Willis Guo

willisg@cs.carnegie.edu | +1 (878) 834-9154 | willisguo.vercel.app | linkedin.com/in/willisguo | github.com/willisguo14

EDUCATION

MS Machine Learning — *Dec. 2025*
Carnegie Mellon University — *Pittsburgh, PA*

BASc Machine Intelligence — *Apr. 2024*
University of Toronto — *Toronto, CAN*

WORK EXPERIENCE

Software Development Engineer Intern — *June 2024 – Aug. 2024*
Amazon Web Services (AWS) — *Vancouver, CAN*

- Developed infrastructure and tools for analyzing PBs of traffic data and performed feature engineering for anomaly detection, increasing detection accuracy of distributed denial-of-service (DDoS) attacks for AWS resources by 4%.
- Designed and implemented a full-stack application for visualizing traffic data and interpreting the DDoS detection system, decreasing the time to diagnose DDoS detection errors by 50%.

Software Development Engineer Intern — *May 2023 – Jul. 2023*
Amazon Web Services (AWS) — *Vancouver, CAN*

- Developed APIs to decouple two internal services responsible for collecting AWS customer payments, reducing SEV2 counts by 60% and deployment times by 50%.
- Created the low-level design, led design meetings, implemented the API infrastructure and logic, performed integration and load tests, and created service alarms and dashboards for operational readiness.

Machine Learning Engineer — *Sep. 2021 – Apr. 2022*
aUToronto — *Toronto, CAN*

- Finetuned YOLO models for traffic light detection, improving accuracy by 10% and subsequently winning 1st overall and in the traffic light challenge at the 2022 invitational SAE AutoDrive Challenge II.
- Led the implementation of an end-to-end machine learning pipeline for speed limit sign detection, achieving 89% accuracy.

RESEARCH EXPERIENCE

Multimodal Foundation Models for Long-Form Video Understanding — *Sep. 2024 – Present*
Carnegie Mellon University (Ruslan Salakhutdinov) — *Pittsburgh, PA*

- Improving instruction-tuning and inference efficiency in vision-language models (VLMs) for long-form video understanding.
- Developed a novel zero-shot VLM method for long-form video question answering leveraging video generation models as world models, improving accuracy by 2% while reducing selected video keyframes by 5x.

Large Language Model (LLM) Reasoning & Knowledge Graph Question Answering (KGQA) — *Sep. 2023 – Apr. 2024*
University of Toronto (Scott Sanner) — *Toronto, CAN*

- Designed a novel LLM-based KGQA method with planning, active retrieval augmentation, and grounded reasoning, reducing LLM hallucinations by 79% and outperforming existing KGQA methods by 8%.
- Invented a novel neuro-symbolic algorithm for logical reasoning with LLMs that is debuggable and repairable, improving LLM commonsense reasoning accuracy by 13%.

Deep Learning for Code Vulnerability Detection — *May 2021 – Aug. 2021*
University of Toronto (Shurui Zhou) — *Toronto, CAN*

- Created a real-world code vulnerability detection dataset with 2,000 training examples by designing a novel automated pipeline for extracting code vulnerability fixes and scraping open-source projects.
- Implemented and re-trained deep learning-based baseline methods, improving detection accuracy by >5%.
- Received the Dean's Undergraduate Research Pivot Fellowship ($8,000) out of 100+ students based on research impact.

PUBLICATIONS

Active Perception for Efficient Inference-Time Long-Form Video Understanding in Vision-Language Models
Martin Ma, Ankit Gupta, **Willis Guo**, Aditya Agrawal, Paul Liang, Russ Salakhutdinov, Louis-Philippe Morency. *In submission*

Verifiable, Debuggable, and Repairable Commonsense Logical Reasoning via LLM-based Theory Resolution
Armin Toroghi, **Willis Guo**, Ali Pesaranghader, Scott Sanner. *EMNLP 2024*

Right for Right Reasons: Large Language Models for Verifiable Commonsense Knowledge Graph Question Answering
Armin Toroghi, **Willis Guo**, Mohammad Mahdi Abdollah Pour, Scott Sanner. *EMNLP 2024*

SKILLS

Languages	Python, Java, Scala, C, SQL, JavaScript, TypeScript, MATLAB
Machine Learning	PyTorch, TensorFlow, Hugging Face, Apache Spark
Software Engineering	PostgreSQL, AWS, ZIO, React, Next.js, Node.js, Docker