

# Yixin (Isshin) Tian

Last Updated on 3rd July 2024

Canadian Citizen | yixino829@outlook.com | +1 (431) 998 2086 | Toronto, ON, Canada

## EDUCATION

### UNIVERSITY OF TORONTO

BASc. IN COMPUTER ENGINEERING  
Sep 2018 - Jun 2023 | Toronto, ON  
Minor: Artificial Intelligence  
Cumulative GPA: 3.67 / 4.0

## LINKS

Blog: // [Yixin's portfolio](#)  
GitHub:// [yixin0829](#)  
LinkedIn:// [yixintian](#)  
Strava:// [yixinruns](#)

## COURSEWORK

### ONLINE

Graph ML (Stanford CS224)  
Intro to Machine Learning

### UNDERGRADUATE

Matrix Algebra  
Convex Optimization  
Natural Language Processing  
Database & Relational Algebra  
Deep Learning (PyTorch) + Proj.  
Operating System  
Distributed Systems + Proj.  
Intro to Machine Learning  
Statistics & Probability

## SKILLS

### DATA SCIENCE & AI

GENAI:  
LangChain • LangSmith • Ollama  
Chroma • Weights & Biases

### DATA TOOLS:

Spark • SQL Tableau • Airflow  
Weights & Biases • MLFlow • DVC

### PYTHON LIBRARIES:

PySpark • NumPy • pandas • polars  
Matplotlib • Seaborn • beautifulsoup4  
PyTorch • scikit-learn • SciPy  
NetworkX • PyG • boto3

### SOFTWARE ENGINEERING

PROFICIENT:  
Python • SQL • C • C++ • JavaScript  
TypeScript • Vim •  $\LaTeX$  • Git

### FAMILIAR:

React • Angular MongoDB • S3 •  
Snowflake • OpenShift • Docker •  
AWS • GCP • Java

## EXPERIENCE

### RBC | DATA SCIENTIST

July 2022 - Present | Toronto, ON

- Climate emission modeling for RBC's agriculture carbon emission target setting
- Designed and built event-driven feature preparation pipeline processing external and internal data from 30+ tables stored in **S3**, **Snowflake**, **Hive** and **MongoDB**

### RBC AMPLIFY | DATA SCIENTIST

May 2022 - Aug 2022 | Toronto, ON

- Co-created ACE Insights (both front & backend using **Angular and Python Flask REST API**), a product for RBC to manage its agricultural financed emission
- Designed and built database schema using **PostgreSQL** and worked with spatial-temporal climate data using **spatial SQL, PostGIS and geopy**
- Designed and built backend logic for **agricultural carbon estimation model** using **Python (Psychog2, Numpy, pandas, Flask-RESTX)**
- Pitched to executive audiences** and won the Industry Disruptor Award (\$20,000)

### KIJJI | DATA ANALYST INTERN

May 2021 - Apr 2022 | Toronto, ON

- Collaborated **cross-functionally** with Product, Software Development, etc. by providing statistical analysis on large datasets including **A/B testing, power analysis, descriptive analysis with Python and R, BI reporting with Tableau**
- Led App Rebuild Dashboard Project for monitoring the beta app launch performance; designed and implemented the **data architecture** using **Spark SQL**
- Trained an **auto-regressive model** with **PyTorch** (96% accuracy) and **random forest model** (87% accuracy) to predict & monitor the health state ("growing", "stable", "declining") of thousands of local communities for Kijiji

## PROJECTS

### LEARNING PATH GENERATOR FASTAPI, GPT-4, AWS, DOCKER

A full-stack ML application leveraging **OpenAI API** to generate a learning path for any input topic. Built backend using **Python FastAPI** and deployed both frontend and backend using Vercel and **AWS Lambda** respectively.

### ED PATH PLUS FLASK-RESTFUL, D3.JS, HEROKU, DOCKER

**Mined graph data** from the course relational database and implemented an interactive **force-directed graph visualization using D3.js** (at 2:00 in demo).

### HOUSE NUMBER IMAGE IN-PAINTING PYTORCH, $\LaTeX$

Trained a **CNN autoencoder** using **PyTorch** that inpaints missing part of house number image and effectively **improves the OCR model performance to 90+%**.

## RESEARCH

### CENTER ON KNOWLEDGE GRAPHS, USC ISI | RESEARCH ASSISTANT

Jan 2024 - Present | Remote

Working on structured data extraction from semi-structured long-context mineral reports by leveraging LLM under the supervision of Prof. Craig Knoblock

### FORCOLAB, UNIVERSITY OF TORONTO | RESEARCH ASSISTANT

Dec 2022 - July 2023 | Toronto, ON

Focused on constructing tree visualizations from computational notebooks to facilitate knowledge sharing between data scientists and business stakeholders under the supervision of Prof. Shurui Zhou