

Kuan-Lin (Gary) Liu

kuanlin.liu@nyu.edu ◇ (929) 332-5669 ◇ [linkedin.com/in/kuanlinliu](https://www.linkedin.com/in/kuanlinliu) ◇ Brooklyn, NY ◇ [garykl.github.io](https://github.com/garykl)

EDUCATION

- | | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------|
| New York University <i>M.S. in Data Science, GPA: 3.8 / 4.0</i> ◦ <i>Related Courses:</i> Deep Learning, Natural Language Understanding, Deep Learning Systems (Cloud, MLOps), Big Data (Hadoop, Spark, Scala), Machine Learning, Database Systems, Introduction to Data Science | New York, NY Expected May 2021 |
| National Taipei University <i>B.B.A. in Statistics and B.A. in Economics, GPA: 3.8 / 4.0</i> ◦ <i>Related Courses:</i> Data Structures, Algorithms, Object-oriented Programming, Data Mining, Dimension Reduction | New Taipei, Taiwan Sept 2014 - Jan 2019 |

TECHNICAL SKILLS

- | | |
|----------------------------------|-------------------------------------------------------------------------------|
| Languages | Python, Scala, SQL, R, Bash Script, C++, HTML, CSS |
| Software & Tools | Hadoop, Spark, AWS, GCP, CUDA, MySQL, Git, Tableau, RESTful API |
| Frameworks & Packages | PyTorch, Scikit-learn, TensorFlow, Pytest, Streamlit, Flask, Dash, Matplotlib |

PROFESSIONAL EXPERIENCE

- | | |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|
| Machine Learning Intern New York University, NY ◦ Designing neural networks with dynamic reaction time to model printed texts into phonemes (NLP, Seq2Seq) ◦ Performing distributed computing (PyTorch Lightning, CUDA) with GPUs to accelerate training by 66% ◦ Brainstorming multiple experimental hypotheses with diverse professors to compare models in different scenarios | May 2020 - Present |
| Data Science Intern Cathay Financial Holdings, Taiwan ◦ Researched on the performance of GNNs (PyTorch Geometric) and sequence models using E-commerce reviews ◦ Built ETL pipelines to speed up text cleaning (Multiprocessing) and create pretrained embeddings (BERT) ◦ Ensured model reliability using unit testing (Pytest) and deploying models with a web app (Streamlit) | July 2020 - Aug 2020 |
| Research Assistant Academia Sinica, Taiwan ◦ Wrangled electricity panel data from 300+ dorms and improved classification accuracy by 7% (R, Google Cloud) ◦ Visualized abnormal records from Hierarchical Clustering and present data insights using R Markdown and ggplot2 | Mar 2018 - July 2018 |
| Teaching Assistant National Taipei University, Taiwan ◦ Instructed 80+ students in team cooperation using version control (Git) and R programming with real-world data ◦ Created an interactive tutorial web app by LearnR and Shiny to intrigue students for better learning | Sept 2018 - Jan 2019 |

SELECTED PROJECTS

- | | |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| Fake Reviews Detection for Restaurants ◦ Led the text preprocessing pipeline to boost efficiency from 20 hours to minutes by Spark NLP and NTLK ◦ Engineered on behavioral features to improve ROC score by 16% (XGBoost), against text features (GloVe, BOW) | Mar 2020 - May 2020 |
| Book Recommendation System for Goodreads.com ◦ Analyzed 4GB user-item interaction data in PySpark and built a Collaborative Filtering recommendation system ◦ Enhanced efficiency of the book-searching system by 10 times and visualized clustering of products with T-SNE | Mar 2020 - May 2020 |
| Automatic Summary Generator ◦ Integrated BertSum and TextRank to capture the importance of sentences in the 1.4GB CNN/DailyMail dataset ◦ Designed a graph-based models by modifying outdated TensorFlow scripts and fine-tuning Pointer-Generator | Mar 2020 - May 2020 |
| Kickstarter's Success Prediction and Product Recommendation ◦ Developed an ETL pipeline using Spark SQL and Scala on HDFS and Spark for faster cleaning JSON files ◦ Built Machine Learning pipelines with Spark's MLlib , and visualized patterns with a Tableau dashboard ◦ Recommended similar products on Amazon by Locality-sensitive Hashing and extracted top words by TF-IDF | Oct 2019 - Dec 2019 |