Shengxiang Sun

+1 4169026176 ◇ owen.sun@mail.utoronto.ca **Google Scholar** ◇ **GitHub** ◇ **LinkedIn** ◇ **X**

EDUCATION

University of Toronto *Sep 2022 – (expected) Apr 2026*

Honours Bachelors of Science in Computer Science

- **GPA: 3.83/4.00**

RESEARCH EXPERIENCE

- **Visiting Research Assistant, Stanford University** *May 2025 – Present*
  Topic: Long-Horizon Contact-Rich Manipulation by Learning from RGB Videos Advisor: **Dr. Weiyu Liu**
- **Visiting Research Assistant, National University of Singapore** *Oct 2024 – May 2025*
  Topic: Autonomous Furniture Assembly by Reading IKEA Manuals with VLMs Advisor: **Prof. Lin Shao**
- **Full-Time Research Assistant, University of Toronto** *May 2024 – Oct 2024*
  Topic: Failure Detection in Vision Language Action Models Advisor: **Prof. Florian Shkurti**

PUBLICATIONS & PREPRINTS

1 [**NeurIPS 2025 (Under Review)**] Qiao Gu, Yuanliang Ju, **Shengxiang Sun**, Igor Gilitschenski, Haruki Nishimura, Masha Itkina, Florian Shkurti, "SAFE: Scalable Failure Estimation for Vision-Language-Action Models"

2 [**RSS 2025**] Chenrui Tie*, **Shengxiang Sun***, Jinxuan Zhu, Yiwei Liu, Yue Hu, Jingxiang Guo, Haonan Chen, Ruihai Wu, Junting Chen, Lin Shao, "Manual2Skill: Learning to Read Manuals and Acquire Robotic Skills for Furniture Assembly Using Vision-Language Models" **[Paper]** **[Website]**

RESEARCH PROJECTS

Long-Horizon Contact-Rich Manipulation by Learning from RGB Videos (Ongoing)
Advisor: Dr. Weiyu Liu, Postdoctoral Scholar, Stanford, CS *May 2025 – Present*

- Evaluating Nvidia FoundationPose on human-performed furniture assembly videos recorded via smartphone.
- Leveraging Nvidia IsaacGym to test FurnitureBench's robotic manipulation performance.

Autonomous Furniture Assembly by Reading IKEA Manuals with VLMs
Advisor: Prof. Lin Shao, Assistant Professor, NUS, CS *Oct 2024 – May 2025*

- Generated a synthetic dataset of over 10,000 furniture parts using a novel automated pipeline in Blender to simulate realistic assembly scenes. Fine-tuned QWEN-2.5B using LoRA to predict furniture part connections.
- Employed vision-language models to generate high-level furniture assembly plans from IKEA manuals, achieving generalization across diverse furniture types and exceeding previous baselines by over 300%.

Failure Detection in Vision Language Action Models
Advisor: Prof. Florian Shkurti, Assistant Professor, University of Toronto, CS *May 2024 – Oct 2024*

- Developed a pipeline using PyTorch and SimplerEnv to fine-tune VLAs (eg. Open-Pi-Zero) on mixed datasets from Open-X-Embodiment, achieving generalization across diverse simulation environments.
- Estimated uncertainty through Nvidia IsaacSim in VLM-based robotic models such as ReKep, producing three diverse uncertainty quantification methods and 50+ samples.
- Tested 2D and 3D part segmentation models on High-Performance Computing Clusters with SLURM.

RESEARCH INTERESTS

My research interests span **Robotics and 3D Computer Vision**, with a focus on **generalizable and safe robot manipulation**. I am particularly passionate about developing algorithms that enable robots to perform complex, long-horizon tasks through simple human instructions, such as "cook the egg."

SCHOLARSHIPS & AWARDS

- **2024 Summer NSERC Math & Computer Science Research Award** (CA$8,000)
- **2022-2024 General In-Course Scholarship** (For maintaining a cummulative GPA of at least 3.7/4.0) (CA$9,000)
- **2023-2024 Dean List Scholar**

WORK EXPERIENCE

Loblaw Digital **Toronto (CA)**
Machine Learning Engineer Co-op - Generative AI Team *Jan 2024 – Apr 2024*

- Enhanced an automated email reply system using Google's Gemini Pro, Python, Docker, CI/CD, Few-Shot and Chain-of-Thought prompt engineering, which resulted in over 3400 correctly automated email replies per week.
- Developed an end-to-end machine learning pipeline for enhanced shopping experience, with OpenAI's GPT-4 Vision, Python, Pandas, SQL, Apache Airflow DAGs, and Google Cloud Platform, which automatically generated product descriptions for 154,286 products sold at Loblaws, Shoppers Drug Mart, and Joe Fresh

New H3C Technologies **Beijing (CN)**
Machine Learning Research Intern *Jul 2023 – Aug 2023*

- Designed training & testing pipelines of Llama2, Dreambooth, InstructPix2Pix on MobaXterm and WebUI, by using PyTorch and HuggingFace, which doubled the team's testing data outputs
- Enabled automated downloads of Python dependencies with bash scripting, reducing installation steps by 40%

EXTRACURRICULAR EXPERIENCE

GenAI Genesis
Hackathon Winner - InterView Team *github.com/InterView* *March 30th, 2024 - March 31st, 2024*

- **First-Place: Best AI in Safety & Responsible AI**
- Developed an AI interview helper for junior HR and hiring managers, using Google's Gemini Pro, LangChain for RAG, and Speech-To-Text API, which achieved 80% accuracy in real-time detection of biased interview questions

HackTheValley - 8
Hackathon Participant - QuickScan Team *github.com/QuickScan* *October 13th, 2023 - October 15th, 2023*

- Trained a CNN and RNN model using TensorFlow, CUDA, cuDNN, and CTC loss function, achieving 83% accuracy in predicting and converting handwritten text to digital text

PROGRAMMING SKILLS & LANGUAGE SKILLS

Proficient Python, LaTeX, HTML **Familiar** PyTorch, Linux, C, Java, Git

Chinese (Native), English (Fluent), French (Intermediate)