

Weizhi PENG

Ontario, Canada

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

Harbin Institute of Technology at Weihai

Major in BEng Software Engineering

Weihai, China

09.2017 - 07.2021

GPA: 85/100 (TOP 10%)

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

Work Experience

Shengqu Game

Machine Learning Internship

Unreal Department - Artificial Intelligence Group

Shanghai, China

08.2023 - 11.2023

- Led the development of a 20-billion-parameter language model (PyTorch) for real-time code completion, prioritizing low-latency inference and high developer satisfaction.
- Streamlined distributed training pipelines using Accelerate and DeepSpeed on a 4x A100 GPU Linux cluster, implementing a custom evaluation framework to maintain model accuracy.
- Deployed the model with PageAttention to achieve 3x faster token output, significantly improving the user experience for coding suggestions.
- Engineered a Visual Studio Code extension integrating company API documentation for GitHub Copilot-style auto-completion and fine-tuned chat capabilities, enabling on-demand API queries directly within the editor.
- Github: <https://github.com/Miraclove/DevAssistant>

BOC International

Quantitative Machine Learning Internship

Fintech Department

Shanghai, China

04.2023 - 07.2023

- Architected and optimized high-frequency trading models built on Transformer, GPT, and BERT architectures for time-series forecasting in financial markets.
- Analyzed large-scale financial data to extract actionable insights, partnering with software engineers to embed ML models within production trading systems.
- Employed rigorous experimentation and hyperparameter tuning, improving model reliability and performance under real-world market conditions.

Research Experience

King's College London

Few-shot Learning for Text Generation

Researcher

London, UK

10.2020 - 06.2021

- In academic research, pre-training models such as T5, GPT2, Bart, etc are used to generate Table to Text generation by Few-shot Learning and memory is utilized to store pre-training knowledge and training instance selection is used to improve training performance and achieve higher performance.
- Pre-trained models are trained and tested using books, songs and other public datasets, evaluated and optimized using BLEU-4, ROUGE-4 and other methods.

University of Birmingham

Development of the COVID Corpus Website

Developer

Birmingham, UK

10.2019 - 06.2020

- Deployed NLP functions on a Django server, providing robust APIs for seamless analysis result retrieval. Spearheaded the NLP module development for the COVID Corpus website using the React framework, working closely with the development team to ensure accurate and effective data visualization.
- Cleaned and preprocessed COVID-19 research data with NLP pipelines for quality analysis. Applied LDA algorithms for effective topic modelling in large datasets.
- Project Demonstration: <https://www.miraclove.com/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