Alan (Jia Lin) Yuan

alan.yuan.jly@gmail.com | linkedin.com/in/jalnyn | github.com/jalnyn | scholar.google.com/

Work Experience

Instacart Feb 2025 – Present
Software Developer *Toronto, Ontario*

Unilever April 2024 – Dec 2024
Machine Learning Engineer - Intern *Toronto, Ontario*

- Train **times-series transfomer** models to **forecast sales** with **PyTorch** increasing accuracy by 5%.
- **Optimize promotion times and prices** with augmented genetic algorithms **GNGSA**.
- Utilized novel algorithm **GNGSA** increasing convergence iterations 70% and hypervolume AUC by 50%.
- Implemented multi-thread processing to increase data pipline speeds by **50%**.

Amazon Jun 2023 – Aug 2023
Software Developer - Intern *Toronto, Ontario*

- Designed a **precompute** layer using **Apache Spark** to increase recommendation speed by **99%** from 150ms
- Built data pipelines to ingest updated data and retrain models in a given time frame
- Created automated data analysis tool to ensure predictions are above **75%** accuracy

Amazon May 2022 – Aug 2022
Software Developer - Intern *Vancouver, British Columbia*

- Engineered a modular microservice in **Java** to send notifications to customer of cashback on select products
- Utilize **AWS** webservices such as **Lambda**, **SQS** and **SNS** to ensure scalability of the notification system
- Integrated service into data pipeline utilizing Amazon internal language **Datapath**

Intel May 2021 – May 2022
Software Engineer - Intern *Toronto, Ontario*

- Developed support software to generate 4000+ of completely random test-cases for edge-case testing

Centivizer Apr 2020 - Sep 2020
Software Developer - Part-time *Toronto, Ontario*

Education

University of Toronto 4.0 cGPA
MSc in Applied Computing (A.I.) *Sept 2023 - Dec 2024*

University of Toronto 3.84 cGPA
HBSc Computer Science Specialist, Major in Mathematics *Sep. 2018 – May 2023*

Projects

WIDE: Websocket-IDE | Link: GitHub Sep 2024 – present

- Architecting and developing a websocket-based IDE using **Rust** and **React** frontend (58 GitHub stars)

Decomposed Face Generation | Link: Github Apr 2024 – May 2024

- Decompose faces into **id, pose, emotion embeddings** and utilize **StyleGAN** to re-generate faces.

PAIR Lab — multiple projects: **RePlan** | **Orbit** Sep 2021 – Aug 2024

- Use of **MPC**, **MultiModal LLM** in completing high level robot tasks prompted by text | **arXiv:2401.04157**
- Contributed to **Orbit**, a robot learning framework built on **NVIDIA Isaacsim**. Published in *RAL* | **project-site**
- Designed and built **GPU** parallelized state systems with low overhead allowing a **4x** speedup over the cpu solution

MultiModal AI Story teller | private repo Jul 2023 – Jul 2024

- Managed **auto-scaling GPU** resources with **Kubernetes** saving up to 90% on **AWS** costs.
- Implemented **Multimodality** using **LLMs** and **Latent Diffusion Models** to build a interactive story teller.
- Utilized NLP techniques to summarize context to reduce context size, reducing inference time by up to **10%**.

CaNetDa: Deep Learning for GeoGuesser in Canada | Link: GitHub Jan 2021 – Apr 2021

- Mined dataset and trained an ensemble of **Computer Vision** models: **ResNet**, **EfficientNet** and **Vision Transformer** resulting in a **47%** improvment over random agent in predicting province of image in Canada
- Scraped **Google Street View** images and used **Google Maps API** to preprocess images for training

Technical Skills

Tech: Deep Neural Networks, Large Language Models, Latent Diffusion Models, Python, C++, JavaScript, Java, Rust
Tools: Huggingface, PyTorch, Git, React, Node.js, MongoDB, SQL, Numpy, GraphQL, Robotics, Vim