

Highlights of Accomplishments

- 8+ years of AI & NLP research: 20+ publications, including EMNLP, EACL, and ACL main conference papers.
- SOTA in Confidence Estimation: LLMs, LRMs & VLMs; EMNLP'25 Outstanding Paper nominee (top 0.4%).
- 4+ years in Healthcare AI: 7 presentations at premier oncology conferences (5 ASTRO, 2 AAPM).
- 4+ years in Finance AI: Mapped science-to-industry funding; SOTA investment prediction on 70M+ graph nodes.
- Industry Collaborations: Led projects with JPMorgan AI Research, Henry Ford Health (HFH), & Cedars-Sinai.

Education & Research Experience

Ph.D. Computer Science - Michigan State University

Aug 2022 - Present (Expected Fall 2027)

AI Research Assistant @ Human Augmentation and Artificial Intelligence Laboratory (Supervisor: [Mohammad M. Ghassemi](#))

Thesis: Signals of Trust: Confidence Estimation in Language, Reasoning, and Vision-Language Models - GPA: 3.95/4 East Lansing, USA

B.Sc. Computer Engineering - University of Guilan

Sep 2017 - Aug 2021

NLP Research Assistant @ Guilan NLP Group Laboratory (Supervisor: [Seyed Abolghasem Mirroshandel](#))

Thesis: Prose2Poem: The Blessing of Transformers in Translating Prose to Persian Poetry - GPA: 3.86/4

Rasht, Iran

Technical Skills

Technical Expertise	Machine Learning, Deep Learning, NLP, Vision-Language Models, Uncertainty Quantification
Programming Languages	Python, C/C++, Java, MATLAB
Data Science Tools	NumPy, Pandas, Scikit-learn, SQL, Jupyter
Machine Learning Tools	PyTorch, Hugging Face, vLLM, DeepSpeed, Unsloth, PyTorch Geometric, TensorFlow
ML Techniques	Probing & Hidden-State Analysis, Calibration, Adversarial Attacks, Fine-tuning (LoRA), RAG, GNNs
Language Models	Llama, Qwen, Gemma, Mistral, Phi, DeepSeek-R1, EXAONE, GPT, Qwen-VL, LLaVA, InternVL
Cloud Computing Tools	Amazon AWS, Microsoft Azure

Selected Publications ([Google Scholar](#))

 **Calibrating LLM Confidence by Probing Perturbed Representation Stability** - EMNLP'25 Main Conference (Top 0.4% of 8,174 submissions, Nominated for Outstanding Paper Award) - [Code](#) & [Data](#)

[Khanmohammadi, R.](#), [Miahi, E.](#), [Mardikoraem, M.](#), [Kaur, S.](#), [Brugere, I.](#), [Smiley, C. H.](#), [Thind, K.](#), & [Ghassemi, M. M.](#) (In collaboration with JPMorgan & HFH) I introduced CCPS, a confidence estimation method that perturbs an LLM's final hidden states and predicts answer correctness from how stable its representations are. Across Llama, Qwen, and Mistral models (8B to 32B) on MMLU and MMLU-Pro, it reduces ECE by 55% and Brier score by 21% and improves accuracy, AUPRC, and AUROC over the strongest prior method, all without modifying model weights.

 **How Reliable are Confidence Estimators for Large Reasoning Models? A Systematic Benchmark on High-Stakes Domains** - EACL'26 Main Conference - [Code](#) & [Data](#)

[Khanmohammadi, R.](#), [Miahi, E.](#), [Kaur, S.](#), [Smiley, C. H.](#), [Brugere, I.](#), [Thind, K.](#), & [Ghassemi, M. M.](#) (In collaboration with JPMorgan & HFH) I introduced RMCB, a benchmark of 347K+ reasoning traces from six LRMs across clinical, financial, legal, and mathematical reasoning, and developed representation-based confidence estimators that achieved SOTA results against 10+ methods. The study uncovered a fundamental trade-off between discrimination and calibration, with no single method dominating both, and a performance ceiling for methods relying only on chunk-level hidden states.

 **Grounded or Guessing? LLM Confidence Estimation via Blind-Image Contrastive Ranking** - ACL'26 Main Conference - [Code](#) & [Data](#)

[Khanmohammadi, R.](#), [Miahi, E.](#), [Kaur, S.](#), [Smiley, C. H.](#), [Brugere, I.](#), [Thind, K.](#), & [Ghassemi, M. M.](#) (In collaboration with JPMorgan & HFH) I introduced BICR, a confidence estimation method for LLMs that detects answers driven by language priors rather than the image. It contrasts hidden states from the real image against a blacked-out one and trains a lightweight probe to reward visual grounding. Across five LLMs and seven baselines spanning VQA, hallucination detection, medical imaging, and financial documents, it achieved SOTA calibration and discrimination simultaneously at 4-18x fewer parameters and zero added inference cost.

 **Vision-Language Model Confidence Is Not a Property of the Answer** - Under review (2026) - [Code](#)

[Khanmohammadi, R.](#), [Miahi, E.](#), [Brugere, I.](#), [Smiley, C. H.](#), [Kaur, S.](#), [Thind, K.](#), & [Ghassemi, M. M.](#) (In collaboration with JPMorgan & HFH) I exposed a critical vulnerability in confidence-gated VLM systems: an attacker can manipulate a model's confidence without changing its answer, making the system reject correct answers and accept wrong ones. I developed answer-preserving attacks that break existing confidence estimators across four VLMs and established that current defenses fail to restore reliability, challenging the assumption that a model's confidence reflects the answer it gives.

