

Ahmed Taha

AI Safety & Evaluation Researcher | ML Engineer | Patent Examiner

(408) 439-9141 | at4058@Columbia.edu | AhmedTaha.io | linkedin.com/in/ahmedtaha-/ | github.com/ahmedtaha100

Education

Columbia University, Fu Foundation School of Engineering

M.S. Computer Science (Thesis Track) | GPA: 3.6

August 2025 – May 2027

Coursework: Machine Learning, High Performance Machine Learning, Neural Networks & Deep Learning, Natural Language Processing

Johns Hopkins, Whiting School of Engineering

M.S. Cybersecurity

August 2024 - May 2026

California State University, Sacramento

B.S. Computer Science

August 2024

Professional Experience

AI Research Student | Columbia University, Irving Medical Center

August 2025 - Present

- <https://scholar.google.com/citations?user=FQ1XbKcAAAAJ&hl=en>
- Developing a RL policy to enhance AI-assisted organ segmentation for radiation imaging under Dr. Yuan Yading
 - A learned controller that proposes the next prompt for interactive contour refinement in oART
- Built MedInsider, a FHIR-style benchmark for evaluating medical LLM agent integrity under institutional pressure; designed 840 paired-twin scenarios across 5 regulatory failure families and 8 incentive conditions.
 - Implemented simulated EHR tools, deterministic action-log scoring, run auditing, backup/restore, and reproducibility manifests for tool-use agent evaluation; Under Review
- Developed SpineFairBench, a counterfactual VLM benchmark auditing age/sex sensitivity in spinal radiology reports; generated 11.8K QC-passed paired X-ray counterfactuals with radiologist validation; Under Review

GS-11 Patent Examiner (AU: 2613) | USPTO

January 2025 – Present

- Evaluated 50+ machine learning and imaging patent applications, determining novelty and patentability through comprehensive prior art searches
- Awarded commendation letter from OPQA Director for office action quality; 104% production

Founder & Software Engineer | District Hut LLC

June 2020 – January 2025

- Led full-stack development projects generating over seven figures in revenue. Built patient-management systems, auto-dealer platforms, and restaurant applications; Registered trademark for company slogan

Open Source Contributions

- Google JAX — Added complex-input support to `jax.scipy.special.gamma` using Lanczos approximation, reflection handling, and SciPy-parity tests
<https://github.com/jax-ml/jax/pull/36521>
- Azure SDK for Python — Fixed `azure-data-tables` kwarg leakage into `async` transports; added `sync/async` regression tests.
<https://github.com/Azure/azure-sdk-for-python/pull/46137>
- Google DeepMind Optax — Implemented BYOL, SimSiam, DINO, and Barlow Twins self-supervised losses with JIT/vmap support and full pytest coverage (PR submitted, under review)
<https://github.com/google-deeppmind/optax/pull/1695>

Project Highlights

Distributed LLM Security Gateway — OpenAI/Anthropic-compatible proxy for LLM security

- Built a distributed Rust-based gateway that inspects prompts and streaming model outputs for prompt injection, jailbreaks, PII/secret leakage, and agentic tool-abuse risks
- Designed low-latency data plane with policy enforcement, classifier caching, tenant-aware sharding, streaming interruption, and fail-open/fail-closed controls
- Implemented supporting control/telemetry architecture with versioned policy registry, structured request logging, anomaly detection, and benchmark evaluation against public LLM security datasets

Mixture-of-Recurions (MoR) — <https://github.com/ahmedtaha100/MoR>

- Implemented a Google NeurIPS 2025 MoR in PyTorch — recursive parameter sharing with token routing (~70M parameters matching 360M vanilla performance)
- Built 2 routing and 2 KV caching strategies with full training stack, pytest suite, and routing heatmaps

Skills

Languages: Python, C/C++, Java, SQL

Frameworks & Libraries: PyTorch, JAX/Flax, TensorFlow, FAISS, LightGBM, FastAPI

Research: Deep Learning (CNNs, Transformers, LSTMs), Computer Vision, Medical Imaging, VLM Evaluation, Diffusion Models, Parameter-Efficient Methods (LoRA, SVD), Side-Channel Cryptanalysis