Moussa Doumbia, Ph.D.

dmoussa37922@gmail.com — Faculty Profile

Research Statement

I am an applied mathematician whose research sits at the intersection of **large language models, responsible AI, and biomedical science**. My work spans fine-tuning transformer models for mental health crisis detection, auditing racial bias in clinical prediction models, applying deep learning to medical imaging, and using NLP to synthesize the biomedical literature at scale. Mathematically, I draw on statistical learning theory, optimal control, and dynamical-systems modeling to ground my AI methods.

My broader research goal is to develop **AI systems that are both capable and trustworthy**—language models and decision-support tools that reason carefully, communicate faithfully, and behave reliably in high-stakes domains such as healthcare. My combination of mathematical rigor, hands-on LLM experience (Hugging Face, transformers, BERT fine-tuning), and applied work in fairness and bias auditing reflects a commitment to responsible AI development that benefits diverse communities.

Education

- **Ph.D., Mathematics** — Howard University, 2017
  Dissertation: Mathematical modeling of malaria transmission dynamics
- **B.S., Mathematics** — University of the District of Columbia, 2007

Academic Appointments

Howard University, Washington, DC

- Senior Instructor, Department of Mathematics — Fall 2025 – present
- Master Instructor, Department of Mathematics — Fall 2023 – Spring 2025
- Lecturer, Department of Mathematics — Fall 2017 – Spring 2023

United States Military Academy, West Point, NY

- Visiting Professor, Department of Mathematical Sciences — June 2024 – June 2025

AI & Machine Learning Research

Active projects:

- *Detecting Mental Health Crises in Social Media Text* — Two-phase BERT fine-tuning on open Hugging Face datasets; evaluating model calibration and out-of-distribution robustness; targeted for **CLPsych 2027** (with student collaborator, in progress).
- *Algorithmic Fairness in Healthcare* — Systematic audit of racial bias in clinical prediction models; intersects with responsible AI evaluation methodology (with student collaborator, in progress).
- *Deep Learning for Sickle Cell Disease Detection* — CNN classification of peripheral blood smear images; focuses on low-resource, high-impact diagnostic settings (with student collaborator, in progress).
- *Reinforcement Learning for Diabetes Management* — Deep Q-Network approach to adaptive treatment recommendations; safety-constrained reward shaping to avoid harmful dosing (with student collaborator, in progress).
- *AI-Driven Extraction of Causal Pathways in Malaria Research* — NLP pipeline over the biomedical literature to surface mechanistic evidence linking malaria to educational outcomes (in progress).
- *Time-Series Forecasting of Malaria Incidence* — Statistical and ML multi-site forecasting models (in progress).

Completed:

- *AI-Based Thematic Analysis of Biomedical Literature for Precision Nutrition* — Published in *The Journal of Nutrition*, 2026 (see Publications).
- *Reviewing the Genetics and Exercise Literature with AI* — NLP-assisted systematic review; submitted to *Springer Nature*.

Technical Skills

LLMs and NLP: Hugging Face Transformers, BERT fine-tuning, spaCy, NLTK, prompt engineering, retrieval-augmented generation (RAG), text classification, named-entity recognition, sequence-to-sequence modeling, model calibration and evaluation.

Deep learning and ML: PyTorch, CNNs, reinforcement learning (DQN), scikit-learn, time-series models, uncertainty quantification, statistical learning theory.

Mathematics and methods: Optimization, inverse problems, optimal control, dynamical systems, differential equations, mathematical epidemiology.

Languages and tools: Python (NumPy, pandas, PyTorch, Hugging Face), R, MATLAB, SQL, Git/GitHub, Jupyter, LaTeX.

Responsible AI: Bias auditing, fairness metrics (demographic parity, equalized odds), model interpretability, safety-constrained RL, dataset documentation.

Grants and Funded Initiatives

- **Co-Investigator, NIH/NIMHD** — *Enhancing Data Science Capability and Application at Howard University*, Grant 3U54MD007597-33S2, $386,250.
- **Co-Investigator, MAA NREUP** — National Research Experience for Undergraduates Program, $28,999; mentored undergraduates on data-science and AI research projects.
- **Virtual Applied Data Science Training Institute (VADSTI)** — Faculty member building data-science training for minority-serving institutions.
- **Microsoft Visiting Data Science Professor Program** — Selected participant, Summer 2021.
- **NSA Social Network Research Program** — Team member, Howard/University of Delaware/NSA collaboration, 2020–2021.

