Jingxuan (Carol) Huang

(+1) 236-991-8860 | jxhuang_carol@outlook.com | Linkedin

EDUCATION

University of British Columbia (UBC)   **Vancouver, Canada**
Master of Computer Science   *Sep.2022 – Jun.2024 (Expected)*

Nanjing University (NJU)   **Nanjing, China**
Bachelor of Information Management and Information Systems   *Sep.2018 – Jun.2022*
- **GPA**: 4.65/5.00 **Rank**: 1/56

WORK EXPERIENCE

Microsoft Research Asia (MSRA) – Beijing, China   *Oct.2021 – May.2022*
Research Intern, Supervised by Weiwei Cui and Yeye He (in Microsoft Research, Redmond, USA)
- Merged heterogeneous tables and experimented on different ways to discover Data Quality rules using knowledge acquired from the data lake.
- Learned and used PySpark to boost the efficiency of data manipulation.
- Produced a detailed documentation to explain code and summarize current results.

Nanjing University – Nanjing, China   *Jul.2019 – Mar.2021*
Research Assistant, Supervised by Prof. Lele Kang
- Wrangled and analyzed a mass of data in PATSTAT, a patent statistical database.
- Constructed measurements of abstract concepts with machine learning methods and used regression analysis to test hypotheses.
- Gave a presentation and published a conference paper in *International Conference on Human-Computer Interaction (HCII)* as first author in 2021, where we discussed the impacts of different search behavior on the innovations of emerging technologies.

PROJECT EXPERIENCE

DataPrep – Simon Fraser University, Canada (Remote Working)   *Apr.2021 – Sep.2021*
Summer Research Intern, Supervised by Prof. Jiannan Wang
- Helped to develop DataPrep, a powerful library to prepare data in Python for data scientists. Aimed to develop new methods of cleaning data with semantics.
- Implemented and optimized new functions for cleaning and validating language, and further added 100+ clean functions for number types to diversify the library.
- Resolved issues from the community and tested existing clean functions to find and fix limitations.
- Proposed and implemented a new strategy to infer column headers with features extracted by functions in the library. Conducted experiments to show its superiority over existing models.

HONOR & AWARDS

- **Stars of Tomorrow** (Award of Excellent Intern), Microsoft Research Asia   *May.2022*
- **Provincial-Level Merit Student**, Education Dept. of Jiangsu Province *(Top 0.4%)*   *May.2021*
- **Duxia Foundation Scholarship**, NJU *(Awarded to only 6 students in NJU)*   *Jun.2020*
- **National Scholarship**, Chinese Ministry of Education *(Top 0.2%)*   *Oct.2019*

TOOLS & LANGUAGES

Python, SQL, Git, Latex, C, C++, R, Spark, Shell