Applied Publications

 **Calibrated Triage, Not Autonomy: Confidence Estimation for Medical Vision-Language Models** - Under review (2026) - [Code](#)
[Khanmohammadi, R.](#), et al. (In collaboration with HFH)

 **Confidence Estimation for Financial Vision-Language Models in Chart and Document Understanding** - Under review (2026) - [Code](#)
[Khanmohammadi, R.](#), et al. (In collaboration with JPMorgan)

-  **Teaching Language Models Safe Clinical Decision Making: Risk-Informed Training to Reduce Catastrophic Errors in Oncology** - ASTRO'26 (oral, top 11%) - IJROBP'26 - IF=6.5 - Khanmohammadi, R., et al. (In collaboration with HFH)
-  **Risk-Controlled Selective Automation of Oncology Chart Review via Representational Stability Calibration** - ASTRO'26 (spotlight) - IJROBP'26 - IF=6.5 - Khanmohammadi, R., et al. (In collaboration with HFH)
-  **Clinical Validation of a Privacy-Preserving LLM for End-to-End CTCAE Toxicity Assessment in Prostate Cancer Radiotherapy** - IJROBP'26 - IF=6.5 - Khanmohammadi, R., et al. (In collaboration with HFH)
-  **Efficient CTCAE Grading for Post-Radiotherapy Toxicities Using Large Language Models: A Privacy-Preserving Approach Using Instruction Fine-Tuning** - IJROBP'25 - IF=6.5 - Khanmohammadi, R., et al. (In collaboration with HFH & Cedars-Sinai)
-  **Integrating natural language processing into radiation oncology: a practical guide to transformer architecture and large language models** - BJR|AI (2025) - Khanmohammadi, R., et al. (In collaboration with HFH & Cedars-Sinai)
-  **Hybrid student-teacher large language model refinement for cancer toxicity symptom extraction** - BHI'25 Khanmohammadi, R., et al. (In collaboration with HFH & Cedars-Sinai & NYU-AD)
-  **Investigating the Temporal Association of Biomedical Research on Small Business Funding: A Bibliometric and Data Analytic Approach** - IEEE TCSS (2024) - IF=4.9 - Khanmohammadi, R., et al. (In collaboration with JPMorgan & NYU-AD)
-  **Iterative Prompt Refinement for Radiation Oncology Symptom Extraction Using Teacher-Student Large Language Models** - ICCR'24 - Khanmohammadi, R., et al. (In collaboration with HFH & Cedars-Sinai)
-  **A Novel Localized Student-Teacher LLM for Enhanced Toxicity Extraction in Radiation Oncology** - IJROBP'24 - IF=6.5 - Khanmohammadi, R., et al. (In collaboration with HFH & Cedars-Sinai)
-  **MambaNet: A Hybrid Neural Network for Predicting the NBA Playoffs** - SN Computer Science (2024) - IF=3.1 Khanmohammadi, R., et al. (In collaboration with Hudl Instat & NYU-AD)
-  **Prose2Poem: The Blessing of Transformers in Translating Prose to Persian Poetry** - ACM Transactions on Asian and Low-Resource Language Information Processing (2023) - IF=2.0 - Khanmohammadi, R., et al.

Honors & Distinctions

Microsoft Applied Sciences Internship Offer

Selected for an Applied Sciences internship on the Excel Agent Team.

May 2026

Redmond, USA

EMNLP 2025 Outstanding Paper Award Nomination

Paper “Calibrating LLM Confidence by Probing Perturbed Representation Stability” nominated for the Outstanding Paper Award, placing in the top 0.4% of 8,174 submissions.

Sep 2025

Suzhou, China

NSF-EMBS-Google Young Professional NextGen Scholar

Selected as one of a limited group of early-career researchers recognized for outstanding accomplishments in biomedical AI and health informatics, with sponsored support and recognition by NSF, EMBS, and Google at IEEE BHI'25.

Sep 2025

Atlanta, USA

Sigma Xi Full Member

Sigma Xi is a Scientific Research Honor Society; Full membership is by invitation only. Learn more [here](#).

Jul 2025

Marathoner - Long Beach Marathon

Oct 2024

ACM ICAIF'23 Workshop Organizing Committee Member

Served as an organizing committee member for the “Large Language Models for NLP in Finance: Applications and Challenges” workshop in ICAIF'23. Presentation material available [here](#).

Nov 2023

Brooklyn, USA

3rd Henry Ford + MSU Cancer Research Symposium

Recognized with the Top Cancer Control Abstract Award and invited to deliver a [talk](#).

Jan 2023

East Lansing, USA

Peer Review & Program Committee Service

Program committee for AAAI'27, IEEE BHI'26, ICML'26 AI4Law Workshop, and CIKM'26 LASS Workshop; 26+ reviews across 8 journals, including Scientific Reports, Communications AI & Computing, Neural Computing and Applications, and Discover AI.

Oct 2022 - Present

Department of Computer Engineering - University of Guilan Scholarship

Awarded a Full Scholarship for graduate studies as an Exceptional B.Sc. Student Talent.

Aug 2020

Rasht, Iran

International University Sports Federation (FISU) Most Active Ambassador

Recognized as one of the most active Student Ambassadors, ranking in the top 10% of 130 ambassadors representing 80 countries, for leadership and contributions to international student sports.

Jan 2020

Languages

English: Fluent

German: Intermediate

Persian: Native