# Yaoning (Kevin) Yu

yyn030600@gmail.com | +1 2064842304 | Homepage | Google Scholar

## RESEARCH INTERESTS

My current research interests include Large-Language-Model (LLM)-based agents, multimodal/vision-language models, and AI security. I am eager to explore other emerging fields, especially in embodied AI, and apply these advances to solve real-world problems that deliver measurable social impact.

## EDUCATION

**University of Chicago | Applied Data Science** *Sep 2023 – Dec 2024*

**University of Washington | Mathematics** *Sep 2019 – Aug 2022*

## PUBLICATIONS/PREPRINTS

**[1] SIPDO: Closed-Loop Prompt Optimization via Synthetic Data Feedback [pdf]**
- Authors: **Y Yu**, Ye Yu, K Wei, H Wang, H Luo, Under Review

**[2] Synthetic Data-Driven Prompt Tuning for Financial QA over Tables and Documents**
- Authors: Y Yu, K Chang, Ye Yu, K Wei, H Wang, H Luo, Under Review *ICAIF 25*

**[3] Large Language Model-based Data Science Agent: A survey [pdf]**
- Authors: P Wang[+], **Y Yu**[+], C Ke[+], X Zhan, H Wang, Under Review ([+]Equal Contribution)

**[4] PREMISE: Scalable and Strategic Prompt Optimization for Efficient Mathematical Reasoning in LLMs [pdf]**
- Authors: Ye Yu, **Y Yu**, H Wang, Under Review

## RESEARCH EXPERIENCES

***Research Intern* | University of California, Los Angeles | California, CA** *June 2025 – Present*

Under Supervision of Dr. Yuan Tian
- Conducted research in security and privacy issues of AI itself, including prompt injection, jailbreak, adversarial attacks, membership inference attacks, etc.
- Developed AI-Agent system for CTF in dynamic situations.

***Research Assistant* | University of Illinois Urbana-Champaign | Champaign, IL** *April 2024 – Present*

Under Supervision of Dr. Haohan Wang
- Designed a self-prompt learning loop to optimize prompt while augmenting data.
- Developed an inverse‑perturbation defense for large vision–language models, leveraging CLIP feature inversions and attention‑map masking to neutralize image‑based jailbreak and prompt‑injection attacks.
- Conducted research in LLM-based agents to enhance the reasoning capabilities.

***Research Assistant* | University of Washington | Seattle, WA** *April 2022 – Oct 2022*

Under the supervision of Dr. George Manucharyan
- Transformed extensive datasets from 1978 to 2021, including ice concentration, sea ice velocity, ocean velocity, and atmospheric wind velocity, into the Ease Grid format
- Created visual representations of these datasets in Ease Grid format to verify data alignment with the original datasets
- Utilized ML and data interpolation to understand deep ocean activity (velocity) with interpolated datasets

## RELEVANT PROJECTS

**CredAIble** *Current*
- AI-driven underwriting document application designed to automate classification, validation, and organization of financial documents by using python (pdf2image, pdfplumber, docx, ReportLab, OCR, etc), Tauri, React, Tailwindcss, and LLM (Large Language Models)-based agents with customized privacy-preserving methods

**TradingHero AI Application [GitHub]** *Spring 2024*
- A comprehensive Conversational AI Streamlit application designed to empower traders and investors with advanced tools for stock market analysis with advanced NLP and ML models.

## INDUSTRY EXPERIENCES

**Industry Data Analyst Intern** | Orient Securities | Shanghai, China *April 2023 – Aug 2023*

**Industry Data Analyst Intern** | Guotai Junan Securities | Shanghai, China *Jan 2023 – April 2023*

**Data Science Intern** | Quant Investment | Shanghai, China *Sep 2022 – Dec 2022*

## SKILLS & INTERESTS

**Programming Languages**: Python, MATLAB, R, SQL

**Libraries/Frameworks/Tools**: Pytorch, TensorFlow, Scikit-learn, Numpy, Pandas, Scipy, Matplotlib, Seaborn, Cartopy, Beautifusoup, Request, Spark (ML Regression, ML Classification), LLMs, LLM-Agents

**Interests:** Music, Reading, Films, Nature