

# AGGREY MUHEBWA, PhD

Seattle, WA 98109 | muhebwa.grey@gmail.com | linkedin.com/in/amuhebwa

## APPLIED AI SCIENTIST | PRODUCTION LLM SYSTEMS | RESPONSIBLE AI DEPLOYMENT

---

Applied Scientist with 3+ years building reliable ML systems for high-stakes domains. Currently engineering production LLM pipelines at Amazon for sustainability impact areas, delivering over 90% precision and 14x faster than commercial tools. Strong research background in deep learning, model interpretability, energy-aware AI and federated learning optimization. Proven ability to balance technical rigor with stakeholder trust - expertise in failure-mode analysis, confidence calibration and cross-functional deployment. Recognized with UN-IPCC Fellowship and \$100K Stanford Precourt Institute for Energy grant for Sustainable AI research.

### Work Experience

---

**Applied Scientist**, Amazon – Seattle, WA August 2025 – Present

Engineered a production-scale LLM monitoring system for near-real-time human rights and carbon risk assessment across 350,000+ global suppliers, compressing 24-month manual audit cycles into continuous, automated intelligence.

- Designed and deployed a production NLP pipeline processing 160,000+ news sources daily across 100+ languages, achieving over 90% precision in violation detection, 14x faster and at a fraction of the cost of \$360K/year commercial alternatives
- Engineered a multi-stage classification system optimized for high-stakes investigations. Pipeline combines LLM-based relevance screening (Claude, Qwen, Mistral) → fine-tuned violation classifiers (BERT, RoBERTa) → context-based entity resolution (RAG). Enables end-to-end mapping of news incidents to specific suppliers across 350,000+ supplier base.
- Reduced false positive rate from 65% to 7% through consensus labeling, feature attribution analysis, and systematic failure-mode documentation, directly increasing investigator adoption and enabling team to focus on verified high-risk cases

**Machine Learning Engineer**, Flume Water – San Luis Obispo, CA June 2025 – August 2025

- Optimized ML algorithms for predicting household water consumption in non-metered homes using smart meter data, improving prediction accuracy from 55% to 85%. Scaled from California to all 48 contiguous US states

**Postdoctoral Researcher**, Stanford University – Stanford, CA Nov 2023 – August 2025

- Awarded \$100K competitive grant from Stanford Precourt Energy Institute to develop a responsible AI model selection framework that enables practitioners to make principled accuracy-vs-emissions tradeoffs. Demonstrated over 30% reduction in deployment carbon footprint while maintaining baseline model accuracy
- Developed a synchronization-based federated optimization algorithm (Kuramoto-FedAvg) that achieves 40% faster convergence than standard methods (FedAvg, SCAFFOLD) under extreme data heterogeneity, enabling stable federated training across non-IID client distributions
- Developed an LLM-based framework for automated classification of levee failure mechanisms from unstructured inspection reports. Combined zero-shot learning (GPT-4) with fine-tuned domain-adapted models, outperforming baseline classifiers by 21%. Enables scalable, reproducible levee vulnerability assessment under shifting climate conditions
- Designed a decentralized federated learning framework for training river discharge prediction models on transboundary hydrology data. Addresses data privacy, heterogeneity, and equitable access constraints across riparian stakeholders. Demonstrated on real-world datasets spanning 5 geographic regions. Federated models achieve comparable prediction accuracy to centralized baselines while keeping raw data local to each custodian
- Developed a causal distillation framework for transferring structured reasoning capabilities from large to compact language models using explanation-based training. Introduced a novel evaluation metric for measuring causal alignment in model-generated explanations. Enables deployment of causally consistent reasoning in resource-constrained environments

**Machine Learning & Remote Sensing Consultant**, World Bank – Washington DC May 2022 – May 2023

- Designed a multi-modal road quality prediction system combining vision transformers applied to multi-resolution, multi-modal satellite imagery with economic development features. Enabled automated

infrastructure assessment across the DRC at scale, replacing manual field inspections and directly informing World Bank funding decisions for over \$100M in infrastructure investments

**PhD Researcher**, University of Massachusetts – Amherst, MA Aug 2018 – Oct 2023

**Electrical & Computer Engineering** (Advisor: Prof. Jay Taneja)

UN IPCC Scholar (2021-2023). 1 of 12 scholars globally selected biannually for climate impact

Dissertation: Incorporating Machine Learning with Satellite Data to Support Critical Infrastructure Measurement and Sustainable Development

- Developed an ML approach for discharge prediction in regions with limited to no observational data. Enables hydrologists to interpret learned model representations as physically meaningful drivers of streamflow across different climate regimes. Addresses a key barrier to ML adoption in hydrology. Published in Water Resources Research (top 1% journal in Environmental Engineering).
- Developed a resource-efficient ML approach for infrastructure monitoring in low-resource settings, achieving comparable accuracy to complex deep learning baselines at 10x lower compute cost through targeted feature engineering, enabling deployment in regions without access to computation infrastructure and satellite coverage
- Deployed ML-driven infrastructure assessment in conflict-affected DRC as World Bank consultant. Navigated unique constraints of limited ground-truth data, security restrictions on field validation, and non-technical decision-makers. Built interpretability layer including uncertainty quantification and failure-mode documentation to enable trusted adoption for over \$100M+ funding allocation.

