

# Wyett “Huaye” Zeng

[wyettzeng@g.harvard.edu](mailto:wyettzeng@g.harvard.edu) | 519-729-8107 | [wyett-zeng.com](http://wyett-zeng.com) | [LinkedIn](#) | [GitHub](#) | [Google Scholar](#)

## Educations

**University of Waterloo & Wilfrid Laurier University**  
Bachelor of Computer Science and Business Administration  
Laurier Alumni Gold Medalist (Major Average 94.79 / 100)

Waterloo, Canada  
Sep 2020 – Apr 2025

**Harvard University**  
Master of Science in Data Science

Cambridge, Massachusetts, USA  
Sep 2025 – Dec 2026 (Expected)

## Summary of Qualifications

- Fluent in programming languages such as **Python, Java, C/C++, Go, Racket, JavaScript**. Experienced with data science packages such as **PyTorch, Keras, Pandas**, and **Numpy**. Knowledgeable in database systems such as **Clickhouse, PostgreSQL**, and **MySQL**. Strong quantitative analytical skills in **SQL, R, Tableau**, and **Power BI**.
- Laurier Student Investment Fund Winter 2025 - fundamental analysis in public equities.
- Actively running a **\$200,000 portfolio** in public equity and fixed income with my own money; risk is managed through diversification in asset classes and geographical allocations. Currently, the NA portfolio has a **25.83% IRR**.
- **200+ citations** for machine learning publications in top venues (ACL & TMLR), including first-author papers [[Google Scholar](#)].

## Work Experiences

### Machine Learning Engineer | GPTZero

Sep 2024 – Aug 2025

- Led the AI-Grader product using transformer-based architectures. On IELTS, a widely used standardized English tests, the model achieves **88% accuracy** within  $\pm 1.0$  band (out of 12) and **97%** within  $\pm 1.5$  bands, rivaling human graders.
- Optimized the grammatical error correction model, reducing edit distance by **50%** and improving GLEU score from 0.7 to 0.8.

### Quantitative Developer | Gradient Boosted Investments Inc.

Jan 2024 – Apr 2024

- Added features in the **Boosted.ai trading algorithm** with Python to optimize daily stock selection for all the company’s clients. The added features expand the algorithm’s capabilities to construct portfolios that align closer with the client’s needs.
- Added features for hedge basket rankings, which ranks multiple hedging strategies with four methods: defensive which focuses on stability and minimizes risks, factor tolerant which aims for maximum performance, idiosyncratic which maximizes idiosyncratic return from the portfolio, and all-weather which takes a balanced approach.
- Rewrote the factor model algorithm, reducing scheduled inference time for **5,000+** customer models by over **90%**. The algorithm uses NumPy, ClickHouse, and PostgreSQL to efficiently compute economic factor values for **20,000+ publicly listed securities** and ETFs daily. The data is then queried by a **Java** backend, post-processed, and displayed to end users.

### Data Scientist | CIBC – Gallant MacDonald

Jan 2023 – Apr 2023

- Developed the quantitative portfolio builder, which can construct a portfolio whose return is within  **$\pm 2.8\%$  of the desired return** using QSolver to provide insight into the more “obscure” alternative investment hedge funds.
- Developed the market analysis report that presents hundreds of market trend graphs to team members in less than 3 minutes. The algorithm is created with Morningstar API, pandas, and Seaborn.
- Partook in numerous due diligence meetings with portfolio managers from big hedge funds such as TCC, Group RMC, and Hamilton Lane. After the meeting, produced a detailed report identifying areas of concern such as liquidity options, distribution schedule, market correlation, fx risks, and interest risk.

## Research Experiences

### Researcher | Wilfrid Laurier University

Sep 2023 – Present

- Completed an **Honours Thesis** on training LLMs to predict market sentiment from economic news by developing a dataset of over 120K entries from Dow Jones News Wires and Wharton Research Data (**Publication under writing**).
- Fine-tuned custom LLMs by combining the Llama base model with custom heads using PyTorch, DeepSpeed, and QLoRA, then transformed the model to a Transformer-based trading algorithm; back tested trade shows **over 15% IRR**.
- Studied topic compositions of financial news over time using methods such as Latent Dirichlet Allocation and Gaussian Mixture Model on the Bert embedding of news text.