

# Keyu Wang

☎ (647)-515-5588 ✉ [keyuwang2002@gmail.com](mailto:keyuwang2002@gmail.com) 🌐 [github.com/asahahaha](https://github.com/asahahaha) 🌐 [Personal Website](#) 🔗 [LinkedIn](#)

## Education

**McGill University, CGPA: 3.96/4.0**

**September 2020 – Present**

*B.Sc. in Computer Science, minor in Geographic Information Science*

*Montreal, QC*

**Courses:** Computer Vision, Applied Machine Learning, Algorithm Design, Model-based Programming, Advanced Geographic Information Science

**Awards:** Dean's Honors List(2x), NSERC Undergraduate Student Research Awards(2x)

## Research

**Mitigating Downstream Model Risks via Model Provenance**

**May 2024 – Present**

*Mila, AI Research Institute*

*Montreal, QC*

- Under the supervision of Prof. Doina Precup, illustrated model provenance risk in healthcare, identified key properties for early warning systems & proposed an open-source, community-led repository for tracking model provenance to enhance transparency and establish a new standard for responsible model management.
- **Pre-print** submitted to NeurIPS 2024 workshops.

**Mapping Western Han: GIS-enabled Historiography for Travel Routes** **February 2023 – January 2024**

*Department of Geography, McGill*

*Montreal, QC*

- Built **project website** from scratch for Historical Geography research project under the supervision of Prof. Raja Sengupta, using JavaScript, HTML, Leaflet.
- Automated map digitization for boundary changes throughout 200 years in early imperial China, using Python & ArcGIS.

## Work Experience

**Data Scientist Intern**

**May 2024 – August 2024**

*Bell Canada*

*Montreal, QC*

- Conducted data analysis on **Google Cloud Platform**, including query optimization and extensive RegEx work for data extraction. Implemented models from **Hugging Face** & followed Agile methodologies to ensure effective collaboration.

**AI & Data Engineering Intern**

**May 2023 – December 2023**

*Bell Canada*

*Montreal, QC*

- Built ML algorithms including clustering for Business Intelligence problems using **Teradata SQL**, Python ML frameworks such as **Scikit-learn**.
- Data analysis & engineering support for building training dataset of a Presence-only model about Competitors' Network Speed Analysis, following ETL pipelines.
- Utilized **Agile** methodologies (Scrum) to ensure effective collaboration with team members, using **JIRA Board & Confluence**.

**Head Instructor Coding & Game Design**

**June 2022 – August 2022**

*Appleby College Camps*

*Oakville, ON*

- Planned course materials for coding and video game design camps, and taught basic Python and GameMaker to over 140 children from 7 to 14 years old.
- Organized weekly showcases & maintained close communication with campers' families for study updates.

## Projects

**You Are At Where You Tweet: GPT Prompting to Geo-locate Twitter Users** | *LLM, AI safety* **April 2024**

- Employed a dataset featuring user-defined locations and tweet texts, the study involved grouping tweets by user, cleaning, and refining input prompts to optimize prediction accuracy.
- Final top-3 accuracy of 47% for worldwide city location inference & 82% for Australian city predictions.

**Logistic & Multi-class Regression on IMDB Movie Reviews and News** | *Machine Learning, NLP* **February 2024**

- Preprocessing of datasets, feature selection through simple linear regression coefficients and mutual information scores, and the development of logistic and multi-class regression models from scratch. We validated our models using small perturbations and monitored cross-entropy to ensure robustness and accuracy.

**Self-implemented KNN & DT Algorithms on Health Datasets** | *Machine Learning, Python, Pandas* **January 2024**

- K-Nearest Neighbors (KNN) and Decision Trees algorithms (without direct use of Scikit-learn) on the health-related datasets. We investigated the strengths and weaknesses of each algorithm in different scenarios, evaluating their performance based on accuracy and AUROC.

## **Frame Prediction for Aerial Objects** | *Computer Vision, OpenCV, MATLAB*

**December 2022**

- Developed object tracking & frame prediction methods for aerial objects in MATLAB based on classic computer vision algorithms.

## **California Gold Rush Historical Storymap** | *JavaScript, HTML, MapBox, GeoJSON*

**October 2022**

- Visualized the history and impacts of immigration to California during 1848-1869 via a storymap.

## **Awards & Scholarships**

---

### **Undergraduate Student Research Award 2024**

\$9000 scholarship issued by Natural Sciences and Engineering Research Council of Canada (NRSEC)

**Undergraduate Student Research Award 2023**      \$9000 scholarship issued by NRSEC

**Perseverance Award for Women in Technology**      \$3000 issued by Quebec government

**McGill University Dean's Honour List \* 2**

## **Skills**

---

**Programming Languages:** Python, Java, C++, C#, HTML/CSS, JavaScript, MATLAB, Bash, OCaml

**IDE/Tools:** VS Code, Eclipse, IntelliJ IDEA, PyCharm, Jupyter Notebook

**Technologies/Frameworks:** Linux, Git, GitHub, GitLab, JUnit, JavaFX, Pandas, Sci-kit Learn

**Databases:** PostgreSQL, DB2, Teradata SQL, Microsoft SQL, SAS

**Spoken Languages:** English, Mandarin Chinese(both native fluency), Japanese (JLPT N2)

## **Leadership / Extracurricular**

---

### **McGill University Chinese Students & Scholars Association**

**September 2020 – Present**

*Vice President Media - Senior Advisor*

- Developed webpages on WeChat Official Accounts for 70+ sponsor stores promotion displays; got 1000+ student users.
- Planned, filmed & edited creative promotion video contents as a leader for club sponsors (such as Liuyishou Hotpot, Presotea, etc.) 3 times each year.
- Co-Directed Montreal Lunar New Year Gala, actively involved in auditions & rehearsals, stage setup, etc.

### **East2West**

**April 2023 – Present**

*Dancer, Editor*

- Active dancer and video editor for a K-pop cover performance crew with a YouTube channel of 1.5M subscribers.