Jinge **Wang**

DATA SCIENTIST · ML & AI

📱 (+1) 304-376-8558 | ✉ jinge0829@gmail.com | 🏠 JingeW.github.io | 🐙 jingew | 💼 JingeW

Summary

Data Scientist specializing in applied AI and multimodal analytics, with a focus on transforming real-world, unstructured healthcare data into validated signals for downstream analysis and decision-making. Experienced in building end-to-end NLP and OCR pipelines, designing rigorous evaluation frameworks, and integrating model outputs into analytical and product workflows through close collaboration with domain stakeholders.

Core Competencies

Engineering: Python, R, OCR, SQL, AWS, Git, Docker, Jupyter, Linux, LaTeX, REST/Flask prototyping

AI & ML: LLMs, NLP pipelines, multimodal modeling, predictive modeling, feature engineering, model evaluation

Analytics: Experiment design, metrics, error analysis, interpretability, statistical validation

Collaboration: Cross-functional delivery, stakeholder communication, prototyping, mentoring, documentation

Work Experience

Polygon Health Analytics LLC *Remote*

DATA SCIENTIST *Oct 2024 – Sep 2025*

- Built end-to-end applied ML workflows on unstructured healthcare data, spanning extraction, modeling, evaluation, and delivery.
- Developed LLM- and OCR-based pipelines to convert clinical documents into structured features for downstream analysis.
- Integrated newly derived features into multimodal datasets, expanding coverage for genomic, imaging, and clinical analyses.
- Designed and evaluated text classification systems, improving accuracy by 30% through algorithm refinement and error analysis.
- Translated ambiguous product and domain requirements into measurable analytical objectives and validated outputs.
- Iterated on models and data pipelines based on evaluation results and stakeholder feedback to improve real-world usability.

West Virginia University *Morgantown, WV*

POSTDOCTORAL FELLOW *Oct 2023 – Sep 2024*

- Applied prompt engineering (few-shot, CoT) to improve AI-based melanoma diagnosis accuracy by 25%.
- Designed reproducible evaluation frameworks to assess accuracy and failure modes in multimodal clinical AI systems.
- Analyzed model limitations and generalization gaps to inform deployment considerations in clinical AI settings.
- Collaborated with interdisciplinary teams to convert domain problems into tractable ML tasks, datasets, and metrics.

West Virginia University *Morgantown, WV*

GRADUATE RESEARCH ASSISTANT *Jan 2018 – May 2023*

- Developed machine learning and deep learning pipelines for large-scale multimodal data analysis.
- Performed feature extraction, feature manipulation, and interpretability analyses to understand model behavior and validate results.
- Applied rigorous statistical methods and experimental design to validate models and ensure reproducibility.

Selected Projects

MAGIC — Multimodal Analysis of Genomics, Imaging and Clinical Data

- Built an OCR- and LLM-based pipeline to extract structured pathology indicators from noisy clinical PDF reports.
- Integrated pathology-derived features into GDC-linked datasets, adding a new informative modality to existing genomic and clinical data.
- Established data foundations for downstream multimodal modeling, including survival analysis and clinical research workflows.
- Prepared structured outputs for interoperability and visualization, including cBioPortal-compatible exports.

Vaccine Vibes — LLM optimized Vaccine Newsletter

- Developed and evaluated LLM-based text classification pipelines for healthcare-related news and content.
- Improved classification accuracy by 30% through prompt engineering, data curation, and systematic error analysis.
- Designed benchmarking datasets and evaluation workflows to support iterative model improvement.

iSMILE — AI-Enabled Mobile Health Prototype for SLE Patients

- Designed an end-to-end mobile health prototype integrating user profiles, EHR-derived data, and AI-driven risk assessment.
- Implemented data flows and analytical logic to support personalized insights and community-oriented features.
- Modeled EHR and SDoH-aligned data structures to support barrier identification and resource recommendation workflows.

