Nan An

ann5@mcmaster.ca • +1 (365) 883-4882 • annan.eu.org

SKILLS

- **Programming Languages:** C/C++, Python, Haskell, SQL, JavaScript
- **Software and Libraries:** Numpy, Pandas, PyTorch, Scikit-Learn, PowerBI
- **DevOps & Systems:** Docker, Git, Linux, MacOS, Windows, Azure
- **Languages:** English (Fluent), Mandarin (Fluent)

EXPERIENCES

Data Scientist Intern Sep 2024 – Present

Health Science Data Branch, Ministry of Health

- Analyzed and managed large-scale Ontario health datasets (>500,000 records).
- Wrote SQL queries to extract insights from relational databases and built dashboards.
- Maintained the ETL pipeline, integrated LLMs for text data cleaning, reducing costs by 90%.
- Built and deployed chatbots using RAG and SQL agents on Azure to enhance data accessibility.
- Implemented predictive models (Random Forest, XGBoost, ARIMA) for healthcare analytics.

Software Developer (Research Assistant) Sep 2023 – Aug 2024

Centre for Software Certification, McMaster University

- Developed and implemented experimental garbage collectors with reduced memory footprint.
- Built unit tests for existing tools and traced memory leaks using Valgrind.
- Maintained build servers and CI/CD, and contributed to research on software verification.

Teaching Assistant Sep 2021 – Aug 2024

Shanghaitech University (2021 - 2023), McMaster University (2023 - 2024)

- Led tutorials and office hours for courses including Probability and Statistics, Calculus, Data Mining, and AI.

EDUCATION

Master of Science in Computer Science Sep 2023 - Present

McMaster University, Hamilton, ON, Canada

Bachelor of Science in Physics (CGPA 3.8/4.0) Sep 2019 – Jun 2023

Shanghaitech University, Shanghai, China

ADDITIONAL EXPERIENCE

Data Scientist (Volunteer) Sep 2024 - Present

Data & Product Team, Front Row Ventures

- Conducted data analysis and built predictive models to support investment decisions.
- Developed dashboards and visualizations to track startup performance and market trends.

- Implemented machine learning techniques to identify promising early-stage companies.

Open-Source Contributor (Neo4j, SGLang, etc.)

- Improved documentation, submitted new feature PRs, and collaborated with other members of the open-source community.

PROJECTS

Data Science & AI Projects

Fraud Detection Graph Neural Network (PyTorch, scikit-learn)

- Developed a graph neural network for fraud detection, achieving >90% accuracy on two large real-world datasets.
- Tackled the class imbalance problem by developing balanced samplers.
- Designed a new graph aggregation method to achieve higher efficiency and accuracy.

Llama (Pytorch)

- Re-implemented Meta's Llama model based on original research papers.
- Developed key components such as tokenization, RoPE embeddings, and multi-head attention.
- Experimented with the MoE (Mixture of Experts) architecture to improve model efficiency.

Software Engineering Projects

Personal Website & Tracking System (JavaScript, Nginx, C, SQL)

- Built a scalable web analytics system using serverless functions, KV cache, and databases.
- Optimized database performance with caching and batched writes.
- Implemented session tracking and bot filtering for accurate user insights.
- Developed an interactive 3D globe with Three.js to visualize analytics data.

Ray Tracing Engine (C++)

- Developed a lightweight ray tracing engine from scratch.
- Implemented support for multiple materials, including diffuse, reflective, and refractive surfaces.
- Optimized the engine with multithreading and SIMD instructions.