Alan (Jia Lin) Yuan

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

WORK EXPERIENCE

Unilever April 2024 – Present

Machine Learning Engineer - Intern

Toronto, Ontario

- Train times-series transformer models to forecast sales in various regions with PyTorch.
- Optimize sales prices and promotion weeks with genetic algorithms utilizing NGSAII.
- Increased model accuracy by 15% by modifying model architecture and finetuning hyperparameters.
- Implemented multi-thread processing to increase data pipline speeds by 150%.

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
- \bullet 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
- \bullet Optimized support tool's Ram templates to reduce false positives and failing cases by around 70%

Centivizer Apr 2020 - Sep 2020

Software Developer - Part-time

Toronto. Ontario

- Designed backend application using Node.JS and SimplePeer to facilitate real-time video streams
- Integrated video system with main **React** site and **NoSQL** database

EDUCATION

University of Toronto 4.0 cGPA

MSc in Applied Computing

Sept 2023 - Dec 2024

University of Toronto

 $3.84~\mathrm{cGPA}$

HBSc Computer Science Specialist, Major in Mathematics

Sep. 2018 - May 2023

PROJECTS

MultiModal AI Story teller | private repo

Jul 2023 – Jul 2024

- Built interactive **socket-based** streaming service serving up to 200+ users a day.
- 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%.

PAIR Lab — multiple projects: RePlan | Orbit

Sep 2021 – Apr 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**
- Worked on meta-controllers to assist multi-policy agents to plan long horizon tasks using sparse rewards
- Designed and built **GPU** parallelized state systems with low overhead allowing a **4x** speedup over the cpu solution

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% improvement 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

Tron UDP Multiplayer | Link: GitHub

Sep 2019 – Dec 2019

- Created a four player game for local networks using the UDP network protocol and C++
- Utilize **epoll** for both client and server to monitor the socket as well as the timer (server) and stdin (client)

BF-Interpreter | Link: GitHub

Mar 2018 - Nov 2018

• Built a BF shell that runs all example BF programs found on Wikipedia in C 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