Dermoscopy — Based Melanoma Classification

- Developed and evaluated multimodal AI systems for melanoma diagnosis using dermoscopic images and clinical context.
- Improved melanoma diagnosis accuracy by 25% via prompt engineering (few-shot, CoT) and systematic error analysis.
- Assessed limitations and safety risks (e.g., hallucinations and overgeneralization) to inform deployment considerations in medical AI.

Education

West Virginia University   *Morgantown, WV*

Ph.D. in Computer Science   *Aug 2016 – May 2023*

- Specialization in Artificial Intelligence, Data Analysis, and Signal Processing.

West Virginia University   *Morgantown, WV*

M.S. in Statistics   *Aug 2013 – Dec 2015*

- Focus on Statistical Modeling, Regression, and Experimental Design.

Anhui University of Finance and Economics   *Anhui, China*

B.M. in Accounting   *Sep 2007 – Jul 2011*

- Completed foundational coursework in management and accounting.

Publications

- **J Wang**; et al. *"ChatGPT-Polished Writing Boosts the Risk of Human-Authored Manuscripts Being Miscredited as AI-Generated."* JAADi. (2025)
- R Cao; **J Wang**; et al. *"Feature-based encoding of face identity by single neurons in the human amygdala and hippocampus."* Nat. Hum. Behav. (2025)
- **J Wang**; et al. *"Limitations and risks of custom GPTs in dermatology. Comment on 'ReconGPT: A novel artificial intelligence tool and its potential use in post-Mohs reconstructive decision-making.'"* JAAD. (2025)
- **J Wang**; et al. *"Preliminary evaluation of ChatGPT model iterations in emergency department diagnostics."* Sci. Rep. (2025)
- Z Feng; G Hu; B Li; **J Wang**. *"Unleashing the power of ChatGPT in finance research: opportunities and challenges"* FIN. (2025)
- **J Wang**; et al. *"Adapting ChatGPT for Color Blindness in Medical Education."* Ann Biomed Eng. (2025)
- **J Wang**; G Hu. *"Boosting GPT-4V's accuracy in dermoscopic classification with few-shot learning. Comment on 'Can ChatGPT vision diagnose melanoma? An exploratory diagnostic accuracy study.'"* JAAD. (2024)
- **J Wang**; et al. *"Bioinformatics and Biomedical Informatics with ChatGPT: Year One Review."* QB. (2024)
- **J Wang**; et al. *"Scientific Figures Interpreted by ChatGPT: Strengths in Plot Recognition and Limits in Color Perception."* NPJ Precis. Oncol. (2024)
- S K Valicharla; **J Wang**; et al. *"Morning Glory Flower Detection in Aerial Images Using Semi-Supervised Segmentation with Gaussian Mixture Models."* Agric. Eng. (2024)
- **J Wang**; et al. *"A critical period for developing face recognition."* Patterns. (2024)
- R Cao; **J Wang**; et al. *"Neural mechanisms of face familiarity and learning in the human amygdala and hippocampus."* Cell Reports. (2024)
- **J Wang**; R Cao; et al. *"Face identity coding in the deep neural network and primate brain."* Commun. Biol. (2022)
- X Xu; X Xiong; **J Wang**; X Li. *"Deformable kernel convolutional network for video extreme super-resolution."* ECCV Wksp. (2020)

Certificates

2023	**Prompt Engineering for ChatGPT,** Vanderbilt University	
2023	**Python Essentials for MLOps,** Duke University	
2023	**Supervised Machine Learning: Regression and Classification,** DeepLearning.AI, Stanford University	
2023	**Introduction to Large Language Models,** Google Cloud	
2023	**Introduction to Healthcare,** Stanford University	
2015	**SAS Certified Advanced Programmer for SAS 9,** SAS Institute	
2015	**SAS Certified Base Programmer for SAS 9,** SAS Institute	

Honors & Awards

2024	**Oral presentation,** TRCCC
2020	**Advances in Image Manipulation workshop and challenges Runner-Up Award,** ECCV AIM