**Machine Learning Engineer, Intern**, Atlas AI – Palo Alto, CA May 2020 – Aug 2020

- Optimized data processing and inference components of an ML pipeline for evaluating spatial anonymization effects on geospatial model performance. Improvements enabled faster experimentation across anonymization thresholds and privacy-performance tradeoff analysis

## Additional Work Experience

---

- Visiting Scientist, Marck, Sharpe & Dorme – Prague, Czech Republic (May 2018 - July 2018)
- Fullstack Software Engineer, Grameen Foundation – Kampala, Uganda (Jan 2015 - July 2016)
- Software Engineer, Mountbatten Limited – Kampala, Uganda (June 2013 - Dec 2014)

## Education

---

**University of Massachusetts Amherst** Aug 2018 – Oct 2023

- PhD in Electrical & Computer Engineering

**Carnegie Mellon University** July 2016 – Dec 2017

- Masters of Science in Electrical and Computer Engineering
- **Relevant coursework:** Machine Learning, Machine Learning for Signal Processing, Deep Learning, Data Science, Data Structures and Algorithms

**Makerere University**, Bachelors of Science in Computer Engineering Aug 2009 – April 2013

## Awards & Scholarships

---

- Awarded \$100,000 grant (as a co-PI) through the Bits and Watts Initiative to conduct research on Socio-psychological perspectives in AI model selection for performance and sustainability (2025)
- United Nations Intergovernmental Panel on Climate Change (IPCC) Scholar (2021-2023)
- Meta (Facebook) Uniting Scholars in Research Recipient (Oct. 2022)
- Google CS Research Mentorship Program Fellow (Class of 2021B)
- College of Engineering Teaching Fellow, University of Massachusetts - Amherst (2020)
- Winner - Nelson Mandela African Institute of Science and Technology Africa Grand Challenge (2018)
- NSF travel grant, International Conference on Computing in Civil and Engineering (2024)
- 9th California Data Summit Grant for Speakers (2024)

## Skills

---

### Core ML/AI:

- LLMs & Foundation Models: Production deployment (Opus, Llama, Qwen), fine-tuning (BERT, RoBERTa), prompt engineering, RAG architectures, multi-LLM consensus systems
- Deep Learning: PyTorch, TensorFlow, Keras; NLP, computer vision, generative AI
- Specialized: Federated learning optimization, model distillation & compression, physics-informed neural networks

### Production & Engineering:

- MLOps: Model monitoring, A/B testing, distributed training, automated ML pipelines
- Cloud & Deployment: AWS, GCP, Docker, API integration, production environments
- Software Engineering: Python (expert), R, SQL, MATLAB; clean code, testing, version control

### Research & Responsible AI:

- Interpretability: Feature attribution (SHAP, LIME), attention analysis, failure-mode documentation, uncertainty quantification
- Model Evaluation: Precision/recall optimization for high-stakes domains, confidence calibration, robustness testing across distributions
- Stakeholder Collaboration: Translating ML for legal/compliance/policy audiences, building trust through transparency

## Selected Publications(18 total, h-index:7, 260+ citations)

---

- **Aggrey Muhebwa** and Khalid K. Osman: A Behavioral Finance Framework for Balancing AI Accuracy and Operational Carbon Emissions. *In the ACM Journal on Computing and Sustainable Societies*, 2025
- **Aggrey Muhebwa**, Lebrand, M. S. M., Steinbuks, J., & Taneja, J. K.. Eyes on the Road: Evaluating Road Quality over Time Using Satellite Imagery in Eastern Democratic Republic of the Congo. *World Bank Policy Commons*
- **Aggrey Muhebwa**, Gabriel Cadamuro, and Jay Taneja. Pixel Perfect: Using Vision Transformers to Improve Road Quality Predictions from Medium Resolution and Heterogeneous Satellite Imagery. *In the ACM Journal on Computing and Sustainable Societies*, 2023
- **Aggrey Muhebwa**, C.J. Gleason, Dongmei Feng, and Jay Taneja. Improving Discharge Predictions in Ungauged Basins: Harnessing the Power of Disaggregated Data Modeling and Machine Learning. *In the Journal of Water Resources Research (WRR)*, 2024
- **Aggrey Muhebwa** & Khalid K. Osman. Promoting Equitable Access to Transboundary Hydrology Data: A Framework for Secure Sharing and Collaborative Machine Learning to Enhance River Discharge Prediction. *Proceedings for the Journal of Computing in Civil Engineering*, 2024
- **Aggrey Muhebwa** and Khalid K. Osman. Using Large Language Models for Systematic Failure Model Identification in Levee Infrastructure. *Proceedings for the Journal of Computing in Civil Engineering*, 2025
- **Aggrey Muhebwa**, Khotso Selialia, Fatima Anwar & Khalid K. Osman: Kuramoto-FedAvg: Using Synchronization Dynamics to Improve Federated Learning Optimization under Statistical Heterogeneity. *[preprint available on arxiv]*

Additional Research:11 papers spanning infrastructure ML, water resource prediction, satellite imagery analysis, and LLM applications. Venues include Water Resources Research, NeurIPS Workshop, and ACM COMPASS

**Full list:** Google Scholar Profile