Pengze Ai

+1 647 896 3206 | pengze.ai@mail.utoronto.ca | Toronto

EDUCATION

University of Toronto Sep 2023 - Jun 2026

Computer Science + Statistics Double Major Bachelor Toronto

- GPA: 3.81/4.0
- Main courses: python (96), Math Expr&Rsng for Cs (100), Computer Organization (90), Neural Networks and Deep Learning, Algorithm Design & Analysis, Time Series Analysis, Numerical Methods
- Dean's List Scholar | Innis College Exceptional Achievement Award(2024)
- Meritorious Winner of Interdisciplinary Contest In Modeling (2021)

PROJECT EXPERIENCE

Graph Neural Networks research Jan 2025 - Present

- LLM-Enhanced Cross-Domain Graph Neural Network Research
  - Investigated **cross-domain generalization** challenges in heterogeneous Graph Neural Networks; identified semantic misalignment issues across domains (e.g., molecular vs. social graphs) as the core bottleneck.
  - Proposed a low-cost **Zero-shot** GNN generalization framework by integrating Large Language Models (**LLMs**) into graph learning pipelines to provide **semantic priors** for node/edge alignment.
  - Conducted a comprehensive **literature review** on LLM-TTT, GraphCL, GraphPrompt, and zero-shot graph learning; synthesized methodological gaps and defined research directions for LLM-enhanced GNNs.
  - Designed the conceptual framework integrating **LLM-based** text descriptions with GNN embeddings to enable **domain transfer** without target-domain labels.
  - Preliminary experiments show improved feature transfer performance in unlabeled target domains, validating the feasibility of **LLM-driven semantic alignment**.
- Drug Trafficker Social Network Graph Construction & Analysis
  - Investigated challenges in illicit drug–related social network analysis, including **high expert annotation cost** and **semantic misalignment** across multimodal data in **heterogeneous graph models**.
  - Designed and built a large-scale heterogeneous social graph dataset with **7,000+ users, 45,000+ posts, and 500+ keywords**, integrating user relations, tweet content, hashtags, and interaction behaviors.
  - Developed custom **Playwright**-based crawlers to collect multi-layer social data, reducing data acquisition cost compared with Twitter API and overcoming API rate and access limitations.
  - Designed a **LLM-assisted semi-automatic labeling workflow**.
  - Constructed a bidirectional edge-building mechanism to **address follower/following visibility limits**, enriching network completeness using reply and retweet edges.
  - Conducted baseline experiments with GCN, GAT, HAN, HetGNN, HGAT, evaluating node classification and relation prediction performance and **highlighting limitations** of both LLM-based and traditional GNN approaches under imbalance and semantic complexity.
  - Findings indicate that conventional heterogeneous GNNs show limited expressiveness on this domain; currently exploring model improvements based on the constructed dataset.

PROFESSIONAL EXPERIENCE

HuleTech (Prof. Yijie Peng's Group, Peking University) May 2025 - Aug 2025

Algorithm Engineer Beijing

- End-to-End Liquor Cabinet Bottle Recognition (Patent Filed)
  - Contributed to a patented **end-to-end** liquor cabinet bottle recognition system designed for real-world commercial

environments with low light, noise, reflections, and occlusions. Unlike traditional label-based recognition methods, this system enables **label-free, batch identification** of all bottles within a cabinet.

- Built the high-quality recognition module on top of an existing YOLO-based bottle detector; responsible for designing algorithms to classify cropped bottle images under **weak lighting** and **cluttered backgrounds**.
- Implemented **ArcFace**-based high-dimensional embedding space, achieving **tight intra-class clustering** and **clear inter-class separation** to enhance recognition robustness.
- Developed a **feature-based incremental learning mechanism**, enabling recognition of new bottle categories by simply inserting a few embedding samples, **without retraining** the full model.
- Designed and deployed an **open-set recognition** pipeline to prevent unknown bottle types from being misclassified as known categories, improving system reliability in real-world deployment.
- Addressed the **key engineering challenge** of "closed-set vs. open-set" trade-off; despite closed-set models performing slightly better on known classes, open-set recognition was selected after evaluating product requirements and commercial constraints.
- Evaluated multiple SOTA baselines (e.g., **Proser**) alongside Softmax + ArcFace, achieving stable and high accuracy with limited training data and computational resources.

- AI **Agent Platform Development** (Dify-Based Secondary Development)
  - Participated in an i**ndustry–academia collaborative project** between **Peking University (Prof. Yijie Peng's team)** and **Kingming Machinery** to support the development of an **enterprise-level AI agent platform**.
  - Assisted in optimizing the platform's underlying architecture and contributed to the **setup of the local development environment**, including Docker Compose deployment of frontend, backend, database, and Redis modules.
  - Contributed to designing a **secure deployment strategy** where frontend/backends are co-located on one server and communicate with DB/Redis via intranet, improving communication security and reducing operational overhead.
  - Assisted in redesigning the platform's **permission system**, separating "Console Roles" and "Client Roles," and implementing a role-centric permission model to **simplify authorization logic** and improve maintainability; Helped define a role–group inheritance mechanism that avoids user-level permission binding and reduces configuration redundancy.
  - AI Agent Workflow Development: Participated in the development of two core AI agent modules:
    (1) **Contract Clause & Sensitive Content Auditing Agent**
    (2) **Enterprise Daily/Weekly Report Generation Agent**
  - Contributed to building multi-node workflow structures, including node allocation, boundary conditions, context planning, and model routing.
  - Assisted in designing a custom conversation history and **context management** mechanism to **address Dify's limitation** where different LLM nodes cannot share memory in multi-step workflows.
  - Helped build local structured variable storage to **retain historical analysis results**, **reducing** reliance on **raw data** and **optimizing token usage**.
  - Contributed to implementing **dynamic context-length computation** and adaptive truncation to **prevent token overflow** and maintain consistency across long LLM interactions.

Tencent May 2023 - Aug 2023

Backend developer LIGHTSPEED STUDIOS Shenzhen

- Developed and optimized internal tools for test engineer teams, improving **workflow automation** and operational efficiency.
- Streamlined the file upload module by enabling front-end direct interaction with storage containers via **RESTful APIs,** significantly **reducing** back-end **data load.**
- Contributed the front-end development (Vue 3.0) of a large language model **(LLM) evaluation** platform, enabling administrators to **design** custom **questionnaires, manage** evaluation workflows, and **monitor** model performance.
- Integrated **real-time** answer **generation** interfaces and implemented **auto-scoring** and **result aggregation** modules, supporting efficient comparison and analysis of multiple LLMs under diverse task scenarios.

ADDITIONAL INFORMATION

- **Skills:** Java, Python, Latex, PyTorch, Docker, Git, RESTful API