Negar Kamali

negar.kamali@u.northwestern.edu - negarkamali.github.io

RESEARCH INTERESTS

• Human-AI Collaboration • AI-Assisted Decision Making • LLMs as Coaches • AI Alignment

EDUCATION

Ph.D. in Computer Science *2022 – present (expected 12/27)* Northwestern University
Evanston, IL, USA; dissertation proposal expected 12/26

Ph.D. in Computational Mechanics *2013-2018* University of Illinois at Chicago
Chicago, IL, USA

M. Sc. in Computational Mechanics *2010-2013* University of Tehran
Tehran, Iran

B.Sc. in Civil Engineering *2006-2010* Tabriz University
Tabriz, Iran

ACADEMIC EXPERIENCE

Northwestern University | *Ph.D. Researcher, Department of Computer Science* Sep 2022 – Present

Member of the Human-AI Collaboration Lab (Kellogg School of Management) and the Epistemic Decisions Lab (CS)

Project: AI-Assisted Decision Making: How Expert and LLM Guidance Shapes Human Judgment

- Designed and preregistered a randomized controlled experiment (three conditions, nine image and video sets, 2,699 participants, 65,000+ pre/post judgments) testing whether just-in-time advice from human experts and from MLLMs improves people's ability to tell real from AI-generated media.
- Estimated causal effects with regression models that account for repeated judgments per participant, measured how the effect of advice changed over the course of a session, and separated gains in discrimination from shifts in response bias (signal detection analysis).
- Tested whether effects generalize beyond specific images using bootstrap confidence intervals and per-stimulus heterogeneity analysis, and built measures of calibration from confidence ratings and participants' predictions of how others would judge the same media.
- Built the LLM advice pipeline (OpenAI, Anthropic, Gemini APIs) and validated it with blind LLM-as-judge checks and prompt ablations, confirming that advice did not reveal the answer and that results did not depend on prompt wording.
- Followed up with a mechanistic interpretability study. Linear probes on open-weight vision-language models detect AI-generated images far better than the models' own answers (AUC 0.92 vs. 62% accuracy), and activation patching and region ablation show the models hold the evidence but do not use it when writing advice (Qwen3-VL, Gemma).
- Supervise an M.S. thesis extending this work: fine-tuning an open vision-language model on expert tips for CCTV footage

Project: Decision-Calibrated Explanations, Reinforcement Learning for Human-Centered LLM Reasoning

- Contributed the human judgment data behind HD-RL, a reinforcement learning method that fine-tunes an open-weight LLM to write explanations that improve human decisions, using a reward model trained on real human choices. In a four-arm human evaluation, its explanations raised accuracy to 65.6%, against 51.9–55.3% for RLVR and RLAIF baselines.

Project: LLM Coach for Cash-Transfer Recipients (with GiveDirectly)

- Led Northwestern's evaluation of a WhatsApp LLM coach that GiveDirectly deployed to cash-transfer recipients in Rwanda: helped design a randomized comparison of two coaching prompts, built an arm-blind LLM-as-judge rubric for topics, coaching style, and plan formation, and analyzed 1,500+ real voice and text conversation turns.
- Designed persona-based simulations and red-team probes grounded in real user data, and presented findings to GiveDirectly's product and leadership teams, who adopted a single-agent design and revised prompts before launch.

Project: Uncertainty Communication in AI-Assisted Decision Tasks

- Conducted a thematic analysis on how AI-generated confidence sets affect human trust, reliance, and labeling accuracy across in-distribution and out-of-distribution tasks.

Project: Participatory Design of AI Decision-Support Tools for Healthcare

- Investigated how end-users interact with and interpret next-day machine learning stress predictions, explanations, and recommendations in a healthcare context.
- Directed participatory design sessions with diverse stakeholder groups to co-design patient-facing AI decision-support tools.
- Devised a prototype Decision Support Tool (DST) showcasing predictions, explanations, bias, uncertainty, risk, and actionable recommendations.

