

Prince Osei Aboagye

Curriculum Vitae

Amazon - Day 1
2112 6th Ave
Seattle, Washington, 98121
915-867-2874
prince@cs.utah.edu

Research Interests

Multimodal & Agents: LLM Post-Training, Vision-Language Alignment, Reasoning (MCTS), Reinforcement Learning.

Foundations & Efficiency: Data-Centric AI, Synthetic Data Generation, Model Compression, Computational Geometry, High-Dimensional Representation Learning.

Work Experience

2024 - Present Amazon, Applied Scientist, Amazon FireTV - Science, Seattle, WA

- Role:
- **Large-Scale LLM Training & Optimization:** Co-led large-scale Transformer training for a production Query Understanding model. Integrated FlashAttention and Mixture-of-Experts (MoE) to maximize GPU throughput. As well as optimized sparse expert routing and applied distributed training strategies (data/model parallelism), resulting in a 10% performance increase in AIA and ASA performance metrics.
 - **Agent-Driven Synthetic Data Generation:** Architected *AFEG* (Agent-Driven Framework for Evaluating Groundedness), a multi-agent system (Appraiser, Inquirer, Inspector) to synthesize adversarial “*weakness-targeted*” test queries. Utilized “*Importance Sampling*” to generate high-error-probability synthetic data, overcoming the “*Data Wall*” of static benchmarks.
 - **Graph-Based Hallucination Detection:** Developed *PGIRE* (Prompted Graph-Inspired Response Evaluation), a novel framework that extracts Knowledge Graph triples (Subject-Relation-Object) from LLM outputs to mathematically verify factual consistency against API evidence. This effectively replaced semantic similarity checks with fact-based grounding for video playback and media discovery agents.
 - **Post-Training Data Curation (SFT & CoT):** Co-spearheaded the data curation pipeline for Large Language Model Supervised Fine-Tuning (SFT). Engineered “*Golden Datasets*” by mining production search logs via Athena/SQL and synthesizing Chain-of-Thought (CoT) reasoning traces to improve model instruction-following and reasoning capabilities.
 - **Media Data Foundation:** Built the “*Ground Truth*” foundation for Video Search by joining large-scale “*Search Screen*” and “*Profile Event*” datasets (millions of rows). Solved complex “*Title Query*” disambiguation problems using multi-level heuristic filtering and fuzzy matching logic to clean noisy interaction data.

2023 - 2024 Visa Research, Research Scientist, Palo Alto, CA

- Research
- **Generative AI for Complex Sequences:** Co-developed *TransactionGPT*, adapting LLMs via specialized tokenization strategies to model continuous variables, achieving SOTA performance in anomaly detection and sequential prediction tasks.
 - **Parameter-Efficient Transformer Fine-Tuning (LECSiV):** Developed an advanced parameter-efficient fine-tuning method extending LoRA through structured SVD-based initialization and rank-aware updates, reducing memory footprint and improving convergence speed in LLaMA-based models. Enabled stable large-scale fine-tuning under constrained GPU memory in distributed training environments.
 - **Multimodal Data Curation & Perception:** Designed visual analytics and *Human-in-the-Loop* workflows for user-guided image captioning, directly facilitating the alignment and curation of large-scale image-text datasets.

- **Data Debiasing & Safety:** Developed an interpretable debiasing framework using Iterative Orthogonalization to surgically remove toxic/biased/stereotypical patterns from representations without degrading model utility, addressing key challenges in organic data curation.
- **Scalable Embedding Alignment:** Developed *Quantized Wasserstein Procrustes*, a scalable optimal transport algorithm to efficiently align distinct high-dimensional distributions (e.g., multimodal/multi-lingual) for foundation model pre-training.
- **Geometric Data Mixing:** Proposed novel *Normalization of Language Embeddings* techniques to restructure vector space geometry, establishing a theoretical framework that significantly enhances alignment accuracy across diverse domains and modalities.
- **Fundamental Metric Learning:** Challenged prevailing paradigms in our *One-Hot Encoding Strikes Back paper*, demonstrating that fixed, orthogonal representations maximize class separability and stability while eliminating the overhead of dynamic prototypes.
- **Adversarial Robustness & Interpretability:** Proved that fixed, coordinate-aligned embeddings force the learning of sparse, interpretable features, significantly enhancing defense against gradient-based adversarial attacks compared to standard Softmax approaches.

Summer 2022 **Research Scientist Intern**, Visa Research

Mentor: Wei Zhang

Summer 2021 **Research Scientist Intern**, Visa Research

Mentor: Wei Zhang

Education

2018 – 2023 **PhD**, *Computer Science*, University of Utah

Advisor **Dr. Jeff Phillips**

2016 – 2018 **MS**, *Applied Mathematics*, University of Texas at El Paso

Advisor **Dr. Michael Pokojov**

2011 – 2015 **BA**, *Mathematics and Economics*, University of Ghana

Advisor **Mrs. Lilian Frempomaa Kyei**

Patents

○ Zheng, Yan, Michael Yeh, Wang Junpeng, Wei Zhang, Liang Wang, Hao Yang, and **Prince Osei Aboagye**. “*Method, System, and Computer Program Product for Normalizing Embeddings for Cross-Embedding Alignment.*” U.S. Patent Application 18/006,649, filed July 6, 2023.

<https://patents.google.com/patent/US20230214177A1/en>

○ Zheng, Yan, Liang Wang, Wei Zhang, Zhongfang Zhuang, Hao Yang, and **Prince Osei Aboagye**. “*Method, System, and Computer Program Product for Embedding Normalization and Orthogonality of Embeddings.*” U.S. Patent Application US2024/0386327A1, published Nov. 21, 2024.

