

Yunzhao (Daniel) Li

yli4333@uwo.ca | +1 647-525-6199 | Toronto | [linkedin.com/in/yunzhao-li23](https://www.linkedin.com/in/yunzhao-li23) | <https://github.com/yunzhaol>

Data Scientist | Quantitative Analyst & Researcher | Data Engineer | Machine Learning Engineer

SKILLS

- **Programming Languages:** Python, R, SQL, Java
- **ML & Visualization:** PyTorch, Scikit-learn, TensorFlow, Pandas, NumPy, Spark, Matplotlib, RESTful APIs, **GitHub Copilot**, Tableau
- **Data Engineering & Databases:** MySQL, Docker, ETL/ELT Pipelines, Azure, **Git/GitHub**, AWS, **Agile**
- **Soft Skills:** Problem Solving, Collaboration, Effective Communication, Project Management, Leadership, Presentation

EDUCATION

- Western University**, London, ON Sep 2025 – Aug 2026
- Master of Data Analytics
 - Relevant Coursework: Artificial Intelligence, Databases, Machine Learning, Unstructured Data, Reinforcement Learning
- University of Toronto**, Mississauga, ON | Annual GPA: **3.76** Sep 2020 – Aug 2025
- Honours Bachelor of Science in Statistics Specialist and Mathematics Minor
 - Relevant Coursework: Advanced Statistical Learning & Modeling, Time Series Analysis, Stochastic Processes, Linear Algebra

WORK EXPERIENCE

- Business Analyst (Internship) | Top Knowledge Co.**, Toronto, ON Jun 2025 – Aug 2025 (3 mos)
- Conducted in-depth data analysis using **VBA/Excel** to categorize customer behaviors and needs, enabling the design and implementation of targeted marketing strategies that successfully increased customer retention rates by **17%**.
 - Designed and maintained **dashboards** that streamlined customer reports and enhanced report accuracy, thereby accelerating compliance adaptation and helping management decision-making, resulting in a **15%** boost in operational effectiveness.
- Data Analyst (Internship) | AstraZeneca** (Central Marketing Department), Shanghai, CN Nov 2022 – Oct 2023 (1 yr)
- Analyzed market dynamics and competitor products using **Python/SQL/Excel**; built evidence-based visuals (**Tableau/Power BI**) for seminar briefings that cut prep time and improved forecast accuracy by **12%**.
 - Improved data reliability for weekly business reviews by diagnosing quality issues across fragmented Excel sources; cleaned, validated, and automated pipelines using **VBA, PowerQuery**, and reproducible workflows, reducing inconsistencies by **28%**
 - Enhanced marketing team **decision-making** by identifying gaps in existing workflows; designed and deployed interactive **Power BI/Tableau dashboards** to streamline communication, enabling faster, clearer stakeholder updates across campaigns.

PROJECT EXPERIENCE

- Electricity Demand Forecasting | **Python, PyTorch, Time-Series ML** Oct 2025 – Nov 2025
- Built a scalable **forecasting pipeline** for **52k+ time-series** entries, integrating preprocessing, temporal splits, and **feature engineering** (lags, rolling windows, cyclical encodings).
 - Benchmarked **Linear/Ridge, MLP, XGBoost** and **CNN** architectures using a reproducible 5-run setup; delivered a top-performing **XGBoost model** that significantly outperformed classical and naive baselines.
- OSFI Risk-Weight Mapping Engine | **SQL, Data Modelling** Sep 2025 – Oct 2025
- Designed **SQL** pipeline mapping **1,000+** bonds to OSFI risk weights using multi-agency ratings (S&P, Moody's, Fitch, DBRS); implemented ranking/CTE logic for **1–3+** rating scenarios and sovereign defaults.
 - Optimized **query design** to improve processing time by **40%**, enabling reliable risk reporting for credit-risk teams.
- LexiGO – Language Learning Desktop App | **Java, OOP, Clean Architecture, Agile** Jul 2025 – Sep 2025
- Developed a **Java Swing application** with **Clean Architecture**, enabling adaptive vocabulary learning via flashcards, spaced repetition, and gamified features (streaks, badges, leaderboards).
 - Implemented core modules and **analytics dashboards** with JSON-based persistence, ensuring robust, maintainable code through **unit and integration testing**, while collaborating in an **Agile** team using **GitHub**.
- Bayesian Analysis of WWII Bombing Target Prioritization | **R, Feature Engineering, Bayesian OLR** Mar 2024 – Apr 2024
- Processed a **178k-record** wartime THOR dataset (filtered to **64k** German missions) and **engineered analytical features** using tidyverse, producing a fully reproducible **data pipeline**.
 - Constructed a **Bayesian ordered logistic regression** model (rstanarm) with complete **MCMC diagnostics** and **posterior predictive checks** to quantify drivers of Allied target prioritization; project selected for **publication in professor's book**.