Univ. of Illinois at Chicago | *Research Assistant* 2013-2018

- Developed an Enriched Reproducing Kernel Particle Method (RKPM) to solve for linear and nonlinear wave propagation PDEs.
- Conducted numerical simulation of wave propagation in multiscale material.
- Wrote several user subroutines for the commercial software Abaqus.

PUBLICATIONS

Conference Proceedings

- "Generative AI Literacy Training Improves Intelligence Analysts' Discrimination of Real and AI-Generated Images", **N. Kamali**, C. R. Gerstner, J. Hullman, M. Groh, *International AAAI Conference on Web and Social Media,* **ICWSM 2027**
- "Characterizing Photorealism and Artifacts in Diffusion Model-Generated Images", **N. Kamali**, K. Nakamura, A. Kumar, A. Chatzimparmpas, J. Hullman, M. Groh, *ACM CHI Conference on Human Factors in Computing Systems,* **CHI 2025**
- "Evaluating the Utility of Conformal Prediction Sets for AI-Advised Image Labeling", D. Zhang, A. Chatzimparmpas, **N. Kamali**, J. Hullman, *ACM CHI Conference on Human Factors in Computing Systems,* **CHI 2024**
  - **Best Paper Honorable Mention Award**

Under Review

- "Calibrating LLM Explanations to Human Decisions with Reinforcement Learning", D. Shu, Z. Guo, **N. Kamali**, M. Groh, J. Hullman, *International Conference on Learning Representations, under review, arXiv,* **ICLR 2027**

Preprints

- "How to Distinguish AI-Generated Images from Authentic Photographs", **N. Kamali**, K. Nakamura, A. Chatzimparmpas, J. Hullman, M. Groh, *arXiv,* 2024

Journal Publications

- "Patients' Expectations of ML-driven JITAI Support for Maternal Stress Management", M. Ulloa, **N. Kamali**, G. Fernandes, E. Soyemi, M. Beltzer, N. Alshurafa, M. Jacobs, *ACM Transactions on Computing for Healthcare,* **HEALTH 2026**
- "Harmonic-enriched reproducing kernel approximation for highly oscillatory differential equations", A. Mahdavi, Sh. W. Chi, **N. Kamali**, *ASCE Journal of Engineering Mechanics,* 2020
- "Influence of Mesoscale and Macroscale Heterogeneities in Metals on Higher Harmonics Under Plastic Deformation", **N. Kamali**, N. Tehrani, A. Mostavi, Sh. W. Chi, D. Ozevin, J.E. Indacochea, *Journal of Nondestructive Evaluation,* 2019
- "Numerical study on how heterogeneity affects ultrasound higher harmonics generation", **N. Kamali**, A. Mahdavi, Sh. W. Chi, *Nondestructive Testing and Evaluation,* 2019
- "Wavelet Based Harmonics Decomposition of Ultrasonic Signal in Assessment of Plastic Strain in Aluminium", A. Mostavi, **N. Kamali**, N. Tehrani, Sh. W. Chi, D. Ozevin, J.E. Indacochea, *Measurement,* 2017

PRESENTATIONS AND TALKS

Invited Talks & Panels

- **AI and Disinformation Summit** (Panelist), *Center for Informed Democracy and Social Cybersecurity, Carnegie Mellon University, Pittsburgh, PA, January 2025.*
- **Laboratory for Analytical Science 2024 Research Symposium** (Invited Speaker), *North Carolina State University, Raleigh, NC, December 2024.*

Posters and Presentations

- "Human-AI Collaboration for Detecting AI-Generated Media: Just-in-Time Advice from Experts and LLMs", **N. Kamali**, A. Kumar, A. Chatzimparmpas, C. Gerstner, J. Hullman, M. Groh, *Poster, ACM CI/HCOMP, Alexandria, VA, USA; NSF AI-SDM Workshop on Human-AI Complementarity, Carnegie Mellon University, Pittsburgh, PA, USA* 2026
- "Generative AI Literacy Training Enhances Detection Accuracy Among Government Analysts", **N. Kamali**, M. Groh, *CODE@MIT, Cambridge, MA, USA* 2025
- "Evaluating Human Perception of AI-Generated Images", **N. Kamali**, A. Chatzimparmpas, J. Hullman, M. Groh, *IC2S2, Philadelphia, PA, USA* 2024
- "Co-Designing Patient-Facing Machine Learning for Prenatal Stress Reduction", M. Ulloa, **N. Kamali**, G. Fernandes, E. Soyemi, M. Beltzer, B. Kaveladze, N. Alshurafa, M. Jacobs, *ISRII, Ireland; CSCW Workshop, Minneapolis, MN, USA* 2024
- "Evaluating the Utility of Conformal Prediction Sets for AI-Advised Image Labeling", D. Zhang, A. Chatzimparmpas, **N. Kamali**, J. Hullman, *Human+AI Symposium, University of Chicago, Chicago, IL, USA* 2023

