

(647) 865-9619  
Toronto, ON  
kyrie807782352@gmail.com

# Ziyi(Kyrie) Xu

Data Analyst / Data Scientist

Portfolio: ziyi-xu  
GitHub: 807782352  
LinkedIn: ziyi-xu

## SKILLS

Programming Languages	Python, R, SQL, NoSQL(Redis), Java, JavaScript, C/C++, Linux OS
ML & AI Frameworks	PyTorch, TensorFlow, scikit-learn, XGBoost, LightGBM, OpenCV
Data Analysis & Visualization	Pandas, NumPy, Matplotlib, Seaborn, Excel, PowerBI
NLP & LLM Tools	GitHub Copilot, ChatGPT, Deepseek, Dify
Development Tools	Git, Jupyter, PyCharm, VScode

## EDUCATION

<b>Master of Management Analytics</b> , Rotman School of Management, University of Toronto	Aug 2025 - Jun 2026 (Expected)
GPA: 3.73 / 4.0   Vector AI Scholarship (2025)   Entrance Scholarship (2025)	
<b>Master of Engineering - Electrical &amp; Computer Eng</b> , University of Waterloo	Sept 2021 - Oct 2022
GPA: 90.50/100 (4.0/4.0)   Specialization in Software and Machine Learning	
<b>Bachelor of Engineering - Mechatronics Eng</b> , McMaster University	Sept 2017 - Jun 2021
GPA: 10.42/12 (3.7/4.0)   Dean's Honour List (2018, 2019, 2021)   McMaster Award, Level 3 (2017)	

## EXPERIENCE

<b>Marketing Analyst / BMO</b>	Jan 2026 — Jun 2026 (Expected)
Practicum - Customer Lifetime Value & Persona Segmentation Models (Ongoing)	Toronto, Canada
• Built a 12-month forward revenue model for 4.5M+ U.S. customers using <b>Lasso</b> , <b>Random Forest</b> , and <b>XGBoost</b> to support value-based targeting.	
• Constructed a 36-month customer-month analytical dataset with 30+ engineered features, addressing revenue volatility and data quality gaps.	
• Applied tenure-based segmentation and expanding-window time validation to improve prediction stability across lifecycle stages.	
• Developed persona clustering ( <b>K-Means/GMM</b> ) integrating predicted CLV outputs to identify high-return customer segments.	
• Designed a pre-campaign <b>ROI</b> simulation framework enabling data-driven campaign selection and marketing prioritization.	
<b>LLM Engineer / Nanjing Yuanquan Software Technology Co., Ltd.</b>	Feb 2025 — Jun 2025
Part-time - AI-driven Global Customer Acquisition & Growth Platform	Nanjing, China
• Collaborated with a 3-member team to prototype a <b>24/7 customer support chatbot</b> , automating inquiries and reducing manual workload.	
• Experimented with multiple <b>LLMs</b> (GPT, DeepSeek, Claude), designing flows and prompts in <b>Dify</b> and tuning parameters (e.g., embedding dimensions) to improve response relevance.	
• Integrated a <b>RAG</b> pipeline with the company's knowledge base for context-aligned domain-specific responses.	
• Implemented LLM-based <b>sentiment classification</b> within <b>Dify</b> flows to dynamically adjust response tone and style.	
• Evaluated prototypes on accuracy, latency, response quality, and error recovery, identifying configurations with deployment potential.	
<b>AI Trainer / Outlier</b>	Apr 2024 — Feb 2025
Part-time - Outlier's Language Writing Model	Remote, Canada
• Evaluated 100+ AI <b>prompt-response pairs</b> using rubrics and pairwise comparisons to assess quality and preference.	
• Scored and ranked model outputs on factual accuracy, coherence, and contextual relevance, generating high-quality training signals for <b>RLHF</b> .	
• Provided structured feedback and annotations to refine datasets, improving <b>LLM</b> alignment, safety, and output quality.	

## PROJECTS

<b>Image Captioning Model Comparison - CNN+LSTM vs ViT+Transformer</b>	Nov 2025 – Dec 2025
• Built an experimental pipeline on the MS COCO dataset 2024 to compare <b>CNN+LSTM</b> and <b>ViT+Transformer</b> image captioning architectures.	
• Evaluated models using <b>BLEU-1/2/4</b> and qualitative examples to analyze differences in caption accuracy, fluency, and error patterns.	
• Analyzed the impact of hyperparameters (e.g., embedding size, vocabulary size, teacher forcing, layer freezing) on model performance and computational efficiency.	
<b>GR8 Summer Stay - Property Recommender</b>	Aug 2025
• Collaborated with a 5-member team to build a property recommender in <b>Python</b> using <b>Sentence-BERT</b> embeddings and an <b>LLM</b> advisor.	
• Developed a CLI and <b>Streamlit</b> front end with semantic matching, <b>SQLite</b> -based user management, chat, wish list, filters, and map features.	
• Integrated an AI travel agent chat (OpenRouter <b>LLM</b> ) for secure, personalized property recommendations and travel advice.	
<b>Tripedia - Tour Sharing Website for Posting and Exchanging Travel Experiences</b>	Jan — April 2022
• Built a full-stack <b>RESTful tour web application</b> in a 6-member team with a <b>Java SpringBoot</b> backend and <b>React.js</b> frontend.	
• Integrated <b>Amazon RDS/S3</b> for secure storage, <b>Spring Data JPA</b> for database access, and <b>Google Map API</b> for location features.	
• Ensured reliability through <b>JUnit5</b> unit testing and <b>SonarQube</b> code analysis (85% coverage), and deployed on <b>AWS EC2</b> with load balancing.	

## COMPETITIONS & ACCOMPLISHMENTS

- **1st Place**, Canadian University Chess Championship 2020 (Under 1800 Section).
- **2nd Place**, Calian Case Competition: Built an AI-driven framework in **Python** using K-Means for customer segmentation and RF, LR, XGBoost, and KNN for churn prediction for a restaurant-arcade brand.
- **2nd Place**, McMaster Engineering Competition 2020: Developed a COVID-19 Tracker with an interactive visualization map.
- **Top 16**, Sumobot Competition 2018: Developed a 10x10cm autonomous robot against 64 teams.
- **Participation Award**, CIBC Case Competition: Proposed an AI-driven investing framework for Simplii under CIBC to engage Gen Z investors.
- **Participation Award**, RoboMaster 2021 Robotics Competition: Developed color and number detection algorithms using **Python OpenCV**.
- **Ongoing Participation**, Scotiabank Case Competition: developing explainable AML models to identify suspicious customers and risk drivers.