

Yuxuan (Wayne) Wang

929-583-8556 • yw5954@nyu.edu • <https://www.linkedin.com/in/wayne-wang-364174231/>

EDUCATION SUMMARY:

New York University, Tandon School of Engineering, Brooklyn, NY
Bachelor of Science **Computer Science**, Math. **GPA: 3.944**

Expected Graduation Date: May 2025

TECHNICAL SKILLS:

Python || **C++/C** || **Linux** || **Github** || **Search Engine Optimization** (query processing, dense/sparse re-ranking) || **Machine Learning** (Deep Learning, Reinforcement Learning) || **DevOps Tools** (Docker, Circle-CI, Clang) || **Database Management** (MySQL, NoSQL) || **Web Development** (Flask, SwiftUI)

RELEVANT COURSEWORK:

Computer Science: Algorithmic Machine Learning and Data Science || Network Security || Computer Networking || Data Structure || Algorithm || Computer Architecture and Organization || Database || Operating System

EXPERIENCES:

Researcher – Evaluation of Graph-Based Vocabulary Mismatch Solution in Information Retrieval June 2024 – Present
Supervised by Professor Torsten Suel @ New York University, Tandon School of Engineering
(Information Retrieval, Search Engine, Query Processing, Database, Python)

- Investigated the effectiveness of graph-based neighbor search and expansion to address vocabulary mismatch in information retrieval, with a primary focus on LADR (Lexically-Accelerated Dense Retrieval).
- Designed an experimental setup to evaluate LADR's efficiency in reranking by using various initial sets of 200–1000 passages obtained through different retrieval methods, including BM25, DeepImpact, DeeperImpact, SPLADEv2, TILDE, and docT5query. This setup allowed for efficient reranking within a pool of 3,000 passages rather than the entire 8.8 million document corpus.
- Testing with multiple dense models (TasB, ColBERT, DeBERTa, and MiniLM) Demonstrated that the pool-based ranking within 3,000 passages achieves results comparable to full reranking of the entire corpus. Additionally, experiments using both KNN and HNSW graph-building methods confirmed that graph-based expansion provides an efficient intermediate solution for handling vocabulary mismatch, allowing cheaper initial retrieval methods to achieve robust results in reranking.

Internship - Industrial Investment Data Analytics

June 2023 - August 2023

Sinosure (Guangzhou) Equity Investment Fund Management Co., Ltd., Guangzhou, China
(Deep Learning, Web Crawler, Database, Python, MySQL)

- Designed a cloud-native, scalable data pipeline using AWS Lambda, Apache Kafka, and MySQL/DynamoDB, enabling real-time extraction, transformation, and analysis of Series B funding data in the Electronic Design Automation (EDA) sector.
- Leveraged advanced ML and NLP techniques (e.g., XGBoost, BERT) to predict funding profitability with 85% accuracy and extract actionable insights from unstructured text, such as market sentiment and key trends.
- Developed interactive dashboards with Tableau and Streamlit, visualizing real-time forecasts, competitor analysis, and sentiment trends, improving decision-making efficiency by 30% through automated insights delivery.

PROJECTS:

ShadowDash (Open Source Development)

September 2024 - Present

(Build Optimization, C, CI/CD Deployment)

- Redefined build configuration by developing a C-based declarative language that replaced .ninja files, enabling direct in-memory graph construction and eliminating inefficiencies in Ninja's character-by-character parsing process.
- Engineered advanced optimizations, including scope encapsulation and parallel compilation, achieving up to a 2x speedup for LLVM manifests and a 1.73x improvement for zlib parsing, making ShadowDash a high-performance alternative for large-scale projects.
- Streamlined CI/CD pipelines by integrating CircleCI, enabling automated testing, benchmarking, and deployment workflows to ensure reliability and maintainability across complex configurations.

DTCC Lab-The Future of Coding with AI (Vertically Integrated Project at Tandon)

September 2023 - May 2024

(Prompt Engineering, Automated Code Generator, Python)

- Designed and optimized AI-driven prompts leveraging the ChatGPT 3.5/4 API to elevate Python code generation, with a primary focus on complex data manipulation and algorithmic problem-solving for LeetCode-style challenges.
- Achieved a benchmark with over 80% of structured prompts under 75 lines executing flawlessly enhancing LLM precision and efficiency in automated coding workflows.

**ContextWIN: Whittle Index Based Mixture-of-Experts Neural Model For Restless Bandits With Contextual Information
Via Deep RL (<https://arxiv.org/abs/2410.09781>)**

September 2023 - December 2023

(Theoretical Computer Science, Probability, Recommendation system, Reinforcement Learning, Python)

- Developed ContextWIN, an advanced architecture enhancing Neural Whittle Index Network for Restless Multi-Armed Bandit problems. Utilized reinforcement learning and a mixture of experts to improve decision-making in dynamic environments, especially recommendation systems.
- Proved the theoretical robustness of both NeurWIN and ContextWIN models, laying the groundwork for future applications in complex decision-making with contextual data.