HONORS & AWARDS

BlueDot Impact | *Rapid Grant for AI Safety Research* 2026
Northwestern University | *Cognitive Science Advanced Research Fellowship* 2024
Northwestern University | *ACM CHI Best Paper Honorable Mention* 2024
Northwestern University | *Todd M. and Ruth Warren and the Chookaszian Family Fellowship* 2022 & 2023
Univ. of Illinois, Chicago | *Chancellor's Student Service and Leadership Award* 2017
Univ. of Illinois, Chicago | *Excellence in Undergraduate Mentoring Scholarship* 2017
Univ. of Illinois, Chicago | *Chicago Consular Corps of Engineers Scholarship* 2017
Univ. of Illinois, Chicago | *UIC Presenter Award* 2016
Univ. of Illinois, Chicago | *Graduate Student Council UIC Award* 2016

TEACHING EXPERIENCE

Northwestern University

- COMP_SCI 367/467: Wireless and Mobile Health (mHealth), Prof. Alshurafa, Winter 2026

University of Illinois at Chicago

- Structural Analysis, Prof. Chi, 2013–2015

INDUSTRY EXPERIENCE

Software Developer | **US API Manager** | *SkyCiv* Jan 2021 - Sept 2022

- Developed cloud-based software for structural engineers, coordinating cross-functional product development across distributed teams.

Automation Expert | *SOM and Arup* 2018 - 2021

SKILLS

- **Research Methods** | Randomized controlled trials, field experiments, within-subject and quasi-experimental designs, A/B testing at scale, survey design, preregistration, thematic and qualitative analysis
- **Statistics** | Causal inference, mixed-effects and OLS regression with clustered errors, logistic regression, bootstrap inference, signal detection theory, power analysis

- **AI Evaluation** | LLM-as-judge pipelines, red-teaming, persona-based simulation, prompt ablations, linear probing and activation patching of vision-language models
- **Programming** | Python (PyTorch, Hugging Face Transformers, scikit-learn, pandas), R, SQL, JavaScript, Git
- **Platforms** | Flask web experiments, SQLite, Prolific, OpenAI, Anthropic, and Gemini APIs, SLURM GPU clusters
- **Generative Media Tools** | Midjourney, Adobe Firefly, Stable Diffusion, Google Veo and Imagen, OpenAI Sora, Kling
- **Relevant Coursework** | Bayesian Statistics, Decision Theory, Prediction for Decision-Making, Explanation and Reproducibility, Differential Privacy, Online Markets

MEDIA COVERAGE

- *New Scientist:* How to Avoid Being Fooled by AI-Generated Misinformation
- *Kellogg Insight:* Can You Tell if These Photos Are AI-Generated?
- *Mashable:* How to identify AI-generated videos
- *Kellogg Insight:* When Put to the Test, Are We Any Good at Spotting AI Fakes?
- *Technology Magazine:* Does Google's Veo 3 Do Enough to Distinguish AI and Reality?
- *Asahi Globe:* What to look for to spot fake images? (In Japanese)

PROFESSIONAL AFFILIATIONS

- Graduate Society of Women Engineers, Professional Development Officer, *Northwestern University*, 2023-2024
- Graduate Society of Women Engineers, Founder and President, *Univ. of Illinois at Chicago*, 2016
- Reviewer for ACM CHI, ACM DIS, and ACM Creativity & Cognition, 2024–2026