

Mohammad Safarzadeh, Ph.D.

mtsafarzadeh@gmail.com | 443.467.5303
LinkedIn | mtsafarzadeh.github.io

Principal Research Scientist at Oracle focused on evaluation, reliability, and adaptation of large language model systems for high-stakes enterprise use cases. Experience spans benchmark design, uncertainty quantification, retrieval-augmented generation, and conflict detection in financial-domain workflows, with additional work in fraud detection, domain adaptation, and healthcare LLMs.

Technical Focus: LLM evaluation, financial AI, RAG reliability, uncertainty quantification, benchmark design, risk-sensitive model analysis, domain adaptation

Experience

Oracle: Principal Research Scientist Nov 2023–Present

- Led evaluation and benchmarking efforts for enterprise LLM systems supporting business-critical Oracle workloads, including financial and database-centered use cases.
- Designed standardized evaluation protocols and diagnostics for robustness, contamination, and failure analysis across complex LLM pipelines.
- Developed methods for conflict detection and resolution in retrieval-augmented generation pipelines for high-stakes financial-domain workflows.
- Designed and published the SQL2NL reverse-generation framework (EMNLP 2025), enabling realistic benchmark construction from production SQL logs and exposing systematic model failure modes.
- Built continual pretraining pipelines for domain-adapted medical LLMs, improving downstream question-answering accuracy by 7%.
- Developed privacy-aware clinical information extraction systems, currently piloted within Oracle Health.

Perceive: Senior Machine Learning Engineer Apr 2022–Nov 2023

- Developed and deployed model compression pipelines (pruning, quantization, distillation, low-rank factorization), significantly reducing model size while preserving accuracy.
- Built gradient- and activation-level monitoring tools to diagnose training dynamics and stability issues in large neural models.

FICO: Scientist II Dec 2021–Apr 2022

- Developed lightweight neural network models for real-time credit card fraud detection under strict latency constraints.

Publications

- M. Safarzadeh, A. Oroojlooy, D. Roth. Evaluating NL2SQL via SQL2NL. **Findings of EMNLP 2025.**
- M. Safarzadeh, H. Laxmichand Patel, A. Oroojlooy, G. Horwood, D. Roth. SPENCE: A Syntactic Probe for Detecting Contamination in NL2SQL Benchmarks. **ACL 2026 Main Conference.**
- X. Zou, R. Sridhar, M. Safarzadeh, D. Roth. When Vision-Language Models Judge Without Seeing: Exposing Informativeness Bias. **ACL 2026 Main Conference.**

Patents (Oracle)

- Intelligent conflict detection and resolution in RAG pipelines. USPTO application filed June 2025.
- Uncertainty quantification for large language models via iterative prompting. Provisional and approved filings, 2025.
- Synthetic data generation via SQL-to-NL for training NL-to-SQL models. USPTO application filed August 2025.
- Syntactic contamination detection for text-to-SQL benchmarks and anchor-guided VLM judging. Provisional filings, December 2025.

Education

Johns Hopkins University	Astrophysics	Ph.D., 2016
--------------------------	--------------	-------------

Awards

Harvard Dean's Competitive Fund for Promising Scholarship	2020
---	------

Research Background

Harvard University, Center for Astrophysics	2018–2021
---	-----------

Postdoctoral research in explainable AI for gravitational-wave detection.

Arizona State University	2016–2018
--------------------------	-----------

Research Profile

35 first-author peer-reviewed publications, 1320+ citations, h-index 26 **Google Scholar**.