Selected Publications

1. Day, J. L., Beckman, J. R., Nelson, R., Weeks, Q., Dorta, J., **Doumbia, M.**, and Thomas, D. M. "Artificial Intelligence-Based Thematic Analysis of Biomedical Literature for Precision Nutrition." *The Journal of Nutrition*, 156(5), 101472, May 2026.
2. **Doumbia, M.** and Yakubu, A. "Asymptomatic Malaria Infections in Pregnant Women of Ngbo in Ohaukwu Local Government Area of Ebonyi State, Nigeria." *Springer Nature — La Matematica*, 2025.
3. **Doumbia, M.** and Yakubu, A. "Malaria Incidence and Anopheles Mosquito Density in Irrigated and Non-Irrigated Villages of Niono in Mali." *Discrete and Continuous Dynamical Systems — Series B*, 22(1), 2017.

Teaching and Curriculum Development

- **Natural Language Processing** — Designed and taught a full-semester course covering text preprocessing, classification, transformers, and large language models (U.S. Military Academy, Fall 2024).
- **Introduction to Data Science** — Created Howard's inaugural Data Science course (Python, data wrangling, visualization, core ML), 2021–2023.
- **New B.S. in Data Science** — Core contributor to Howard's new major with concentrations in Artificial Intelligence, Health Analytics, and Business Analytics.
- **Coursera: Linear Algebra for Data Science Using Python** — Co-created Howard's global online course on computational linear algebra through Python (ongoing enrollment).

Honors, Fellowships, and Recognitions

- **AWS-MLU Faculty Fellow (Nominee)**, Amazon Web Services Machine Learning University, 2026 — Recommended through the Howard AI Network to co-lead faculty workshops, partner with the University of Michigan's Distributed Teaching Collaborative Consortium, and develop AWS-integrated course materials ($10,000 annual stipend).
- **ASCEND-AI Faculty Learning Community**, Howard University, 2026 — Selected cohort focused on AI literacy, responsible AI in teaching and learning, and AI-integrated curriculum design.

- Outstanding Faculty, Department of Mathematics, Howard University, 2021–2022.
- Edward A. Bouchet Fellowship, Howard University, 2008–2013.
- Visiting Scholar, Mathematical Biosciences Institute, Ohio State University, 2011.

Invited Talks and Professional Service

- **Joint Mathematics Meetings (JMM), 2026** — Invited talk: *From Timbuktu to Howard: Reflections on a Journey Through Mathematics and Legacy — Celebrating 50 Years of the Ph.D. Program in Mathematics*.
- **APHA 2026 Annual Meeting and Expo** — Abstract submitted (ID 594107): *Together Advancing Health Interventions: AIHEC and AIM-AHEAD Building Capacity in AI/ML among Tribal Colleges and Universities* (under review).
- **2024 AMS Spring Eastern Sectional Meeting**, Howard University — Session Moderator.
- **NSF-CBMS Regional Conference on Mathematical Biology**, Howard University, 2018 — Panel Organizer.

Student Research Mentored

Howard Mathematics / Data Science Open House, April 2026:

- *Detecting Mental Health Crises in Social Media Text: A Two-Phase BERT Fine-Tuning Approach* — Bitania T. Molla. Targeted for CLPsych 2027.
- *Algorithmic Fairness in Healthcare: Analyzing Racial Bias in Clinical Prediction Models* — Sydney Helstone.
- *Reinforcement Learning for Diabetes Management: A Deep Q-Network Approach* — Alana King.
- *Deep Learning for Sickle Cell Disease Detection: CNNs for Peripheral Blood Smear Image Classification* — Shirjak Thokar.

Earlier projects:

- *Advanced EDA and Machine Learning for COVID-19 Data Analysis*, Summer 2023.
- *COVID-19 Prediction using the SEIRAF Model* — Computational epidemiological modeling in Python, Summer 2023.