Anglin CAI

Email: cai@u.northwestern.edu | angl1n.github.io | Evanston, Illinois

Education

Northwestern University **09/2025 - 06/2027 (Expected)**

- M.S. in Computer Engineering (Thesis, supervised by Prof. Manling Li)
- GPA: 4.0/4.0

China University of Mining and Technology-Beijing **09/2021 - 06/2025**

- Degree: B.S. in Information & Computing Science
- GPA: 3.67/4.0 (Top 10%)
- Undergraduate thesis: Multimodal Time-Series Forecasting for Industrial Sensors

Research Interests

- Multimodality, Trustworthiness, Machine Learning

Publications

- Canyu Chen*, **Anglin Cai***, Joan Nwatu, Yale Li, Jessica Hullman, Rada Mihalcea, Kathleen McKeown, Manling Li. "Fairness Failure Modes of Multimodal LLMs." **NeurIPS 2026**. Also at TrustNLP@ACL 2026 (Oral). * Equal contribution
- **Anglin Cai**, Zhiqi Wang, Xiaozhi Yang, Yurun Lin, Dongwon Lee, Yuchen Yang. "A Plausible Explanation Is Enough to Mislead Multimodal Judges." In submission.

Awards

- **Best Paper Award** at ACL 2026 TrustNLP Workshop.
- **NeurIPS 2026 Travel Award.**

Research Experience

NYU Abu Dhabi | Abu Dhabi, UAE **06/2026 - Present**

Research Assistant | Project lead (Supervisor: Prof. Jiahao Yu)

- ***(In preparation) Visual-Evidence Faithfulness across VLM Answering, Judging, and Agentic Tasks***
- Evaluating and analyzing VLMs' faithfulness of visual evidence across deployment roles with multiple perturbations.

Pennsylvania State University | University Park, PA **02/2026 - Present**

Research Assistant | Project lead (Supervisor: Prof. Yuchen Yang & Prof. Dongwon Lee)

- ***A Plausible Explanation Is Enough to Mislead Multimodal Judges***
- Designed and built **JUPE**, a benchmark of **21,470 judgment instances across three visual domains**, using controlled interventions that hold images, questions, and answers fixed to isolate how explanations influence multimodal LLM judges.
- Identified **plausibility capture** across **14 multimodal judges**: ordinary explanations reduced rejection of incorrect answers by **12.7–13.8 percentage points** on average; applied signal detection theory to reveal a shift toward acceptance without a clear improvement in discriminability.

Northwestern University | Evanston, IL **06/2025 - 02/2026**

Research Assistant | NU-MLL-Lab (Supervisor: Prof. Manling Li)

- ***Fairness Failure Modes of Multimodal LLMs***
- Built **MultiBBQ**, a multimodal fairness benchmark comprising **2,460 QA pairs across four demographic categories**, using controlled image synthesis and visual grounding to reduce textual shortcuts in evaluation.
- Designed complementary **Fairness Score and Bias Score** metrics to distinguish appropriate abstention, evidence-based reasoning, and stereotypical responses across ambiguous and disambiguated contexts.
- Evaluated **28 multimodal LLMs**, uncovering divergent refusal behavior, fairness degradation under image noise and quantization, bias amplification over backbone LLMs, and limited effectiveness of common mitigation strategies.

Institute of Automation, Chinese Academy of Sciences | Beijing, China **02/2025 - 05/2025**

Research Assistant, National Key Laboratory of Pattern Recognition - Multimodal Physio-Psychological Sensing Group

Contactless ECG Reconstruction and Multimodal Emotion Sensing via mmWave Radar

- Formulated an mmWave ECG extraction workflow using signal decomposition to suppress respiration; implemented downsampling, denoising, and R-peak detection in Python to produce interpretable heartbeat markers.
- Optimized clinical data pipeline, reducing setup time by 83% (60m to 10m) and increasing patient throughput by 5x via independent deployment of mmWave and micro-motion sensors.

Skills

- **Programming:** Python, C/C++
- **ML / Data:** HuggingFace Transformers, PyTorch, scikit-learn, NumPy, pandas, Matplotlib
- **Tools:** Git, Linux, Jupyter, TensorBoard, Weights & Biases, VS Code / PyCharm, LaTeX / Overleaf