Fanglu (Louie) Yang

416-829-9261 | louieyang27@outlook.com | LinkedIn | Personal Website | Data Science & ML

Education

University of Toronto Sep 2023 – 2027

HBSc – Co-op Specialist in Statistics – Statistical Machine Learning and Data Science Stream
Major in Computer Science, Minor in Economics

Awards: E-Fund Entrance Scholarship of Excellence - 10,000 CAD, 2024 Dean's list

Work Experience

Baker Tilly Canada May 2025 – Aug 2025

Data Science Co-op – **LLM, LangChain, Embedding, Streamlit, Python, Alteryx, Power BI** Toronto, Canada

- Designed and deployed a **SOC 2 audit automation system** with interactive ML-driven control testing dashboard and **LLM chatbot** for policy search and evidence interpretation, cutting manual workload by **90%**.
- Developed an end-to-end **ML pipeline (LangChain, Gemini, Python)** that uses LLM-driven judgment to assess evidence across **80+** controls, increasing audit flagging accuracy by **85%** and reducing review time by **70%**.
- Implemented **RAG** with **vector search and embeddings (LangChain, FAISS)** to power dynamic SOC testing queries and policy lookups, enhancing interpretability and decision accuracy by **85%**.
- Engineered **recursive OCR + anonymization system (Tesseract, SpaCy)** for text extraction and entity masking and integrated Gemini to automate compliance decisions, cutting document processing time by **75%**.
- Designed **Alteryx workflows** for automated data cleaning, merging, and transformation and performed interactive **Power BI dashboards**, streamlining analysis speed for audit GL review and tax compliance reporting by **80%**.

Shanghai Institute of Technical Physics, Chinese Academy of Sciences Apr 2021 – Nov 2021

Research Assistant – **OpenCV, Arduino, 3D-Modelling & Printing** Shanghai, China

- Developed a **computer vision pipeline** using **OpenCV** and **MediaPipe** to recognize real-time piano hand gestures via 21-point hand landmark extraction, improving **gesture detection accuracy** by **65%**.
- Enhanced piano player **gesture stability** by **40%** through **10+** iterative 3D-printed designs and integrated **Arduino-based** feedback mechanisms with ultrasonic sensing and buzzer alerts for real-time posture correction.
- Awarded **3rd Prize** out of **15,000+ students** in two national tech innovation competitions.

Projects

TTC Delay Analytics and Prediction - Datathon: 1st Winner (Best Overall) Mar 2025

UofT SDSS (Students in Data Science and Statistics) Datathon

- Evaluated **10 datasets(150k+ entries)** of Subway, Streetcar and Bus delay data using Python (Pandas, Numpy) and achieved **over 99% accuracy** on lateness prediction with **Correlation Heatmap, Logistic Regression, Classification, and XGBoost**.
- Collaborated **analytical analysis** with Data Scientists and UI/UX Designer to integrate **PowerBI** and **Figma** for interactive dashboard design with **50+ dynamic filtering**, gauge charts, data tables, bar and line graphs.
- Won **Best Overall Project** out of **100 students** in consideration of Visualization, Modeling, and Business Insights.

Car Evaluation Prediction ML Model Using XGBoost Classifier Feb 2025

Analyzed a dataset with **1000+** entries from the UCI Machine Learning Repository

- Achieved **98%** data completeness by handling missing values and applying label encoding via **Pandas and NumPy**.
- Trained an **XGBoost classifier** with **98.92% accuracy** using hyperparameter tuning via **GridSearchCV**.
- Visualized feature importance using **Matplotlib** via bar charts highlighting top predictors for car acceptability.

Statistical Analysis on Research Project Investments by the MCU Mar 2024 – Apr 2024

Analyzed a 20-year (2004–2024) dataset from the Ministry of Colleges and Universities (Coursework Project)

- Implemented **RStudio** for data cleaning, wrangling, statistical analysis (bootstrap, clustering, cross-validation, logistic regression, random forest), and data visualization (ggplot) on a dataset with **5000+ entries**.
- Conducted **A/B testing** from different approaches, recommending Govt of Ontario to diversify research investment.

Skills

Languages: Python (Pandas, NumPy, Matplotlib, Scikit-learn, Seaborn, Keras, PySpark, XGBoost, GridSearchCV, Tensorflow, spaCy), C, Java, JavaScript, HTML/CSS, R, SQL

Tools: LLM, Langchain, Embedding, RAG, Alteryx, NLP, Git, VS Code, Microsoft (Word, Powerpoint, Outlook, Access, Teams), Big Data, Data Visualization, React, RStudio, Excel, CSV, Power BI, Tableau, Microsoft Azure, Jupyter Notebook, Anaconda, OpenRefine, PostgreSQL, MySQL, Arduino, Data Analysis (Regression, Machine Learning, A/B Testing).