Kecen Yao

New York University
+86 13660739596 ky2138@nyu.edu linkedin.com/KecenYao github.com/KecenYao

Education

University of Toronto **Incoming**
Master of Science in Applied Computing *Toronto, Canada*

New York University Shanghai **September 2020 – June 2024**
Bachelor of Science in Computer Science & Data Science (Concentration in AI) *Shanghai, China*

- Cumulative GPA: 3.87/4.00, Major GPA: 3.97/4.00
- Selected Awards:
  - Dean's List of NYU 2021-2023
  - NYU Shanghai Hackathon Best Result Award

New York University **September 2022 – June 2023**
NYU Study Abroad Program *New York, NY, United States*

- Cumulative GPA: 3.96/4.00, Major GPA: 3.96/4.00

Relevant Coursework

- Algorithms (A)
- Special Topic: NLP (A)
- Computer Systems Org (A)
- Multivariable Calculus (A)
- Data Structures (A-)
- Applied Internet Tech (A)
- Operating Systems (A-)
- Probability and Statistics (A)
- Machine Learning (A)
- Reinforcement Learning (A)
- Database Design & Implementation (A)
- Linear and Nonlinear Optimization (A)

Experience

Machine Learning Intern @ UniDT Technology (Shanghai) Co., Ltd **August 2023 – December 2023**
Supervisor: Dr. Xiaohua Xuan & Dr. Kehuan Shi, Algorithms Department *Shanghai, China*

- Constructed and deployed a Large Language Model (LLM) called "Wisdom-Interrogatory" specifically designed for legal queries.
- Achieved massive legal data regularization, cleaning, and analysis; trained and fine-tuned the LLM.
- Improved the LLM's deploying website through developing the web development framework. Used Vue.js for improved user interaction on the front end and established a comprehensive database and API management on the back end.

Research Assistant & Tandon UGSRP member - AI4CE Lab @ New York University **December 2022 – Present**
Supervisor: Professor Chen Feng, assistant professor of Civil and Urban Engineering at NYU Tandon *New York, NY, United States*

- Built a webcam data downloading pipeline and down-sampled video loading pipeline independently, contributing to a new street webcam streaming video database targeting crowd detection. Benchmarked the database using state-of-the-art energy-saving detection algorithms.
- Explored the role of uncertainty in object counting with its statistical significance and utility in enhancing counting accuracy.
- Implemented inductive conformal inference to generate pixel-wise density ranges aligned with ground truth densities at specified probabilities (e.g., 90%).Developed a method to integrate pixel-wise density uncertainties into the counting process, refining counting regressors through local uncertainty adaptation.

Research Assistant - NYU Multimedia and Visual Computing Lab @ New York University **July 2023 – September 2023**
Supervisor: Professor Yi Fang, associate professor of Electrical and Computer Engineering at NYU Abu Dhabi and NYU Tandon *New York, NY, United States*

- Successfully integrated pre-trained language model like GPT-4, visual language model like CLIP, and task-driven model such as GR-ConvNet, to develop a zero-shot learning system that can generate gestures for robot grasping and placing tasks using RGBD images.
- Developed a robust pipeline that enabled autonomous robotic self-checking through the implementation of Visual Question Answering (VQA), empowering the model to assess and validate its actions.
- Explore robotic simulation using ROS (Robot Operating System) and Gazebo, successfully running robotic arm simulation demos.

Selected Projects

Advanced Deep Reinforcement Learning Algorithm Analysis in BipedalWalker Environment | *Python, Pytorch, RL* **September 2023 - January 2024**

- Conducted an extensive comparative study of prominent Deep Reinforcement Learning (DRL) algorithms, including Proximal Policy Optimization (PPO), Deep Deterministic Policy Gradients (DDPG), Twin Delayed DDPG (TD3), and Soft Actor-Critic (SAC), within the OPENAI Gym's BipedalWalker environment.
- Enhanced and applied new performance metrics for more accurate evaluation of algorithm efficacy, demonstrating distinct advantages and limitations of each algorithm through both quantitative and qualitative analysis.

Neutral Summarization for Framing Biased Articles | *Python, Pytorch, LLM* **January 2023 - May 2023**

- Developed an innovative approach to automatic neutral summarization using the T5 encoder and fine-tuned T5 decoder, proposing a solution for mitigating media framing bias caused by selective journalistic writing, aiming to foster social solidarity and reduce division among readers.
- Examined the model performance, demonstrating that it outperforms baseline model (Bert) by demonstrating excellence in Avg. Framing Bias metric (arousal) and Salient Info metrics (BLEU and ROUGE1-R), underscoring the model's ability to produce highly balanced and informative content.
- Proposed an innovative single-input-document neutral summarization architecture, which can be applied to the real world sector.

School Forum - Web Development Project | *JavaScript, HTML, MongoDB, Vue* **January 2023 - May 2023**

- Developed a full-stack web development project incorporating front-end and back-end technologies, including Express, Handlebars, and Mongoose.
- Innovated a successful and fully functional online forum website encompassing essential features such as user authentication (login/register), post creation, commenting, deletion, and liking/disliking posts.
- Implemented robust data management via storing and retrieving data from a MongoDB server to ensure seamless user interactions on the platform.

NYC Taxi Prediction - Machine Learning Projects | *Python, Pytorch, Sklean* **January 2022 - May 2022**

- Established commonly used frequently-used machine learning models to forecast future taxi cab demand in New York City.
- Conducted thorough data cleaning and preprocessing of New York City taxi data which were obtained from the official website, utilizing Pandas and NumPy to achieve effective data manipulation and employing standard machine learning techniques such as Test-Train-Split and Cross-Validation.
- Implemented supervised learning clustering algorithms in Python, such as Linear Regression, Tree-based models, KNN (K-Nearest Neighbors), and Recursive Neural Network (RNN). Generated an outstanding model performance, with the best-performing model (RNN) attaining an impressive R2 score exceeding 0.99, demonstrating exceptional predictive accuracy.

Teddy Bear - Self-developed 3D game | *C#, Unity* **January 2022 - May 2022**

- Developed a Unity-based 3D parkour game independently which featured self-developed scripts, models, settings, audios, etc.
- Created the innovative Gameplay, which involves jumping, wall sliding, scene traversal, item collection, and death checkpoint.

Technical Skills

Programming Languages: Python (proficient), C/C++, C#, JavaScript, HTML, CSS, SQL, MATLAB
Technologies/Frameworks: Linux, Git, Docker, MongoDB, Unity
Machine Learning: PyTorch, KNN, CNN, RNN, Transformer, LSTM