Yipeng Liu

Tel: +1 (515) 450-0253 | Email: yipengl@iastate.edu | LinkedIn | GitHub | Google Scholar

EDUCATION

Iowa State University — Ames, IA, USA
Ph.D. - Statistics — *Aug. 2025-May 2030*

University of Michigan — Ann Arbor, MI, USA
M.S. - Data Science — *Aug. 2023-May 2025*

Central University of Finance and Economics — Beijing, China
B.S. - Data Science and Big Data Technology — *Sept. 2019-Jun. 2023*

PUBLICATIONS

- S. Marino, Y. Wang, **Y. Liu**, M. Tang, ID. Dinov, M. Burns, R. Cassidy, J. Nanni, Y. Yuan, T. Chen, A. Sinha and B. Pandian, "Medical Data Sharing and Synthetic Clinical Data Generation – Maximizing Biomedical Resource Utilization and Minimizing Participant Re-identification Risks", *npj Digital Medicine. 8, 526*, (2025).

- **Y. Liu**, C. Wang and W. Sun, "A Fast and Effective Classification Method for Missing Data", *International Symposium on Advances in Informatics, Electronics and Education*, (2022).

RESEARCH EXPERIENCE

Research Assistant (Supervisor: Brendon Watson) — Ann Arbor, MI, USA
Watson Neural Lab, U-M (joint with L'Etoile Lab, UCSF) — *Aug. 2024-Apr. 2025*
- Extracted and partitioned large .nd2 files using Matlab and conducted tracking on neurons of interest.
- Utilized image convolution techniques in Python to pre-process and analyze nematode video files of over 30 GB.
- Deployed nonparametric methods to identify sleeping modes of C. elegans using fluorescence curves with over 150 features.

Research Assistant (Supervisor: Ivo Dinov and Simeone Marino) — Ann Arbor, MI, USA
Statistics Online Computational Resource (SOCR) Lab, U-M — *Jan. 2024-Feb. 2025*
- Contributed to the DataSifter Project focusing on balancing data utility and privacy for complex biomedical data.
- Developed and deployed over 20 privacy and utility metrics for obfuscated EHR/clinical data across various data types such as time series/tabular/longitudinal data.
- Generated quality and diagnostic reports for both privacy and utility metrics for 5 levels of synthetic data.
- Embedded the package functionality in RShiny web apps.

Research Assistant (Supervisor: Shuhang Guo) — Beijing, China
School of Information, CUFE — *Jan. 2021-Nov. 2021*
- Implemented Apache Superset based on Python/MySQL to design visualization interface and dashboards.
- Integrated deep-learning-based financial crisis forecast models with a three-tier financial index system.
- Conducted digital transformation development research and data analysis reports on over 30 indicators.

PROFESSIONAL EXPERIENCE

Software Test and Algorithm Engineer Intern — Beijing, China
Beijing HiRain Technologies Co., Ltd. — *Oct. 2022-Feb. 2023*
- Completed research on more than 10 ETL tools and algorithm research on graph databases like Neo4j and Nebula.
- Maintained intelligent drive knowledge graph based on Kettle.
- Led the design of road topology generation algorithms, and performed road network annotation and quality screening on more than 600 satellite maps.

Analyst Intern — Beijing, China
China International Capital Co., Ltd. — *Sept. 2022-Oct. 2022*
- Conducted due diligence on target companies, their industries, and the entire supply chain.
- Researched real industries such as and contributed to meeting minutes and industry reports.
- Compiled the last 3 years of annual report data of 4 companies and built DCF model respectively.

Teaching

STAT 5101: Statistical Methods for Data Analysis, ISU, Lab Instructor	*2025 Fall*
Linear Algebra, CUFE, Tutor	*2021 Spring*

Awards

Social Practice Excellence Scholarship, CUFE	*2023 Spring*
Academic Research and Innovation Excellence Scholarship, CUFE	*2022 Fall*
Social Practice Excellence Scholarship, CUFE	*2021 Fall*

Competitions

College Innovation and Entrepreneurship Training Program, National Excellent Project	*Jun. 2022*
Mathematical Contest in Modeling, Honorable Mention	*Feb. 2022*
Contemporary Undergraduate Mathematical Contest in Modeling, Second Prize of Beijing	*Sept. 2021*

Services

MDS Program Ambassador, U-M Department of Statistics	*May 2024*
Minister of the Publicity Department, CUFE Math Club	*Sept. 2020 – Jan. 2021*

Technical Skills

Programming: R (ggplot2, dplyr, tidyverse, reticulate, rpy2, rshiny, etc.), Python (Pandas, NumPy, Matplotlib, Scikit-learn, Seaborn, TensorFlow, PyTorch, etc.), Matlab, Java, SAS, Stata, Scala, etc.
Frameworks: Hadoop, Spark, etc.
Miscellanea: MySQL, LaTeX, Markdown, etc.