Yanfei Wang

yanfeiwang29@gmail.com |Google Scholar| 25 SW 5th Terrace, Gainesville, FL 32601

AREAS OF SPECIALIZATION

My research sits at the intersection of causal inference and longitudinal clinical modeling, using real-world data such as electronic health records (EHRs) and claims to answer decision-focused questions about disease risk, progression, and treatment strategies. Within this theme, my work centers on three core areas: (1) **Causal AI for real-world evidence**, developing principled estimation frameworks that combine modern representation learning with target trial emulation, time-varying treatment definitions, and causal survival estimands to produce clinically interpretable effect measures; (2) **interpretable and trustworthy machine learning for longitudinal healthcare data**, emphasizing transparent model structure, uncertainty quantification, and evaluation under dataset shift, missingness, and irregular observation, so that predictive signals align with actionable clinical decisions; and (3) **bias-aware, deployment-oriented methodology**, focusing on diagnosing and mitigating confounding, selection, and censoring biases, assessing transportability across health systems, and translating model outputs into decision interfaces that make assumptions, limitations, and intended use explicit.

Education

PhD of Biomedical Informatics 2019-2023
University of Texas Health Center Houston
- Advancing Precision Medicine: Unveiling Disease Trajectories, Decoding Biomarkers, and Tailoring Individual Treatments

Master of Statistics 2016-2018
George Washington University

Bachelor of Mathematics 2011-2015
University of Colorado Denver

Peer-Reviewed Journal Articles

- **Wang Y.**, Wang Q, Zhou M, et al. Integration of Genetic and Imaging Data for Alzheimer's Disease Diagnosis and Interpretation. *Advanced Science*. 2025;12(41): e07629.
- **Wang Y.**, Guo Y, Tan AC, et al. A real-world cohort study of immune-related adverse events in patients receiving immune checkpoint inhibitors. *npj Precision Oncology*. 2025;9(1):346.
- **Wang Y.**, Niu S, Doonan BP, et al. Real-world analysis of cardiovascular adverse events and risk factors after immune checkpoint inhibitor therapy. *Cardio-Oncology*. 2025;11(1):107.
- **Wang Y.**, Zhou, X., et al. Revealing Chronic Disease Progression Patterns Using Gaussian Process for Stage Inference, *Journal of the American Medical Informatics Association*, 2023.
- **Wang Y**., Zhou, X., et al. Phenotype-Genotype analysis of Caucasian patients with high risk of osteoarthritis. *Front Genet*. 2022 Aug 29; 13:922658.
- **Wang Y**., Zhou, X., et al. Causal Discovery in Radiographic Markers of Knee Osteoarthritis and Prediction for Knee Osteoarthritis Severity with Attention-Long-Short-Term Memory. *Front Public Health*. 2020 Dec 18; 8:604654.
- Feng, Y., Yi, J., Wen, J. **Wang Y.**, et al. COV2Var, a function annotation database of SARS-CoV-2 genetic variation. *Nucleic Acids Research*, 2023: p. gkad958.
- Wu S, Zhang J, **Wang Y**., et al. metsDB: a knowledgebase of cancer metastasis at bulk, single-cell and spatial levels. *Nucleic Acids* Research, 2024: gkae916.
- Liu X, Wang Q, Zhou M, **Wang Y**., et al. DrugFormer: Graph-Enhanced Language Model to Predict Drug Sensitivity. *Advanced Science*, 2024;11(40):2405861.
- Luo, R., Chyr, J., Wen, J. **Wang Y.**, et al. A novel integrated approach to predicting cancer immunotherapy efficacy. *Oncogene* 42, 1913–1925 (2023).
- Deng Y., You L., **Wang Y.**, Zhou X. A Coarse-to-Fine Framework for Automated Knee Bone and Cartilage Segmentation Data from the Osteoarthritis Initiative. *J Digit Imaging*. 2021 Aug;34(4):833-840.
- You L., Deng, Y., **Wang Y.**, et al. A novel sagittal craniosynostosis classification system based on multi-view learning algorithm. *Neural Comput & Applic* 34, 14427–14434 (2022).

