Wyett "Huaye" Zeng

wyettzeng@g.harvard.edu | 519-729-8107 | wyett-zeng.com | LinkedIn | GitHub | Google Scholar

Summary of Qualifications

- Languages: Java, Python, C++, C, C#, SQL, Go, Bash, JavaScript, HTML/CSS
- Tools: gRPC, Protobuf, GraphQL, AWS, Azure, PyTorch, LLamaIndex, LangChain, Pandas, NumPy, Slurm, Clickhouse, Docker
- Relevant Courses: Object Oriented Programming (95), Data Structure (99), Algorithms (99), OS (96), Machine Learning Statistics (97), Intro to AI (100), Reinforcement Learning (91), ML Dev Operations (In Progress), Unsupervised Learning (In Progress).
- **200+ citations** for machine learning publications in top venues (ACL & COLM), including first-author papers [Google Scholar].

Education

Harvard University — Cambridge, Massachusetts, USA
Master of Science in Data Science — Sep 2025 – Dec 2026 (Expected)

University of Waterloo & Wilfrid Laurier University — Waterloo, Ontario, Canada
Bachelor of Computer Science and Bachelor of Business Administration Double Degree — Sep 2020 – Apr 2025
Laurier Alumni Gold Medalist (Major Average: 94.79/100)

Work Experiences

Software Developer Intern (AI/ML) | GPTZero — Sep 2024 – Aug 2025 | Toronto, Canada

- Spearheaded the writing coach product, implemented a new **AWS Lambda** function to parse large user documents and used RAG with LlamaIndex to generate relevant feedback, increased **user screen time by 30%** and satisfaction rating by **15%**.
- Launched the AI-Reviewer product, developing an end-to-end ML server with a multi-step agent workflow using **LlamaIndex** for retrieval, grading, and feedback. The system processed **1,000+** assignments in its first month—surpassing all company target.
- Migrated the writing feedback system from Prompt Flow and **Flask** to LlamaIndex and **Quart**, redesigning API routing and introducing asynchronous request handling, enhanced scalability and reduced processing time on internal API endpoints by **30%.**
- Led the AI-Grader product using transformer-based architectures with **PyTorch**. On IELTS, a widely used standardized English tests, the model achieves **88% accuracy** within ±1.0 band (out of 12) and **97%** within ±1.5 bands, rivaling human graders.

Software Developer Intern | Boosted.ai — Jan 2024 – Apr 2024 | Toronto, Canada

- Rewrote the core factor model algorithm, significantly reducing 5,000+ customer models' scheduled inference time by over **90%**, resulting in weekly savings of **500+ hours** of computation time on **AWS EC2**. The algorithm leverages **NumPy**, **ClickHouse**, and **PostgreSQL** to efficiently compute and update daily economic factor values for **20,000+** publicly listed securities.
- Developed the investment style matching feature for 150+ institutional clients using **GraphQL, gRPC,** and **Protobuf**. The feature automatically analyzes client portfolios and reports the fitness between the portfolios and their selected investment style.
- Developed and deployed multiple enhancements for the signature Boosted.ai trading algorithm spanning analytics server, database retrieval, and backend server using **Python, Java, gRPC**, and **Protobuf**, increasing client satisfaction scores by over 6%.
- Developed new AI commentary features leveraging **LangChain** for prompting LLMs to deliver tailored portfolio analysis and insights on macroeconomic topics for 150+ institutional clients worldwide, achieving over **85%** user adoption within two weeks.

Data Scientist Intern | Canadian Imperial Bank of Commerce — Jan 2023 – Apr 2023 | Toronto, Canada

- Set up and deployed a **PostgreSQL** server on **AWS** to centralize financial data storage for alternative investments, enabling parallel query execution and caching that accelerated data retrieval by **3×** and improved synchronization throughput by **45%**.
- Developed a client report generation application using **PyQt5**, integrating data from multiple external partners through **REST API** calls, cutting monthly processing time **from 25 to 6 hours** while completely remove any potential for human errors.
- Developed a market analysis program that integrates streaming data from Morningstar APIs, generates **300+** interactive market trend graphs for team members in minutes, and delivers short-term forecasts using GRU and LSTM models built with **TensorFlow**.

Software Engineer Intern | Siemens — Jan 2022 – Apr 2022 | Ottawa, Canada

- Developed and maintained the Android-based NXS application using **Java** and **Android Studio**, strictly adhering to **object-oriented design** patterns such as MVC and Singleton to ensure modularity, maintainability, and reliable hardware interfacing.
- Fixed a critical Glucose conversion issue by resolving incorrect asynchronous saving logic with **Realm database** and **Reactive Java**, and restructured the data handling pipeline to improve stability and decreased screen loading time by **70%**.

Publications

- **(First Author, ACL 2025)** AceCoder: Acing Coder RL via Automated Test-Case Synthesis [paper][website][huggingface].
- **(COLM 2025)** ScholarCopilot: Training LLMs for Academic Writing with Accurate Citations [paper][website][demo].
- **(TMLR 2024)** MANTIS: Interleaved Multi-Image Instruction Tuning [paper][website][demo].