Anni(Annie) Zhou

+1-470-601-1243 | azhou60@gatech.edu | Annie's Website

LinkedIn | Github | Google Scholar

Summary

Passionate researcher with 8+ years of experience in predictive and statistical modeling, machine learning, and data analysis. Enthusiastic about advancing AI methodologies for real-world applications. Proven track record in research, publication, and collaboration across diverse teams.

Skills

- **Programming Languages & Tools:** SQL, PostgreSQL, Python, C/C++, R, Matlab, Git
- **Data Science & Machine Learning:** Scikit-learn, TensorFlow, PyTorch, Pandas, NumPy, JAX
- **Specialized Area:** Large Language Models, Generative AI, Computer Vision, AI Security, Statistical Modeling, Deep Learning, Reinforcement Learning, Machine Learning, Data Analytics, Recommendation System

Experience

- **Computer Vision & Image Processing** *Jan. 2025 - Present*
  Research Assistant, mentored by Dr. Yuchen Liu North Carolina State University
  - Designing and implementing a deep learning framework using Convolutional LSTM and encoder-decoder architecture to predict plant growth at the pixel level, modeling complex spatio-temporal leaf dynamics from sequential RGB and labeled images.
  - Developing a **multi-task learning** model that jointly performs future plant image forecasting and environmental condition classification (e.g., drought vs. control), enhancing prediction accuracy through shared spatiotemporal features.
  - Pioneering a novel Pixel-of-Interest (POI) based bi-polar regression metric, enabling physically meaningful evaluation of growth predictions beyond conventional image similarity metrics like SSIM and MSE.
  - Conducting comparative analysis across state-of-the-art video prediction models (ConvLSTM, PredRNN, PhyDNet, SimVP), achieving top performance on the KOMATSUNA dataset through robust preprocessing, segmentation, and augmentation pipelines.
- **Reinforcement Learning & Deep Learning & Uncertainty Quantification** *Aug. 2018 - Jan. 2025*
  Research Assiatant, mentored by Dr. Raheem Beyah and Dr. Rishikesan Kamaleswaran Georgia Tech
  - Conducted a comprehensive series of research to improve AI for healthcare.
  - Proposed a series of novel advanced **trustworthy & robust machine learning & deep learning** methodologies with uncertainty quantification.
  - Analyzed and preprocessed big healthcare dataset under guidance of clinicians.
  - **First authored three peer-reviewed publications** detailing the methodologies and findings.
- **Privacy-preserving Distributed Online Learning Recommendation System** *2016-2018*
  Research Assistant, mentored by Dr. Pan Zhou at Wuhan National Laboratory for Optoelectronics (WNLO) China
  - Designed a privacy-preserving collaborative recommendation system, achieving 98.06% accuracy using online learning methodologies.
  - Addressed privacy challenges through Differential Privacy techniques, ensuring user data confidentiality.
  - Conducted rigorous theoretical analysis of regret bounds, achieving sublinear regret convergence to the optimal policy.
  - Authored **three peer-reviewed publications** detailing the methodologies and findings.

Education

- **Georgia Institute of Technology** *Aug. 2018 - Jan. 2025*
  Ph. D. in Electrical and Computer Engineering (Defended Jan 2025, Diploma Expected May 2025) Atlanta, United States
- **Huazhong University of Science and Technology** *Sept. 2014 - Jun. 2018*
  B.S. in Electrical and Computer Engineering Wuhan, China
- **Fine-Tuning for LLMs** *Feb. 2025 - Mar. 2025*
  Certificate LinkedIn Learning

Projects

- **Fine-Tuning Large Language Models for NLP Tasks [Code]** *Jan. 2025 - Present*
  Tools: Python, Transformers, PyTorch, TensorFlow, LoRA, Hugging Face, etc.
  - Developed and implemented advanced NLP models by fine-tuning large language models (Flan-T5, DistilBERT) for various tasks, including text summarization, sentiment analysis, translation, and question answering.
  - Employed Low-Rank Adaptation (LoRA) technique to efficiently fine-tune transformer models, significantly reducing computational overhead and training time.
  - Integrated transfer learning and prompt-engineering methodologies, demonstrating improved performance on benchmark datasets such as SST-2 and SQuAD v2.
  - Built interactive chatbot prototypes using custom fine-tuned models to enhance conversational accuracy and fluency.