- Wang Q, Feng Y, **Wang Y.**, et al. AntiFormer: graph enhanced large language model for binding affinity prediction. *Briefings in Bioinformatics*, 2024; 25(5): bbae403

Scientific Presentations

- Abstract "Unraveling the role of CTLA-4 variants in acute kidney injury after ICI treatments." at ASCO Annual Meeting 2024
- Poster presentation at University of Florida AI Symposium 2025
- Guest Speaker at GMS6804 Translational Bioinformatics 2025

Research Experience

Assistant Scientist 2026-Present

- **Causal AI & Target Trial Emulation**: Led development of causal AI methodologies for longitudinal EHR and claims data, designing target trial emulation pipelines for time-updated, add-on, switching, and combination strategies, with explicit exposure definitions (overlap, lag, adherence) for drug effect estimation.
- **Causal Discovery and Disease Progression Modeling**: Developed interpretable longitudinal disease progression models that align predictive outputs with causal questions and decision contexts, emphasizing transparent feature construction, uncertainty quantification, calibration, and clinically meaningful explanations rather than black-box performance.
- **Decision-Capable Digital Twin Research**: Advanced decision-capable digital twin methodology for state updating and strategy evaluation under routine-care data constraints, incorporating observation-process modeling, feedback loops, and explicit decision interfaces that communicate assumptions, uncertainty, and failure modes.
- **Grant Writing, and Team Science**: Wrote and contributed to competitive grant proposals, developed analytic sections, power and feasibility narratives, and data harmonization plans, and coordinated multi-disciplinary collaborations among clinicians, statisticians, informaticians, and data engineers.

Postdoctoral Associate 2024–2025

- **Statistical Modeling & Machine Learning**: Developed mathematical frameworks, including hybrid scoring, Bayesian zero-inflated models, and variational inference, to analyze sparse, high-dimensional biomedical data. Applied Bayesian hierarchical inference, MLE, EM, and posterior sampling for robust estimation and uncertainty quantification.
- **Causal Inference & Treatment Effect Estimation**: Combined traditional methods (propensity scores) with neural network–based counterfactual modeling (MLPs) to estimate individualized treatment effects and uncover real-world clinical heterogeneity.
- **EHR Data Standardization**: Led the mapping of UF Health clinical data to NCI-compliant standard elements, building scalable SQL-based ETL pipelines to enable interoperable, oncology-focused analytics across systems.
- **Genetic Risk & Association Analysis**: Performed GWAS and polygenic risk score (PRS) analyses using biobank-scale datasets (e.g., All of Us, UK Biobank) to quantify genetic susceptibility and integrate findings into clinical risk modeling.
- **Survival & Time-to-Event Analysis**: Applied Kaplan-Meier, log-rank tests, and Cox regression to evaluate treatment outcomes and stratify patient risk using longitudinal genetic and clinical data.

Graduate Research Assistant 2019-2023

- **Predictive Modeling & Causal Discovery:** Developed deep learning models (e.g., attention-based LSTM) and causal discovery methods (e.g., PC, FCI, and Granger causality) to predict chronic disease severity and identify early pathological signals using longitudinal EHR data from national cohorts.
- **Modeling Disease Heterogeneity:** Applied Gaussian Process models and interpretability techniques (SHAP, LIME) to characterize individual variability in disease expression across large-scale datasets (ABCD, OAI, TCGA), informing precision health strategies.
- **Genomic Analysis & Causal Inference:** Conducted GWAS and Mendelian randomization to uncover genetic risk factors and causal pathways underlying chronic diseases, supporting the development of targeted interventions.

- **Counterfactual Modeling & Domain Adaptation:** Designed a dual-network architecture for individualized treatment effect estimation, integrating representation learning and domain adaptation to improve causal inference from real-world COVID-19 treatment data.

Scientific Appointments

- Ad hoc reviewer for Scientific Reports; BMC Cancer; IEEE ICH
- Member of AMIA
- Programme Committee for International Digital Public Health Conference 2026 and IEEEICHI2026
- Reviewer for 2026-27 PHHP (Public Health and Health Professions) AI PhD Fellowship