

# Kuan-Lin (Gary) Liu

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

## EDUCATION

---

### New York University

*M.S. in Data Science, GPA: 3.80 / 4.0*

New York, NY

Expected May 2021

- *Related Courses:* Natural Language Understanding, Advanced Machine Learning, Big Data (Hadoop, Spark, Scala), Database Systems, Introduction to Data Science

### National Taipei University

*B.B.A. in Statistics and B.A. in Economics, GPA: 3.83 / 4.0*

New Taipei, TW

Sept 2014 - Jan 2019

- *Related Courses:* Data Structures and Algorithms, Object-oriented Programming, Data Mining, Dimension Reduction, Experimental Design, Time Series, Econometrics, Mathematical Statistics, Categorical and Linear Regression

## TECHNICAL SKILLS

---

### Programming

Python, Scala, SQL, R, C++, MATLAB, SAS

### Software & Tools

Hadoop, Spark, MySQL, AWS, RESTful API, Git, Tableau, HTML, CSS

### Packages

PyTorch, Scikit-learn, NLTK, Spacy, Flask, Selenium, BeautifulSoup, Pandas, NumPy

## PROFESSIONAL EXPERIENCE

---

### Machine Learning Intern

*New York University*

May 2020 - Present

New York, NY

- Designing neural network architecture with dynamic reaction time to model English/Italian printed texts into phonemes, which simulates human's reading comprehension and Dyslexia problems (**PyTorch**, **NLP**, **Seq2Seq**)
- Collaborating closely with professors from diverse backgrounds to brainstorm about state-of-the-art research papers

### Data Science Research Intern

*Cathay Financial Holding Co., Ltd.*

July 2020 - Present

Taipei, Taiwan

- Conducting research on **Graph Neural Networks** to build connections between words using Amazon reviews
- Transforming texts to word vectors via pretrained embeddings, and training models on AWS (**PyTorch Geometric**)

### Research Assistant

*Research Center for Humanities and Social Sciences, Academia Sinica*

Mar 2018 - July 2018

Taipei, TW

- Wrangled electricity panel data from 300+ dorms and improved classification accuracy by 7%. (**R**, **Google Cloud**)
- Visualized abnormal records with **Clustering** and present data insights during weekly reports using R Markdown

### Teaching Assistant

*National Taipei University (Courses: Programming for Data Science)*

Sept 2018 - Jan 2019

New Taipei, TW

- Instructed 80+ students in collaborating with **Git** and R programming data wrangling using real-world data
- Created an interactive tutorial web app by LearnR and Shiny to intrigue students for better learning

## SELECTED PROJECTS

---

### Fake Reviews Detection for Restaurants in NYC

Mar 2020 - May 2020

- Boosted efficiency of the text preprocessing pipeline from 20 hours to minutes by parallel computing (**Spark NLP**)
- Focused on feature engineering to improve ROC score by 16% (**XGBoost**), compared to word vectors (**GloVe**)

### Automatic Text Summary Generator: News Application

Mar 2020 - May 2020

- Integrated **BertSum** and **TextRank** to capture the importance of sentences in the 1.4GB CNN/DailyMail dataset
- Designed a graph-based summarization pipeline to reduce word repetition by fine-tuning a pretrained Seq2Seq model

### Book Recommendation System for Goodreads.com

Mar 2020 - May 2020

- 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

### Predicting Kickstarter's Success and Recommending Products from Amazon

Oct 2019 - Dec 2019

- Built machine learning pipeline with Spark's MLlib and BigDL, and visualized patterns with a **Tableau** dashboard
- Developed an ETL pipeline using **Spark SQL** and **Scala** on HDFS and Spark for faster cleaning JSON files