Daoming Liu

ldm924940564@icloud.com | 236-869-9179 | www.linkedin.com/in/daoming-liu

EDUCATION

University of British Columbia **Vancouver, BC**
M.S. in Data Science: Computational Linguistics *Sept 2024- Excepted July 2025*
- **CGPA:** 4.30/4.33 | **Relevant Coursework**: LLM Agent Workflow, Multimodal LLM, SQL Database Systems and Cloud Computing

University of Washington **Seattle, WA**
B.S. in Statistics: Data Science *2020 - 2024*
- **CGPA:** 3.73/4.00 | **Relevant Coursework:** Probability and Statistics, Geospatial Data Analysis, Supervised and Unsupervised Learning

SKILLS & TOOLS

- **Programming Languages:** Python (PyTorch, Scikit-Learn, transformers, LangChain, numpy), TypeScript, SQL, R
- **Applications:** Hugging Face, LangChain, FastAPI, Docker, GCP, AWS, GitHub, Weights & Biases, MySQL, MongoDB, Tableau, Power BI, Jupyter, LangSmith, Wordsmith, Postman, Notion, Linear, Next.js

PROFESSIONAL EXPERIENCE

BC Children's Hospital **Vancouver, BC**
NLP Engineer Intern *May 2025 – Jun 2025*
- Developed an end-to-end Retrieval-Augmented Generation (RAG) pipeline to answer health-related queries, modularizing data preprocessing, retrieval, and generation components for evaluation and experimental iteration.
- Designed a configurable data preprocessing pipeline supporting multiple chunking strategies and embedding models, enabling custom data transformations for semantic search and rapid experimentation.
- Developed and tested REST API endpoints using FastAPI and Postman, enabling backend-to-frontend integration and supporting RAG-based query workflows.
- Collaborated with cross-functional teams and stakeholders to gather requirements and design technical solutions, ensuring alignment between RAG system capabilities and business needs.

Adauris **Vancouver, BC**
Software Engineer Intern *Jan 2025 – Apr 2025*
- Designed and refined structured prompts for AI article-to-podcast products, improving output alignment with target style and structure, and reducing token costs by 50%.
- Built RESTful endpoints (FastAPI, Pydantic) to integrate backend AI services with frontend interfaces, tested endpoints using Postman to ensure reliable API workflows for podcast generation pipelines.
- Collaborated with cross-functional teams in agile sprints (Linear), aligning product goals with client expectations and delivering MVPs that incorporated early stakeholder feedback.
- Contributed to CI/CD pipelines (Git, GitHub) by conducting thorough code reviews and standardized documentation, enabling fast and reliable AI application deployments.

University of British Columbia **Vancouver, BC**
NLP Research Assistant *2024 - 2025*
- Developed Docker-based containerized environments for research servers, streamlining dependency management and GPU configuration to accelerate LLM fine-tuning workflows and improve reproducibility.
- Fine-tuned a vision-language model for sign-language understanding on large-scale datasets, implementing a contrastive-learning strategy and achieving a 5× BLEU score improvement over the baseline.
- Customized and extended the SentenceTransformer IR evaluator for multimodal contrastive learning, enabling direct evaluation of cross-modal embeddings and accelerating model iteration cycles.

Seer Big Data **Shenzhen, China**
Data Analyst Intern *Aug 2024 – Jan 2025*
- Conducted geospatial analysis on 10,000+ mental health records and visualized geospatial clusters, enabling data-driven decisions for cross-functional teams.
- Built a semi-automated PDF digitization workflow using chain-of-thought LLM prompting and manual validation, achieving 90%+ extraction accuracy.
- Analyzed multi-year student PE data using clustering and regression models, uncovering key patterns and sharing insights via interactive dashboards.

Hirebeat **Jersey City, NJ**
Data Analyst Intern *2024 Summer*
- Built a resume chatbot pipeline with RAG architecture (LangChain), improving data management scalability and increasing response accuracy over traditional prompt engineering methods.
- Researched and evaluated emerging tools to enhance LLM agent workflows, investigating methodologies for building proprietary financial analysis LLM pipelines.

UW Center for Environmental Health Equity   **Seattle, USA**
Data Analysist   *2023-2024*
- Retrieved wildfire-related large disaster data through public APIs; used R to perform data preprocessing, data visualization, and geospatial data transformation, providing teams with cleaned data for immediate analysis
- Led hypothesis-driven EDA and statistical testing for geospatial health datasets using R and Tableau, surfacing key features to support academic insights and clustering analyses
- Geo-referenced 50+ African American community maps from the 1860s to modern-day geographic coordinates; utilized R, ArcGIS, and QGIS to create geospatial shapefiles for data visualization and demographic analysis
- Conducted geospatial and statistical analyses to evaluate the health impacts of unconventional oil and gas developments in Texas, contributing as a co-author to a longitudinal study examining the psychological well-being of local residents