Nicholas Wall

Email : walln@hey.com

https://www.walln.dev

Education

Southern Methodist University — Dallas, TX

- *Master of Science in Computer Science - Lyle Discovery Scholar* — *Jan. 2023 – Dec. 2023*
  Machine Learning and Artificial Intelligence Specialization

Southern Methodist University — Dallas, TX

- *Bachelor of Arts in Computer Science - SMU Distinguished Scholar* — *Aug. 2019 – Dec. 2022*
  Machine Learning and Artificial Intelligence Specialization

Experience

- **IBM** — Dallas, TX
  Software Development Intern — *May 2022 - August 2022*
  - **Applied Research**: Collaborated with research teams to develop systems using custom state-of-the-art models for knowledge extraction and conversational AI. This involved putting research artifacts into practice, developing production-grade software, and performing experiments.
  - **Knowledge Extraction and Visualization**: Designed a framework for working with directed cyclic graphs in Python, converting unstructured data into knowledge bases. Developed a React-based tool for visualizing and manipulating these knowledge graphs with thousands of nodes, enhancing data interpretation for 2 high-priority projects.
  - **Training Models**: Trained BERT-based models for various natural language processing tasks, including co-reference resolution, named entity recognition, and relationship extraction. These models performed on par with state-of-the-art methods on internal benchmarks.
  - **Dataset Curating**: Created and curated datasets to improve machine learning model performance through detailed annotations and cleaning. This improved word error rate performance by more than 10% for Automatic Speech Recognition (ASR) applications.

- **IBM** — Dallas, TX
  Application Development Intern — *May 2021 - January 2022*
  - **Generative Chat**: Created a chat-based application to streamline complex workflows for thousands of users. The chat system integrated with user-uploaded images using computer vision systems, leveraged retrieval augmented generation, and had a mobile application frontend.
  - **Realtime Dashboard**: Built a react-based real-time dashboard to monitor generative chat applications, reducing monitoring latency by a factor of 10 and enabling non-technical users to perform administrative tasks. Utilized MongoDB and Node.js to create the monitoring service.
  - **Speech-to-Text**: Trained and deployed Speech-to-Text machine learning models leveraging technology such as wav2vec. Optimized the models for low-quality phone calls, achieving word error rates under 15, comparable to state-of-the-art.
  - **AI Education**: Guided a team of more than 10 colleagues with less technical experience in applying AI for Natural Language Processing through dataset construction and developing models. This enabled a better understanding of the user experience as well as higher-quality datasets for our models.
  - **Backend Development**: Added new features and integrated AI systems into event-driven microservice-based applications. This involved designing for Domain-Drive-Design patterns and distributed systems. Projects were deployed onto Kubernetes clusters and worked with containerization technology on our public cloud.
  - **Optimization Algorithm**: Implemented a demo application using mathematical and constraint programming methods as well as high-performance computing for decision optimization models.

Projects

- **Hooper**: Open source full-stack LLM-based chat experience for basketball news and analytics with generative UI to render dynamic HTML based on the context. Includes fine-tuned Llama 3, Next.js, and serverless architecture.
- **Scratch**: High-quality annotated implementations of machine learning papers in Python with JAX, such as Mamba, Swin Transformers, Llama, attention variants, and accelerator kernels.

Programming Skills

- **Languages**: Python, Javascript, C++, SQL, Rust **Technologies**: AWS, JAX, PyTorch, React, CUDA