Keynote Speech

How Much Confidence Do We Have for GenAI-Based Reasoning – A Perspective from Reliability Inference

Speaker,

Steven Li, PhD, Professor

Western New England University

Speaker Biography:

Professor Steven Li is a full professor with the Department of Industrial Engineering and Engineering Management at Western New England University, Springfield, MA, USA. His research expertise includes AI/ML, data analytics, reliability/resilience engineering, and the trustworthiness and security of agentic AI. He earned his Ph.D. in Industrial Engineering from the University of Washington. Prof. Li currently serves as Editor-in-Chief of IEEE Transactions on Reliability, the flagship journal of the IEEE Reliability Society. Prof. Li served as the president of IEEE Reliability Society 2022-2024. He is the President-Elect of IEEE Systems Council 2026.

Abstract/Outline

This talk discusses the trust and confidence of using GenAI-based models in quantitative reliability reasoning. Using synthetic datasets generated from classic reliability models, we design multiple experimental factors, including sample sizes, prompts, and various AI models (Claude, DeepSeek, Gemini, ChatGPT, and xAI) to examine how various factors affect the accuracy in reliability estimations. The confidence levels and robustness of each AI model's performance are assessed through a factorial experimental design framework, and it is interesting to observe the estimates of unknown parameters under the classic Weibull reliability model do not converge over increasing sample size as what is commonly observed in the traditional statistics-based reliability inference. We also observe that ChatGPT and Deepseek perform the best in reliability estimation compared with other AI models. ChatGPT and Deepseek also show robustness in predicting reliability regardless of the types of prompts being used. The findings shed light on the reliability and confidence of using GenAI models for inferential reliability predictions and reasoning.

Desired Learning Outcomes from the talk

- Applications of GenAI in new product design for performance and reliability improvement;
- Implications of using GenAI in real-life engineering and product design process;
- Trust and confidence of GenAI reasoning in general

Target Audience

GenAI researchers, Industry professionals focusing on systems engineering, reliability practitioners, job seekers in GenAI and prompt engineering

Topics and keywords: *AI, GenAI, Trust of AI reasoning, AI confidence assessment, High-impact decision making*