(Simone) **Simeng Li**

47 Changhong Road, Fancheng District, Xiangyang, Hubei Province, China 441002
Phone: (86) 18810591026 · Email: simone@bu.edu
Homepage: https://simengli1998.github.io/
GitHub: https://github.com/SimengLi1998
LinkedIn: https://www.linkedin.com/in/simeng-li-simone

EDUCATION

Master of Science in Business Analytics Jul. 2021 – May 2022
Boston University, Boston, USA | GPA 3.24/4

- Main courses: *Analyzing Data, Business Analytics Toolbox, Supervised and Unsupervised Machine Learning, Business Experimentation and Casual Methods, Marketing Analytics, Financial Analytics, Data Ethics, and Advanced Analytics Topics, Projects and Practicum*
- Project Management Officer of Capstone Project with Spinnaker Analytics

Second Degree(postgraduate) in Info Management & Info Systems Sept. 2020 – Jul. 2021
Beijing Institute of Graphic Communication, Beijing, China | GPA: 85.8/100

- Courses: *Design and Production of App, Digital Media Technology and Content Management, Statistics, Introduction to Information Management and Information Systems, Database Technology and Application, Principles of Management, Business Data Analysis, Big Data Technology and Application*

Bachelor of Engineering in Mechanical Engineering Sept. 2016 – Jul. 2020
Beijing Institute of Graphic Communication, Beijing, China | GPA: 3.87/5

- Relevant courses: *Advanced Mathematics I and II, Linear Algebra, Probability and Statistics, Computer and Information Technology, VB Program Design, Electronic Process Practice I, Basics of Control Engineering, Engineering Graphics I and II, Engineering Graphics Mapping, Industrial Robot Technology, Computer-Aided Engineering Calculations, Virtual Reality Design*
- Consistently among Top 4 of the Class
- Excellent Student Scholarship + Excellent Student Organizer Award, 2017, 2018, 2019

PUBLICATIONS

- Simeng Li, Hao Fan*, Yifan Wang, "Dynamic Monitoring of Google Play Store Ranking Changes: Data Mining of Important Features", The 5th International Conference on Big Data & Artificial Intelligence & Software Engineering (http://www.icbase.org/), September 20-22, 2024
- Simeng Li, "Mitigating Illusions in LLMs: The Power of Knowledge Graphs", Technology and Information journal, Paper 430073, Dec. 2023
- Yuansheng Qi*, Simeng Li, Yinping Gu, "Application of Intelligent Robots in the Printing Industry", Journal of the 4th China (Guandong) International Technology Exhibition, Apr. 2019

*Note: *corresponding author*

WORKING PAPERS

- Simeng Li, Hao Fan*, "Exploring the Similar Recommendation Mechanism in Google Play Store: A Multimodal Data Fusion Approach", planned for publishing in a SCI

Journal in March 2025
- Simeng Li, Hao Fan*, "Dynamic Knowledge Weight: Automatic Enhancement of Prompt Expression", planned for publishing at an academic conference in 2025

WORK EXPERIENCE

Data Mining Specialist Jul. 2022 - Present
Wuhan Dobest Information Technology Co., Wuhan, China
(www.wedobest.com.cn)

As an award-winning member of the Data Science Group, Platform Division, monitor and analyze the activities of the company's 1,906 game products and more than 15 million daily active users in 160 countries and regions, who post a daily average of 150,000 comments. Independently develop algorithms, models and tools and provide guidance to 12 members of the Overseas Publicity Group for optimizing product release, exposure and rankings in Google Play Store, for the purpose of improving user experience, acquisition and retention. Successfully completed a series of notable projects:

