

Jiande Wu, Ph.D.

Applied AI / ML Scientist | LLM Systems | Public Health Modeling

jiandewu@gmail.com | [Website](#) | [GitHub](#) | [LinkedIn](#) | [Google Scholar](#) | [ORCID](#)

Research Profile

Applied AI/ML scientist developing predictive models, LLM applications, and reproducible workflows for biomedical, EHR, surveillance, and digital trace data. Research focuses on infectious disease forecasting, LLM and AI-agent workflows, computational genomics, and scalable scientific computing.

Selected Projects

Transformer-Based Multi-Source COVID-19 Forecasting | PyTorch | [Paper](#) | [Code](#)

- Co-developed and retrospectively evaluated a Transformer-based, multi-source forecasting model across 46 South Carolina counties and three Omicron waves; achieved 74.0% to 82.6% median county-level Percentage Agreement (PA) for one-week case forecasts and 83.2% to 86.3% for one-week death forecasts.
- Demonstrated 34% higher PA, 64% lower RMSE, and 68% lower MAE than a persistence baseline for five-week forecasts in the reported aggregate comparison.

resplens | Python, MCP | [Code](#)

- Reliability-aware AI toolkit for US respiratory surveillance data, with MCP tools and evaluated document retrieval.

AI Frontier Digest | Python (standard library only), Gemini API, GitHub Actions | [Site](#) | [Code](#)

- Built an automated pipeline that collects AI news every four hours from 33 sources (31 researcher, lab, and news RSS/Atom feeds plus Hacker News and Hugging Face Daily Papers, filtered by points/upvote thresholds) and uses an LLM to write Chinese titles, summaries, categories, and 1–5 importance ratings.
- Constrained LLM output with per-batch JSON Schemas (enum-restricted item IDs and categories), fell back to plain JSON mode when the API rejected the schema, and re-validated every field locally; items from a failed batch stay unseen and are retried on the next run.
- Used the LLM to merge reports of the same event across sources and to write longer explanations from article full text, treated as untrusted content; daily picks are ranked with per-source caps and a coverage bonus for multi-source events.
- Runs on scheduled GitHub Actions with unit tests gating every update and a monthly feed-health check that fails on dead or stale feeds.

Epidemic Forecasting Agents | Python, AutoGen, Ollama

- Prototyped an AutoGen multi-agent team (seven role-based LLM agents and a human-in-the-loop proxy in a round-robin group chat) that drafts requirements, data contracts, forecasting architecture, and a release plan for an epidemic forecasting product; runs on hosted OpenAI-compatible APIs or local Ollama models.
- Compared LangGraph, AutoGen, CrewAI, and LlamaIndex Agents on reproducibility, auditability, stability, interpretability, data efficiency, and engineering cost.

ObsAwareData | Python | [Data](#) | [Pipeline](#)

- Weekly COVID-19, influenza, and RSV surveillance data for the 50 US states and DC (NHSN, NSSP, NWSS, Census), with reliability flags and source-release provenance, built from an archive of CDC data releases.

Technical Skills

- **Programming:** Python, R, SQL, C/C++/STL, MATLAB, Linux shell scripting
- **Machine Learning:** PyTorch, scikit-learn, XGBoost, Transformer-based modeling, predictive modeling, deep learning, feature engineering, evaluation, validation, and fine-tuning
- **LLM and AI Systems:** Prompt engineering, retrieval-augmented generation (RAG), tool/function calling, structured outputs, MCP, workflow orchestration, and AI-agent prototyping
- **Biomedical and Public Health Data:** EHR data, surveillance data, digital trace data, genomic and epigenomic data, data integration, quality control, and reproducible analysis
- **Scientific Computing:** MPI, CUDA, GPU computing, parallel and distributed algorithms, Linux clusters, mathematical modeling, and simulation

Professional Experience

Data Manager / Data Scientist

Jul 2023 – Present

Center for Public Health Modeling and Response, Clemson University

- Manage extraction, integration, quality control, and documentation for approximately 100,000 EHR records across five data sources and 46 counties or regions.
- Develop end-to-end AI/ML pipelines and lead one-week infectious disease prediction and surveillance workflows, including early warning systems for outbreak detection and response.
- Co-developed the multi-source COVID-19 forecasting model published in *Epidemics* (see Selected Projects).
- Develop LLM-assisted and AI-agent workflows for data documentation, knowledge extraction, data quality review, querying, preprocessing, analysis automation, workflow orchestration, model monitoring, and report generation.
- Design reusable prompt patterns and evaluation approaches covering accuracy, usefulness, relevance, and safety for LLM-assisted research workflows.

Senior Bioinformatics Specialist

Oct 2017 – Jul 2023

LSU Health Sciences Center

- Developed and maintained bioinformatics and computational genomics workflows for large-scale biomedical research data analysis.
- Designed and optimized algorithms and pipelines for genomic data processing, quality control, statistical analysis, and computational interpretation.
- Applied parallel and distributed computing methods to improve the scalability and performance of data-intensive scientific workflows.
- Supported interdisciplinary research teams with reproducible analysis, data management, technical documentation, and interpretation of computational results.

Postdoctoral Researcher

Jul 2016 – Jun 2017

Clemson University

- Conducted large-scale statistical analysis using Stata to support behavioral and public health research.
- Designed and implemented the Behavioral Survey Measurement & Analysis Tool (BSMAT), a PHP/MySQL application for research data collection, management, and analysis.

Research Assistant, Department of Electrical Engineering

Jan 2008 – Dec 2014

University of New Orleans

- Designed, implemented, and evaluated computationally efficient particle-filter algorithms for target tracking.
- Developed parallel and distributed algorithms using MPI, CUDA, C/C++, Linux clusters, and GPU computing environments.
- Teaching Assistant for Probability Theory, Numerical Methods, and Mathematical Modeling.

Selected Publications

- **Wu, J.**, Tanim, S., Woo, M., Ahammed, T., Bleichrodt, A. M., & Rennert, L. (2025). [A deep learning approach for enhancing pandemic prediction: A retrospective evaluation of transformer neural networks and multi-source data fusion for infectious disease forecasting](#). *Epidemics*, 53, 100865.
- **Wu, J.**, Zhang, L., Kuchi, A., Otohinoi, D., & Hicks, C. (2022). [CpG site-based signature predicts survival of colorectal cancer](#). *Biomedicine*, 10(12), 3163.
- **Wu, J.**, & Hicks, C. (2021). [Breast cancer type classification using machine learning](#). *Journal of Personalized Medicine*, 11(2), 61.
- **Wu, J.**, Mamidi, T. K. K., Zhang, L., & Hicks, C. (2019). [Integrating germline and somatic mutation information for the discovery of biomarkers in triple-negative breast cancer](#). *International Journal of Environmental Research and Public Health*, 16(6), 1055.
- **Wu, J.**, Mamidi, T. K. K., Zhang, L., & Hicks, C. (2019). [Deconvolution of the genomic and epigenomic interaction landscape of triple-negative breast cancer](#). *Cancers*, 11(11), 1692.
- Jilkov, V. P., & **Wu, J.** (2015). [Efficient GPU-accelerated implementation of particle and particle flow filters for target tracking](#). *Journal of Advances in Information Fusion*, 10(1), 73–88.

Education

- **Ph.D., Electrical Engineering**, University of New Orleans, 2014
- **M.S., Electrical Engineering**, University of New Orleans
- **M.S. and B.S., Computer Science and Engineering**, North China Electric Power University