- **Deep Reinforcement Learning & Uncertainty Quantification [Code]** *2023 - 2024*
  Tools: Deep Learning, Reinforcement Learning, Data Analytics, Imputation and Preprocessing, etc. Georgia Tech
  - **Designed and built from scratch** a novel reinforcement learning algorithm integrating conformal prediction and reinforcement learning, improving adaptive decision-making under uncertainty for early sepsis prediction.
  - Proposed a linear contextual bandit extension offering faster convergence and enhanced interpretability for datasets with low variability.
  - Conducted large-scale experiments on ICU datasets (27,000+ hours of electronic **time series** health records) **using JAX**, achieving high prediction accuracy and reliability in sepsis identification and demonstrating improved patient-specific reward approximation.
  - **Collaborated with a multidisciplinary team** to translate research outcomes into actionable clinical insights.

- **Natural Language Processing & Machine Learning [Code]** *2023*
  Tools: Pytorch, Tensorflow, BERT, SVM, NLP, PCA, etc. Georgia Tech
  - Conducted an in-depth analysis of racial bias and hate speech on Twitter during the COVID-19 pandemic using natural language processing (NLP) techniques.
  - Built and compared various classification models (Logistic Regression, SVM, Random Forest, BERT) for hate speech detection on social media data.
  - Employed topic modeling (LDA) and sentiment analysis to uncover underlying themes and public sentiment trends. Audited model performance across racial groups to assess fairness and potential bias in ML-driven content moderation.
  - Demonstrated strong data preprocessing, feature engineering, and interpretability techniques within a real-world, socially impactful context.

- **Reinforcement Learning & Uncertainty Quantification [Code]** *2022 - 2023*
  Tools: Bayesian Modeling, Reinforcement Learning, Data Analytics, Imputation and Preprocessing, etc. Georgia Tech
  - **Designed and built from scratch** a novel Bayesian-based reinforcement learning framework with uncertainty quantification, achieving up to 98% coverage.
  - **Optimized parallel computing** techniques, **enhancing computational efficiency and accelerating** processing 24 times.
  - Implemented **advanced imputation methodologies** such as MICE, GANs, Random Forests, etc., to address missing clinical data, ensuring scalability and robustness of the predictive models.
  - Conducted extensive experiments on **large-scale training datasets** of up to 2,000 ICU patients, demonstrating improved patient-specific personalized model selection.
  - Worked closely with clinicians to identify critical features and develop clinically relevant preprocessing workflows.

- **Reinforcement Learning & Machine Learning [Code]** *2020 - 2021*
  Tools: scikit-learn, Reinforcement Learning, Feature Engineering, Data Analytics, Imputation and Preprocessing, etc. Georgia Tech
  - Achieved 90.82% accuracy using a novel online learning-based prediction framework combining various machine learning algorithms such as XGBoost, Random Forest, and Logistic Regression.
  - Processed and analyzed large-scale electronic health records for 40,336 patients using PostgreSQL, ensuring data integrity and completeness.
  - Provided explainable results through AI tools such as **LIME** and **SHAP**, facilitating clinician acceptance and trust in predictive outcomes.

Professional Activities & Service

Peer Reviewer of *IEEE Global Communications Conference*, and *Expert Systems With Applications*.

Teaching

- **CS/ECE 6263 Introduction to Cyber-Physical System Security** *2019 - 2022*
  Graduate Teaching Assistant, Georgia Institute of Technology [Course Website]