

TASSALLAH AMINA ABDULLAHI

she/her/hers ♦ Providence, RI

tassallah_abdullahi@brown.edu ♦ tassabdul.github.io

RESEARCH PROFILE

Research Scientist / Engineer specializing in the evaluation, alignment, and behavioral control of large language models. My work focuses on building robust evaluation pipelines (including LLM-as-a-Judge and benchmarks), analyzing model failure modes, and designing safety guardrails across multilingual, multimodal, and high-stakes settings.

Core Interests: Model Alignment & Behavior, LLM Evaluation, Responsible AI, Human–AI Interaction, Multilingual NLP.

RESEARCH AND INDUSTRY EXPERIENCE

Humana

Senior Responsible AI Data Scientist, R&D

Remote, USA
2026 – Present

- Build evaluation frameworks that assess LLM capability, reliability, safety, and behavioral alignment.
- Design and optimize LLM-as-a-Judge evaluation pipelines to measure model quality and reduce evaluator bias.
- Develop guardrails and responsible-AI methods to identify, mitigate, and monitor generative-AI risks.
- Translate qualitative failure-mode analysis into actionable evaluation and model-improvement strategies.
- Conduct AI risk assessment and governance research supporting safe deployment of foundation models.

Optum AI, UnitedHealth Group

Graduate Research Intern – AI/ML

Remote, USA
Jun 2025 – Oct 2025

- Built retrieval-augmented generation systems improving factual grounding and reliability of LLMs.
- Designed adversarial evaluations targeting hallucination and prompt-injection failure modes.
- Fine-tuned embedding models with contrastive learning to improve semantic retrieval.
- Built evaluation frameworks combining automated metrics, adversarial testing, and expert review.

Brown University

Ph.D. Candidate, Computer Science

Providence, RI
Sept 2021 – Present

- Research on foundation-model evaluation, behavioral control, alignment, and multilingual NLP.
- Study how prompting strategies, personas, and context shape model behavior and decision-making.
- Develop benchmarks for LLM safety, reasoning reliability, and behavioral consistency.

SELECTED PUBLICATIONS

Model Evaluation, Safety, and Behavior

- **The Persona Paradox: Medical Personas as Behavioral Priors in Clinical Language Models.** ACL ARR 2026.
- **Benchmarking is Broken—Don’t Let AI Be Its Own Judge.** NeurIPS 2025 (Position Paper).
- **UbuntuGuard: A Culturally-Grounded Policy Benchmark for Equitable AI Safety in African Languages.** ACL 2026.
- **AfriVox: Probing Multilingual and Accent Robustness of Speech & Multimodal LLMs.** EACL 2026.

Foundation Models, Reasoning, and Retrieval

- **K-Paths: Reasoning Over Graph Paths for Drug Repurposing and Interaction Prediction.** KDD 2025.
- **Retrieval-Augmented Zero-Shot Text Classification.** ACM SIGIR ICTIR 2024.

Generative AI in Healthcare

- **AfriMed-QA: A Pan-African Multi-Specialty Medical QA Dataset.** ACL 2025. Best Social Impact Paper Award.

- **Afrispeech-Dialog: A Benchmark for Spontaneous Healthcare Conversations in African English.** NAACL 2025.

EDUCATION

Brown University <i>Ph.D. Candidate in Computer Science</i> Thesis: <i>Towards Trustworthy Clinical AI: Knowledge Grounding, Inference Reliability, & Behavioral Control</i> Advisors: Carsten Eickhoff and Ritambhara Singh; Sc.M. Computer Science (GPA 4.00/4.00), May 2023	Providence, RI Sept 2021 – May 2026
University of Cape Town <i>M.Sc. Computer Science (Distinction)</i>	Cape Town, South Africa May 2021
Bayero University Kano <i>B.Sc. (Honors) Computer Science</i>	Kano, Nigeria May 2014

TECHNICAL SKILLS

Programming	Python, Java, R, SQL
ML Frameworks	PyTorch, TensorFlow, Hugging Face Transformers, scikit-learn
LLM Methods	Evaluation, LLM-as-a-Judge, RAG, embeddings, fine-tuning, safety/guardrails, prompting
ML Systems	FAISS, LangChain, Azure, Google Cloud
Languages	English, Hausa, Yoruba, Pidgin English

AWARDS AND LEADERSHIP

- ACM SIGHPC Computational and Data Science Fellowship (2022–2025); Brown University Diversity Fellowship.
- ACL 2025 Best Social Impact Paper Award; National Research Foundation Scholarship, South Africa.
- Led multilingual healthcare-AI dataset collaborations across African institutions; mentored undergraduate researchers.