Hillary Nana Yaa Owusu

LinkedIn · (240) 946-1078 · www.hillaryowusu.com · hnyowusu@cs.umd.edu

Education

University of Maryland — **College Park, MD**
Ph.D. in Computer Science | Advisor: Naomi Feldman — *Expected 2028*
M.S. in Computer Science, 2026

Kwame Nkrumah University of Science and Technology (KNUST) — **Kumasi, Ghana**
B.S. in Telecommunications Engineering — *2021*

Research Interests: LLM Reliability, Mechanistic Interpretability, Cognitive Bias in AI, Reasoning Robustness, Trustworthy AI

Research Experience

CLIP Lab, University of Maryland — **College Park, MD**
Ph.D. Researcher — *2023–Present*

Conduct interdisciplinary research spanning NLP, cognitive science, and computational social science to understand reliability failures in large language models. Study how misleading context, model confidence, and post-training shape cognitive biases and decision-making. Use behavioral and mechanistic analyses to identify how these failures emerge internally and develop evaluations that surface reliability risks before deployment in decision-support settings. Implement and prototype models using Python, R, C, and Java; leverage PyTorch, TensorFlow, NumPy, pandas, and scikit-learn for model development and analysis.

Selected Papers

[Google Scholar]

1. Anchoring Depends on Confidence and Post-Training in Language Models. *ACL 2026.* [Paper]
2. Localizing Anchoring Pathways in Language Models. *Preprint; under review.* [Preprint]
3. Bias-Aware AI Chatbot for Engineering Advising. *arXiv, 2025*; mentored undergraduate research project. [Preprint]

Selected Project

CouncilAgent / AnchorBreak — **MLSS NYC 2026 Hackathon**
Agentic Framework for Reducing Anchoring Bias in Language Models — *2026*

Co-developed a multi-agent deliberation framework that coordinates independent model proposals, structured critique, and evidence-based revision to reduce anchoring effects in language-model responses. Contributed to the research formulation, system design, benchmark integration, and empirical evaluation; the framework improved performance over the baseline by 62.7% and achieved 89% of human-best benchmark performance. [Project]

Teaching & Mentorship

University of Maryland — **College Park, MD**
Program Coordinator, Technical & Leadership Preparedness Program — *2024–Present*

Design and deliver a multi-module AI and machine learning workshop series for undergraduates with varied technical backgrounds. Translate concepts in probability, linear algebra, machine learning, data analysis, and large language models into accessible lessons, hands-on exercises, and applied projects. Mentor students conducting AI research, including a project on bias-aware systems for academic advising.

Professional Service & Recognition

Peer Review: NeurIPS 2026; ACL 2026; NeurIPS Reliable ML Workshop 2025; ICML Workshops on AI4Good and Technical AI Governance Research

Selected Participant: Machine Learning Summer School (MLSS NYC), 2026 — selected among approximately 200 participants from more than 1,200 applicants ($\sim$16.7% selection rate)

Award: ACM-W Scholarship