP Pipelines + LDA):** Cleaned and preprocessed large-scale COVID-19 research datasets with custom NLP pipelines for high-quality analysis, applying Latent Dirichlet Allocation (LDA) for effective topic modeling.
- **Project Demonstration:** <https://pengwz.info/service/nlp-analysis/>

Additional Skills

- **Computer Language:** Python, Java, C++, C, SQL, JavaScript, HTML, Haskell
- **Software:** Pytorch, Linux, Visual Studio, Eclipse, MATLAB, Wireshark, SQL Server, React, NumPy, Pandas, Django