- Teamed up with 3 coworkers to develop Zhiqite (https://zhiqiteai.cn), our company's integrated AI Generated Content system that features texts, images, audios, and videos. Specialized in prompt engineering for the Applications module, which had 105 creative applications. For the Solutions module, rejuvenized outdated training data in LLMs by incorporating Serper's real-time search, LangChain, and Few-Shot Learning. Additionally, utilized FastAPI and ChatGPT-3.5 to develop a marketing copywriting tool, which cut down writing time from 2 hours to 2 minutes per piece, with an average cost of $0.005 per request and 90% availability.
- In a team of 12, led the development of a user comment analysis system capable of handling over 100 million comments in 1,906 game products and multiple languages. Designed a multi-task function with a single LLM invocation to output results about sentiment, intensity, intention, keywords, and comment translation and to optimize token and time costs. Deployed the FastAPI-based system on Tencent Cloud and achieved 99% success rate in real-time comment monitoring, at a cost of merely USD $0.002 per comment request.
- Organized a team of 8 engineers to optimize the display of the company's 1,906 game products in Google Play. To improve dominance analysis, increased the pretreatment features from 28 to 63, and then eliminated features that had an expansion factor of above 10. The seventh and final version of solution increased the visibility of the company's products in Google Play by 30%-272% in America and by 19%-27% worldwide.
- Independently developed an automatic tagging algorithm to process 16,800 reminder themes for English Block Puzzle, an edutainment game. The project completely replaced the manual tagging task that would take 5 editors a month with instant automatic tagging. Leveraged WordNet for word annotation and used the bart-large-mnli model for semantic tagging. Achieved an accuracy rate of 72% in theme categorization, with potentials for even higher accuracy with further training.

Algorithm Development Intern (part-time) Jan.– Jul. 2021
Zhangyou Technology Co., Wuhan, China
(www.zhangyou.com)

- Actively assisted the development and optimization of an automatic multi-point serving system for game products on various platforms with modules including the dispatching hub, ad materials and copy pretreatment, video and NLP algorithms, Wide&Deep and DeepFM models, digitalization, efficiency enhancement, and cost control.
- Joined the development of core algorithms to forecast in-game purchasing behavior, with

a precision rate of above 90%.

GRADUATE SCHOOL PROJECTS

Boston University Jul. 2021 – May 2022

1. Supervised Machine Learning
   - With guidance by Dr. Georgios Zervas, *Severance forecast of traffic accidents in Boston*, by using data pretreatment, exploratory data analysis, Seaborn data visualization, and scikit-learn machine learning library.
   - With guidance by Dr. (Shawn) Zusheng Jin, *Capstone Project*: As PMO, led a group of five graduate students to work with Spinnaker Analytics, a Boston-based decision solution analytics company, to forecast the stock market by using long short-term memory and Arima model of time series forecasting in Python.
   - With guidance by Dr. Mohammad Soltanieh-ha, *Airbnb price forecast*, by using Tableau, Google Cloud Platform, Big Query, and scikit-learn.
2. Unsupervised Machine Learning
   - With guidance by Dr. Brock Tibert, *Customer clusters analytics*, by using sentiment analysis, natural language processing, principal component analysis, t-distributed stochastic neighbor embedding, Uniform Manifold Approximation and Projection, k-means clustering, and H-cluster.
3. Advanced Data Analytics
   - With guidance by Dr. Andrey Fradkin, *Casual analysis*: Analyzed the topic, "Does raising questions in the classroom lead to better command of knowledge?", by using comparative experiment, questionnaire, t-test, balance check, and Regression in R.
   - With guidance by Dr. Dokyun (DK) Lee, *Analysis of Credit Card Applications*: by using proxy variables analysis, ethical frameworks (consequentialism, deontology and axiology), the feedback loop, and gender and racial discrimination analysis.
   - With guidance by Dr. Peter Wysocki, *Stock ROI analysis of listed companies in America*: by using pretreatment of missing and abnormal data, correlation analysis, ordinary least squares regression, and Seaborn and A/B testing data visualization.
4. Deep Learning
   - With guidance by Dr. Gordon Burtch, *Forecast Walmart sales ahead of holiday fluctuations*: Cleaned data and performed EDA for attribute understanding. Utilized feature selection, k-fold cross-validation, and time-series models like ARIMA and LSTM for overfitting mitigation. Optimized via grid search to accurately predict future departmental sales.

UNDERGRADUATE PROJECTS

Beijing Institute of Graphic Communication Sept.2020 – Jan. 2021

- With guidance by Dr. Shuo Liu, Coordinated data preparation for the "Study of Public Opinion Guidance Strategies under Emergent Public Health Incidents: Case Study of Covid-19 Situation in Beijing".
- With guidance by Dr. Shuo Liu, Led a group of 5 students in data processing and analysis for the "Study of Innovative Cultural Industries under the Development of New Technology".

CERTIFICATIONS AND SKILLS

- Google Analytics Individual Qualification Certificate, Feb. 2022 and Nov. 2023
- Prompt Engineer Qualification, Certified by iFLYTEC and Datawhale, Oct. 2023
- Experienced user of Python, R, SQL, Basic Linux, Git, FastAPI, Google Cloud, Tableau, Hive, Kafka, Machine Learning, Deep Learning, Prompt Engineering, and LangChain