Hua Yang

(+1) 984-373-9587 | hyang45@ncsu.edu | Hua Yang | HuaYang.me

Education

NC State University, Raleigh, NC **GPA: 3.7/4.0**
Ph.D., Computer Science, Aug 2024 – Dec 2027
Relevant Coursework: Software Engineering, Algorithms, Generative AI for Software Engineering

South China University of Technology **GPA: 4.0/4.0**
M.S., Computer Science, Sep 2021 – Jun 2024
Relevant Coursework: Machine Learning Algorithms, Software Testing

Central South University **GPA: 4.0/4.0**
B.S., Electronic Engineering, Sep 2017 – Jun 2021
Relevant Coursework: Object-Oriented Programming, Networks, Data Structures and Algorithms

Skills

Languages: Python, Java, C#, CSS, HTML, JavaScript
Tools/Frameworks: Git, PyTorch, HuggingFace, CI/CD, AWS, REST API, Flask, Bootstrap
Databases and Operating Systems: MySQL, Oracle, SQLite, Windows, Ubuntu
Computer Networks: TCP/IP, SSH, Socket Programming, DNS, HTTP,

Work Experience

Machine Learning Intern
AI Large Model for Intelligent Cognition Center, Pazhou Lab (Guangzhou) *May 2022 – May 2023*

- Engineered a real-time multimodal sleep disorder detection system, achieving top-5 ranking among 171 national teams and securing a $10,000 award.
- Improved multi-domain model accuracy (55% → 59%) across 7 datasets (text, networks, multimodal) while achieving a 110× runtime speedup, setting a new state-of-the-art benchmark.

Projects

Causal Study on Membership Inference Attacks in Code LLMs *Sep 2024 – May 2025*

- **Tech: Python, PyTorch, Hugging Face, CUDA, Causal Inference**
- Evaluated semantically equivalent code transformations (SECT) as evasion strategies against membership inference (MI) in Code LLMs.
- Fine-tuned CodeGPT, StarCoder, and CodeGen on transformed datasets, showing minimal accuracy loss (less than1.5%) but significant MI reduction (up to 10.19%).
- Applied causal analysis to confirm variable renaming as the strongest factor in disrupting MI.

Causal Analysis of PII Leakage in Large Language Models for Code *Sep 2024 – May 2025*

- **Tech: Hugging Face, OpenAI API, GPT-5 HTML, Python**
- Revealed heterogeneous privacy risks in code LLMs through training-dynamics and causal analysis.
- Built PII dataset from real repositories and fine-tuned representative LLMs.

Movie Recommendation Web App Integrating LLMs *Sep 2024 – Dec 2024*

- **Tech: CSS, HTML, CI/CD, SVM, CNN, XGBoost, MySQL**
- Led a 4-person team to develop a web app for movie recommendations, mentoring junior developers and ensuring **smooth collaboration**.
- Designed and optimized machine learning models (SVM, CNN, XGBoost) for AI-driven movie recommendations, including issue tracking, pull requests, and CI/CD automation.

Publications

- **Yang, H.**, et al. *A Causal Perspective on the Role of Training Dynamics for Interpreting Privacy Risks in Code Models Submitted to FSE 2026* (under review).
- **Yang, H.**, et al. *How Do Semantically Equivalent Code Transformations Impact Membership Inference on LLMs for Code? Submitted to ICSE 2026* (under review).
- **Yang, H.**, Chen, C. P., Chen, B., & Zhang, T. (2024). *Improving the Interpretability through Maximizing Mutual Information for EEG Emotion Recognition. IEEE Transactions on Affective Computing.* (Top Journal in AI for Science).
- **Yang, H.**, Chen, C. P., Chen, B., & Zhang, T. (2023). *Facexplainer: Generating model-faithful explanations for graph neural networks guided by spatial information.* In *2023 IEEE BIBM* (pp. 718-725). IEEE. (Acceptance Rate: 19.5%, Explainable AI).