# Yusen Jiao

Website: y29jiao.github.io/Yusen-Jiao.github.io/ | Cell:+1 (416) - 991 - 1019 | Email: y29jiao@uwaterloo.ca

## Education:

**University of Waterloo**, Waterloo, ON **Date of Graduation**: **September 2023**

• Master of Engineering in *Systems Design Engineering: AI and Machine Learning*

• Cumulative GPA: 3.9 out of 4.0

**University of Waterloo**, Waterloo, ON **Date of Graduation**: **April 2022**

• Bachelor of Honor Mathematics, Major in *Information Technology System Management* and Minor in *Statistics*

**Relevant Courses:** Advanced Image Processing, Statistical Pattern Recognition, Database Management, Experimental Design, Applied Linear Models and Process Improvement, Computational Statistics and Data Analysis, Applications Software Engineering

## Research Interest:

I am Interest in leveraging **Machine Learning** techniques to enhance and **optimize Image Processing** Algorithms.

## Publications:

Yifei Zhang, **Yusen Jiao**, Jiayi Chen, Zhaoyang Li, Jieyu Zhang and Frederic Sala. "Just Select Twice: Leveraging Low Quality Data to Improve Data Selection". In *submission: Conference on Neural Information Processing Systems (NeurIPS), Datasets and Benchmarks Track, 2024.*

## Research Experience:

**Research Assistant**, University of Wisconsin-Madison **2024/01 – Present**

• Co-authored a research paper aimed at advancing data valuation techniques in Data-Centric Machine Learning, submitted to the NeurIPS, Datasets and Benchmark Track.

• Played a key role in designing and implementing our data valuation framework, which significantly enhances the identification and utilization of high-quality data.

• Conducted the experimental validation of our framework using optimal transport-based methods across multiple imaging and NLP datasets, confirming its superiority in optimizing dataset quality.

• Extended our novel data valuation framework to incorporate additional valuation methods, demonstrating the framework's capability to enhance performance across data valuation methods.

**Research Assistant**, University of California, Berkeley **2023/09 – 2024/01**

• Contributing to a project at UC Berkeley's Deep Learning Lab, focused on enhancing the capabilities of a 3D Large Language Model.

• Focused on optimizing the model's performance to accurately describe intricate scene features, enhancing its interpretative abilities.

• Continuously updating project strategies by reviewing latest research in deep learning and 3D modeling, and contributing insights in lab meetings and seminars for collaborative research advancement.

**AI-powered English Speaking Tutor**, Waterloo, ON **2023/04 – 2023/08**

***Research Objective: allow users to freely communicate with the software in English to achieve the effect of practicing speaking***

• Incorporated the ChatGPT model as the core of our newly developed AI system, with the main purpose of leveraging its powerful conversational capabilities; designed special prompts that guide the generation process, ensuring that the answers generated are understandable to users.

• Employed and trained the VITS model via a dataset consisting of 12,000 unique voices, enabling it to produce a remarkable human-like voice effect, ensuring that the generated voice is indistinguishable from that of a real human and eliminating any perception of artificiality typically associated with machine-generated voices.

**Detecting Violation of Helmet Rule for Motorcyclists,** Waterloo, ON **2023/01 - 2023/04**

Process and analyze the given video to identify the motorcycle rider who violated the traffic rules:

• Pre-processed and trained the videos using YOLOv5: The motorcycle rider who violated the traffic rules was accurately identified with achieving an accuracy of over 85%**.**

• Developed a customized CNN-based neural network model based on Tensorflow framework, with training and testing via the pre-processed videos; built an SVM model as a benchmark to show the priority of the developed CNN model.

• Developed a Variational Auto-Encoder (VAE) to augment the original video training datasets, enhancing the training process and improving the classification performance of the CNN-based model.

• Derived the strengths and weaknesses analysis of the models considered in this study, and reflected on potential avenues for future improvement options.

**Waterloo Region House Purchasing Analysis,** Waterloo, ON **2020/05 - 2020/09**

• Implemented and compared performance review of single and multiple linear regression models on the dataset of Waterloo Region House Purchase records; Analyze the correlation and determinant of self-explanatory variables and response by examining the corresponding statistical significance.

• Led group of 3 on constructed efficient R scripts for model simulation and result validation; successfully predicted optimal purchasing information with an appreciable R-square and provided a recommendation based on customized weight inputs.

## Professional Experience:

**Data Analyst,** Beijing DiDi Infinity Technology and Development Co., Ltd, **2021/04 - 2021/08**

• Established the provincial SQL database, performed statistical data analysis and visualization for Hebei Province in a group of two; preprocessed and categorized the gathered information from an internal database.

• Developed an automated data framework for business intelligence and operations insight using Python and SQL, resulting in a significant 20% improvement in data validation efficiency and accuracy.

• Modified a deterministic model for predicting traffic flow, embedded in the above data framework, into a Bayesian prediction model by adding a priori mechanism to better capture the stochastic nature of traffic.

• Addressed various requirements aimed at enhancing specific task completion degree via Python, R and SQL: optimized the commission algorithm for drivers in Hebei Province; provided recommendations for product bundles and promotions to increase sales by 13%; helped groups develop effective strategic measures to expand the user base, resulting in a notable 17% growth in the number of users.

## Brief Personal Description:

Throughout my graduate studies, I have consistently maintained a high GPA, reflecting my solid understanding of machine learning fundamentals. I actively sought opportunities to enrich my research experience, rapidly advancing my expertise. With each project, I not only dedicated myself fully but also rapidly evolved in my research capability. This journey of exploration led me to identify my true passion in the fields of computer vision and image processing, areas where I have successfully applied my knowledge to develop innovative solutions for real-world problems.

## Skills

• **Programming Skills:** Python, SQL, R, C, Java

• **Analytical Skills:** Machine Learning Algorithm Development, Statistical Modeling, Numerical Analysis, Database Development and Management

• **Software Proficiencies:** MySQL, R Studio, Minitab, MS Suite