

SARA MALVAR

malvar.sara@gmail.com • +55 61 981156756

Applied Scientist & AI Engineer (Ph.D.) with 10+ years of experience translating research into production-grade AI systems. Senior AI Applied Scientist at Microsoft, specializing in LLM evaluation, benchmarking, and GenAI systems. Leads enterprise-scale AI platforms (e.g., Microsoft Frontier/Copilot Tuning), partnering across product and engineering to deliver measurable business impact. Inventor on multiple AI patents and mentor to 300+ data scientists.

Experience

Microsoft | M365

Sr. AI Applied Scientist

Sep 2024 – Present

- Led LLM evaluation strategy for Copilot Tuning/Frontier Copilot, the turnkey fine-tuning product in Copilot Studio, defining quality bars across skills/recipes and unblocking GA readiness for Fortune 500 customers including Marriott, McKinsey, Koch, EY, Pearson, and Bayer
- Architected a DAG-based, agent-agnostic LLM-as-a-judge framework adopted as the single evaluation backbone across all agent recipes; delivered 45+ production changes in 6 months spanning rubric generation, hallucination & PII-leakage detection, and partial-trajectory scoring; reduced eval cost and turnaround ~2x via ~40% response-cache reuse.
- Built grader-as-a-service APIs powering reinforcement and supervised fine-tuning pipelines, adopted by all recipe engineering teams to self-serve evaluation.
- Designed a self-calibrating evaluation system for LLM judges, introducing a multi-objective calibration framework with meta-metrics and optimizing grading behavior via Pareto-based prompt evolution, improving reproducibility and alignment across models and reducing manual calibration.
- Partnered cross-functionally with PM, applied science, and recipe engineering to translate enterprise customer feedback (Finance, HR, Legal, Marketing, Accounting workflows) into evaluation requirements; mentored 1 direct report and onboarded an additional applied scientist onto the framework.

Microsoft Research

Research Software Development Engineer

Aug 2021 – Sep 2024

- Co-built and shipped FarmVibes.AI / TerraVibes, Microsoft Research's multi-modal geospatial AI platform, owned the methane detection workstream end-to-end (Sentinel-2 satellite imagery, deep-learning cloud/shadow models, source attribution), delivered into the platform used by enterprise customers including Schlumberger, Land O'Lakes, Bayer, ITC, and Digital Green, and won 1st place ("Hack for Space") and 2nd place ("Hack for Sustainability Executive Challenge") at the Microsoft Global Hackathon.
- Led the LLM fine-tuning and RAG pipeline effort that became the team's flagship deliverable: shipped fine-tuning, Q&A, and RAG pipelines into the Azure Machine Learning private registry; work was showcased to Microsoft's CTO and the President of Microsoft Research and spotlighted by Microsoft's CEO in the Build 2024 opening keynote.
- Drove two flagship enterprise R&D engagements (Dow and Unilever) as technical lead: shipped 3 ML solutions (causal ML, transformer-based SMILES models, tuned baselines) that outperformed Dow's in-house emulsion-stability models, securing a renewed multi-year engagement.
- Generated significant AI research output: named inventor on 10+ patents and author on 9 peer-reviewed papers, spanning LLM fine-tuning, retrieval-augmented generation, causal machine learning, and deep learning for remote sensing

Research Centre for Gas Innovation (RCGI)

Jul 2019 – Aug 2021

Postdoctoral Researcher

- Applied ML across 6+ interdisciplinary projects at the intersection of energy, environment, and materials, including deep learning for materials discovery, NLP for public-sentiment analysis on energy policy (Shell partnership), and ML-driven seismic imaging (STMI program with Shell). Published peer-reviewed work and translated research prototypes into tools used by industry partners.
- Bridged physics-based modeling with machine learning to convert sensor and simulation data into predictive models for active-matter rheology and bioprocess optimization. Mentored junior researchers and co-supervised graduate students.

Udacity, Data Science Dojo, Awari, Alura

2018 – Aug 2024, concurrent

Data Science & ML Instructor

- Taught and mentored 300+ students and professionals in machine learning, deep learning, and applied LLMs across 4 platforms, including 1:1 mentorship across 9 Udacity nanodegrees (Data Scientist, ML Engineer, Deep Learning, AI for Healthcare), live webinars at Data Science Dojo, and 3 Awari cohorts. Top-rated mentor on MentorCruise for coaching on LLMs and applied AI.

Education

Ph.D. in Applied Sciences

University of Pennsylvania/University of Sao Paulo

2015 - 2019

Ph.D. visiting researcher

University of Tokyo

SELA Scholarship recipient (Brazil–Japan government academic exchange)

2018

M.Sc. in Applied Sciences

University of Brasilia

EMBRAER Best Master's Dissertation Award

2014 - 2015

B.Sc. in Electrical Engineering

University of Brasilia

Graduated with honors, EMBRAER Best Undergraduate Project Award

2009 - 2013

Leadership & Community

- Technical lead for evaluation and grading infrastructure across Copilot Tuning, coordinating rubric, grader, and integration work across applied science, PM, and engineering teams.
- Mentored 10+ interns and early-career scientists, with mentees producing production code, patents, and peer-reviewed publications; primary onboarding point for new applied scientists on the team.
- Led hiring and interview loops for applied scientist roles, providing signal on technical depth, judgment, and research-to-production readiness.
- Frequent internal speaker and reviewer on LLM evaluation, fine-tuning, and agent quality, influencing best practices across teams.

Skills & Tools

AI / ML: Python, PyTorch, TensorFlow, Keras, scikit-learn, XGBoost, deep learning, classical ML, causal ML, statistical modeling.

GenAI / LLMs: LangChain, LangGraph, LlamaIndex, DSPy, Hugging Face Transformers, PEFT, vLLM, RAG, fine-tuning, LLM evaluation, LLM-as-a-judge, agent orchestration, agent evaluation, tool use, structured outputs, prompt injection defense.

MLOps / Production AI: Azure ML, MLflow, Docker, CI/CD, model deployment, monitoring, inference pipelines, model registry, experiment tracking.

Data / Analytics: SQL, pandas, experiment design, benchmarking, metric design, error analysis, data quality, data validation, A/B testing, multivariate testing, hypothesis testing, causal inference.