Hongtao Hao

+1 (608) 982-2209 | Madison, WI | hongtaoh@cs.wisc.edu | github.com/hongtaoh | hongtaoh.com

[An avid blogger/educator with more than 100 tutorials on coding, ML, and statistics published on my personal website.]

Education

University of Wisconsin-Madison Madison, WI
PhD in Computer Sciences *Sept 2021 — Dec 2025*
- Research: Machine Learning/AI for Healthcare; LLM applications in Data Science

Experience

Research Intern (LLM & HCI) Jun 2023 — Aug 2023
Robert Bosch LLC *Sunnyvale, CA*
- Prototyped an LLM-powered data analysis interface (Python, FastAPI, Streamlit) enabled by prompt engineering.
- Built a scalable pipeline to systematically evaluate LLM outputs.

Publications & Projects

1st Author • **Joint Progression Model (JPM) for Mixed Pathologies** *Under Review (ML4H 2025)*
- Boosted the accuracy of disease progression modeling for mixed pathologies by 28% compared to SOTA.

1st Author • **Bayesian Event-Based Models (BEBMs) for Subtypes** *Under Review (ML4H 2025)*
- Innovated on EBMs to handle heterogeneous patient subgroups, improving performance by 27% compared to SOTA.

1st Author • **Stage-Aware Event-Based Modeling (SA-EBM) for Disease Progression** *MLHC 2025* (Link)
- Improved disease progression modeling accuracy by 40% compared to four benchmark algorithms.

4th Author • **Investigating Interaction Modes and User Agency in Human-LLM Collaboration for Domain-Specific Data Analysis** *CHI Extended Abstracts, 2024* (Link)
- Studied interaction paradigms and user agency for effective human-LLM collaboration in domain-specific data analysis.

3rd Author • **Can LLMs Infer Domain Knowledge from Code Exemplars?** *IUI Companion, 2024* (Link)
- Conducted a preliminary study on the ability of LLMs to extract domain-specific knowledge from code exemplars.

Maintainer • **LLM-Powered Movie Recommendation** (Link)
- Built a scalable recommender system using retrieval-augmented generation (RAG) and the TMDB dataset.
- Integrated LangChain retrievers and vector stores to enhance personalization.

Maintainer • **Chinese Women's Partner Choice Data Lab** (Link)
- Developed an interactive one-page Svelte app that estimates how many men fit user-selected partner criteria.

1st Author • **Thirty-Two Years of IEEE VIS: Authors, Fields of Study and Citations** *IEEE VIS, 2022* (Link)
- Analyzed authorship, fields, and citation networks across three decades of IEEE VIS publications.

Maintainer • **Deep Learning for American Time Use Survey (ATUS)** (Link)
- Predicted time-use patterns from demographics using TensorFlow.js; deployed as an interactive web app.

Research Assistant • **COVID-19 Fatality Rate** (Link)
- Automated a Covid-19 data dashboard with CI/CD and improved visualization design, ranking among the Top 10 Most Liked notebooks on Observable (2020).

Maintainer • **pysaebm: Open-Source Python Package** (Link)
- Developed and maintain an open-source Python package for event-based modeling, with extensive documentation.

Skills

- Programming Languages: Python, JavaScript/TypeScript, HTML/CSS, Java, Bash, R, Julia, SQL
- ML: PyTorch, scikit-learn, Pandas, NumPy, Numba, LangChain, JAX, Prompt engineering (Tree-of-Thoughts, instruction tuning)
- Technologies: FastAPI, Next.js, React, D3.js, Svelte, Tailwind CSS, MongoDB, UNIX, Docker, Google Cloud Platform, Azure, CI/CD, HPC, LLMOps (evaluation pipelines, scalable deployment)