Keyu Wang

(647)-515-5588 keyuwang2002@gmail.com github.com/keyuwww Personal Website LinkedIn

Education

Harvard University **September 2025 – May 2027 (Expected)**
Master of Science in Data Science *Cambridge, MA, USA*

McGill University, cGPA: 3.96/4.0 **September 2020 – December 2024**
B.Sc. in Computer Science, minor in Geographic Information Science *Montreal, QC, Canada*

Courses: Applied Machine Learning, Natural Language Processing, Computer Vision, Database Systems, Algorithm Design, Discrete Mathematics, Statistics, Advanced Geographic Information Science

Awards & Scholarships

Undergraduate Student Research Award (USRA) 2024
$8,700 | Issued by Natural Sciences and Engineering Research Council of Canada (NSERC)
Undergraduate Student Research Award (USRA) 2023 $9,000 | Issued by NSERC
Research Assistantship Stipend $2,550 | Issued by Social Sciences and Humanities Research Council (SSHRC)
Perseverance Award for Women in Technology 2022 $3,000 | Issued by Quebec Government
McGill University Dean's Honour List 2022
McGill University Dean's Honour List 2021

Publications

1. **K. Wang***, J. Li*, S. Yang, Z. Zhang, D. Wang. *When Truth Is Overridden: Uncovering the Internal Origins of Sycophancy in Large Language Models.* [Paper]

2. S. Yang, S. Zhu, Z. Wu, **K. Wang**, J. Yao, J. Wu, L. Hu, M. Li, D.F. Wong, D. Wang. *Fraud-r1: A Multi-round Benchmark for Assessing the Robustness of LLMs Against Augmented Fraud and Phishing Inducements.* In **Findings of the Association for Computational Linguistics (ACL), 2025**. [Paper]

3. **K. Wang**, A.N. Iranzad, S. Schaffter, M. Risdal, D. Precup, J. Lebensold. *Mitigating Downstream Model Risks via Model Provenance.* In **NeurIPS 2024 Workshop on Socially Responsible Language Modelling Research**, **2024**. [Paper]

Work Experience

Research Intern **May 2025 – August 2025**
Okinawa Institute of Science and Technology (OIST) *Okinawa, Japan*

- Joined the Machine Learning and Data Science Unit under **Prof. Makoto Yamada**, focusing on fractal video pretraining on brain-inspired vision foundation model.

Machine Learning Engineer **Feb 2025 – May 2025**
Moonarch *Remote*

- Developed document extraction pipeline for startup company profiles, getting structured insights from unstructured documents, enabling accurate parsing of mixed-format business documents containing images and embedded text.

Visiting Student Researcher **Jan 2025 – May 2025**
King Abdullah University of Science and Technology (KAUST) *Thuwal, Saudi Arabia*

- Conducted research in the Provable Responsible AI and Data Analytics (PRADA) Lab under **Prof. Di Wang**.
- Explored LLM sycophancy behaviors, designing a prefix framework to test models like Llama 3.2 and Qwen 2.5 and conducted mechanistic analysis to explain the reason why.

Research Intern on AI Safety **May 2024 – December 2024**
Mila, AI Research Institute *Montreal, QC, Canada*

- Under the supervision of **Prof. Doina Precup** with collaborators from Google & Meta, illustrated model provenance risk in healthcare, identified key properties for early warning systems & proposed an open-source, community-led repository for tracking provenance to enhance transparency and establish a new standard for responsible model management.
- As the 1st author, **paper** is accepted to NeurIPS 2024 workshop: Socially Responsible Language Modelling Research.

Data Scientist Intern **May 2024 – August 2024**
Bell Canada *Montreal, QC, Canada*

- Conducted data analysis on **Google Cloud Platform**, enhancing service assurance outcomes by streamlining query performance and utilizing advanced RegEx to extract and refine data efficiently.
- Reduced the initial dataset to 6% of its original size, focusing on high-value data, and delivered actionable insights using effective data visualization, while collaborating within the team using Agile methodologies.

AI & Data Engineering Intern **May 2023 – December 2023**
Bell Canada *Montreal, QC, Canada*

- 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, using **JIRA Board & Confluence**.

Undergraduate Research Assistant **February 2023 – January 2024**
Department of Geography, McGill *Montreal, QC, Canada*

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

Head Instructor Coding & Game Design **June 2022 – August 2022**
Appleby College Camps *Oakville, ON, Canada*

- 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

McGill School of Computer Science Scheduling Website | *Web Development* **December 2024**

- As one of the 2 backend developers who made a user-friendly platform that simplifies appointment scheduling.

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 accuracy.
- Final top-3 accuracy of 47% for worldwide city location inference & 82% for Australian city predictions.

Frame Prediction for Aerial Objects | *Computer Vision, OpenCV, MATLAB* **December 2022**

- Developed object tracking & frame prediction methods for aerial objects in MATLAB using classic (non-AI) 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.

Presentations

Socially Responsible Language Modelling Research (SoLaR) workshop at NeurIPS 2024 **December 2024**

- Given poster presentation about the "Mitigating Downstream Model Risks via Model Provenance" paper at NeurIPS workshop in Vancouver, Canada.

Undergraduate Computer Science Research Symposium (UCORE) | *LLM, AI safety* **September 2024**

- Given poster presentation about the "Mitigating Downstream Model Risks via Model Provenance" project at a school-level research poster session with 60+ attendees.

GIS (Geographic Information Science) Day 2022 | *Historical Geography* **November 2022**

- As an outstanding student project, invited to give an oral presentation at GIS Day hosted by McGill for presenting a story map visualizing the history and impacts of immigration to California during 1848-1869.

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, TensorFlow, PyTorch
Web Development : XAMPP, React, Node.js
Databases: PostGreSQL, DB2, Teradata SQL, Microsoft SQL, SAS
Video Production : Adobe Premiere Pro, Adobe Lightroom
Spoken Languages: English, Mandarin Chinese(both native fluency), Japanese (JLPT N2)

Leadership / Extracurricular

East2West **April 2023 – Present**

Dancer, Editor

- Danced & edited long videos for a K-pop cover performance crew with a YouTube channel of 1.5 million subscribers, achieving on average 20K views per video.
- Participated in the "2024 K-pop Cover Dance Festival" and won 3rd place in the Canada division.

K-Rave McGill **September 2021 – December 2024**

Dance Leader

- Led K-POP dance projects by organizing practice logistics, teaching dances, and giving members weekly individual feedback for improvement to ensure performance quality.

McGill University Chinese Students & Scholars Association **September 2020 – April 2024**

Vice President Media - Senior Advisor

- Directed, filmed & edited creative promotion video content for club sponsors (such as Liuyishou Hotpot, Presotea Montreal, etc.) 3 times yearly as a part of the signed discounting contract, to benefit all McGill Chinese students from receiving 10 to 25% off discounts.
- Co-directed Montreal Lunar New Year Gala 2022 with 100+ performers and 15 staff, actively involved in venue booking, auditions, rehearsals, & stage setup.
- Developed webpages on WeChat Official Accounts for 70+ sponsor stores promotion displays; got 1000+ student users.
- Organized member photoshoot for 70 internal members, involving studio visiting & bookings, creative ideas, pose directing & photography.