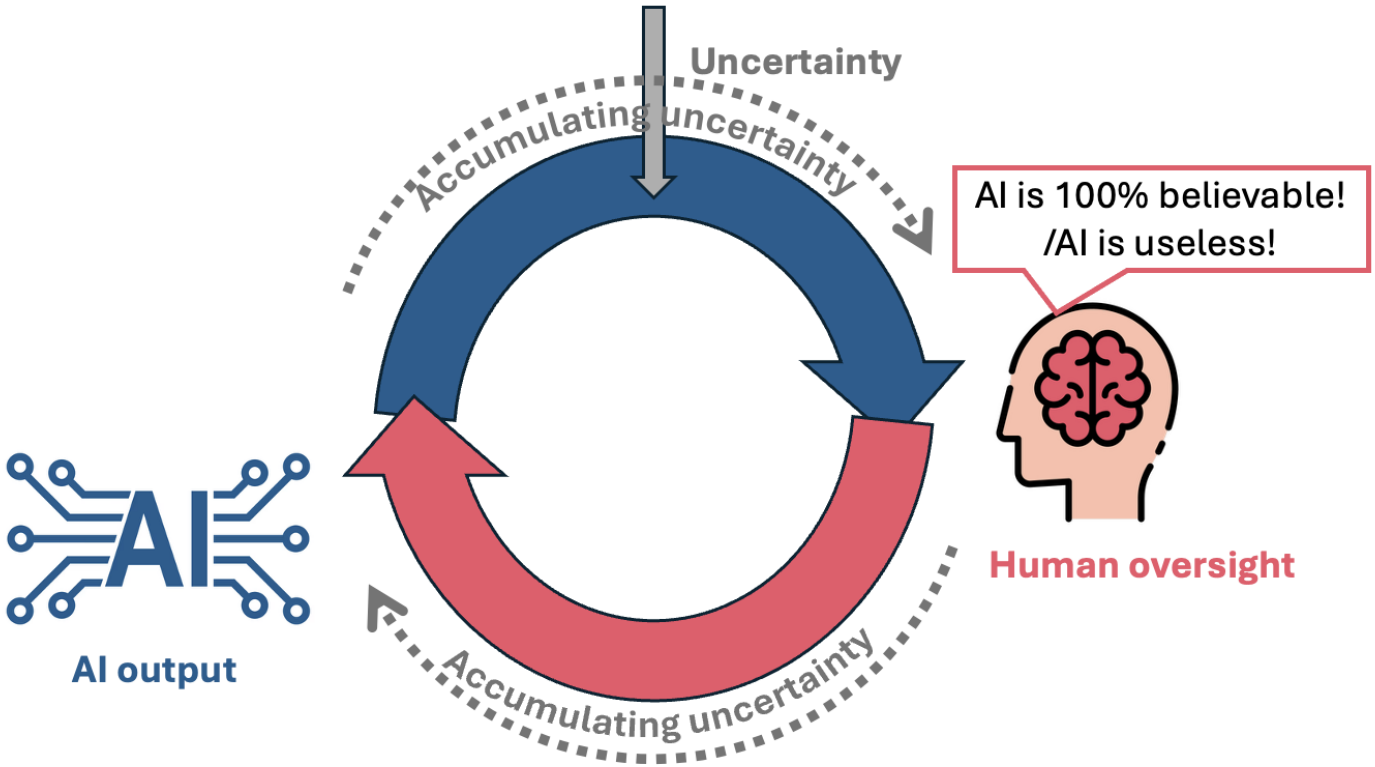
Uncertainty in Human-AI Collaborative System

Uncertainty disrupts Human–AI trust



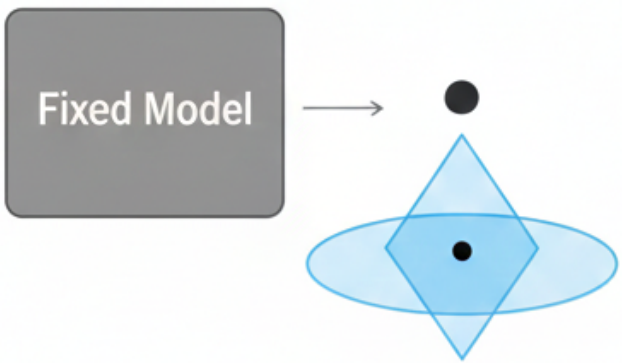
Human-AI interaction cycle

My Goal

Build trustworthy AI systems that quantify and control uncertainty

Research Summary

UNCERTANITY QUANTIFICATION (Post-hoc)



Idea: Conformal prediction for statisically valid prediction sets

UAI '23, NeurIPS '24, AAAI '26

UNCERTANITY CONTROL (Training)

Idea: Uncertanty-aware training & Federated coordination

AAAI '25, ICML '25, UAI '25, AAAI '26

Ongoing Projects

I am currently exploring:

- Uncertainty-aware preference selection for LLM fine-tuning (under review @ ICLR 2026)
- Semantic-aware text generation to preserve global consistency

I am actively thinking about:

- How reliability signals can reduce LLM hallucinations
- How agentic LLMs should estimate and use uncertainty

Happy to discuss or collaborate!

yuanjie.shi@wsu.edu