<https://patents.google.com/patent/US20240386327A1/en>

○ Zheng, Yan, **Prince Osei Aboagye**, Zhongfang Zhuang, Michael Yeh, Junpeng Wang, Liang Wang, Javid Ebrahimi, and Wei Zhang. “*Method, System, and Computer Program Product for Unsupervised Alignment of Embedding Spaces.*” U.S. Patent Application US2025/0238480A1, published July 24, 2025.

<https://patents.google.com/patent/US20250238480A1/en>

Ph.D. Dissertation

Thesis 2023: Understanding the Geometry of Structured Vectorized Representations. **Prince Osei Aboagye**, Doctoral Dissertation, The University of Utah, 2023. **Link:** Ph.D. Dissertation

Publications

- arXiv Preprint 2025:** *TransactionGPT*. Yingtong Dou, Zhimeng Jiang, Tianyi Zhang, Mingzhi Hu, Zhichao Xu, Shubham Jain, Uday Singh Saini, Xiran Fan, Jiarui Sun, Menghai Pan, **Prince Osei Aboagye**, Yan Zheng, Liang Wang, Wei Zhang, and others. **Link to paper:** <https://arxiv.org/abs/2511.08939>
- ICDM 2024:** *Matrix Profile for Anomaly Detection on Multidimensional Time Series*. Chin-Chia Michael Yeh, Audrey Der, Uday Singh Saini, Vivian Lai, Yan Zheng, Junpeng Wang, Xin Dai, Zhongfang Zhuang, Yujie Fan, Huiyuan Chen, **Prince Osei Aboagye**, Liang Wang, Wei Zhang, Eamonn Keogh. **Link to paper:** <https://www.arxiv.org/abs/2409.09298>
- IEEE PacificVis 2024:** *Visual Analytics for Efficient Image Exploration and User-Guided Image Captioning*. Yiran Li, Junpeng Wang, **Prince Osei Aboagye**, Michael Yeh, Yan Zheng, Liang Wang, Wei Zhang. **Link to paper:** <https://arxiv.org/abs/2311.01016>
- KDD 2024:** *RPMixer: Shaking Up Time Series Forecasting with Random Projections for Large Spatial-Temporal Data*. Chin-Chia Michael Yeh, Yujie Fan, Xin Dai, Vivian Lai, **Prince Osei Aboagye**, Junpeng Wang, Huiyuan Chen, Yan Zheng, Zhongfang Zhuang, Liang Wang, Wei Zhang. **Link to paper:** <https://dl.acm.org/doi/10.1145/3637528.3671881>
- Preprint 2024:** *Has Your Pretrained Model Improved? A Multi-head Posterior Based Approach*. **Prince Osei Aboagye**, Yan Zheng, Junpeng Wang, Uday Singh Saini, Xin Dai, Michael Yeh, Yujie Fan, Zhongfang Zhuang, Shubham Jain, Liang Wang, Wei Zhang. **Link to paper:** <https://arxiv.org/abs/2401.02987>
- Preprint 2024:** *PDT: Pretrained Dual Transformers for Time-aware Bipartite Graphs*. Xin Dai, Yujie Fan, Zhongfang Zhuang, Shubham Jain, Chin-Chia Michael Yeh, Junpeng Wang, Liang Wang, Yan Zheng, **Prince Osei Aboagye**, Wei Zhang. **Link to paper:** <https://arxiv.org/abs/2306.01913>
- Preprint 2024:** *One-Hot Encoding Strikes Back: Fully Orthogonal Coordinate-Aligned Class Representations*. **Prince Osei Aboagye**, Hasan Pourm Mahmoodaghababa, Yan Zheng, Chin-Chia Michael Yeh, Junpeng Wang, Huiyuan Chen, Xin Dai, Liang Wang, Wei Zhang, Jeff Phillips. **Link to paper:** <https://openreview.net/forum?id=1C0Gu2BMpf¬eId=xJm0jrT8ru>
- ICLR 2023:** *Interpretable Debiasing of Vectorized Language Representations with Iterative Orthogonalization*. **Prince Osei Aboagye**, Yan Zheng, Jack Shunn, Chin-Chia Michael Yeh, Junpeng Wang, Zhongfang Zhuang, Huiyuan Chen, Liang Wang, Wei Zhang, Jeff Phillips. **Link to paper:** <https://openreview.net/forum?id=TKQ1sxd9P4>
- ICLR 2022:** *Normalization of Language Embeddings for Cross-Lingual Alignment*. **Prince Osei Aboagye**, Jeff Phillips, Yan Zheng, Junpeng Wang, Chin-Chia Michael Yeh, Wei Zhang, Liang Wang, Hao Yang. **Link to paper:** <https://openreview.net/forum?id=Nh7CtbyoqV5>
- AMTA 2022:** *Quantized Wasserstein Procrustes Alignment of Word Embedding Spaces*. **Prince Osei Aboagye**, Yan Zheng, Michael Yeh, Junpeng Wang, Zhongfang Zhuang, Huiyuan Chen, Liang Wang, Wei Zhang, Jeff Phillips. **Link to paper:** <https://aclanthology.org/2022.amta-research.15/>

Academic Research Projects

Music Genre Classification & Audio Feature Analysis, University of Utah

- Focus:
- **Media Perception & Feature Extraction:** Engineered an end-to-end audio pipeline on the GTZAN dataset to extract structured features for model training.
 - **Deep Learning Architectures:** Optimized Deep Neural Networks (DNNs) in Keras/TensorFlow, configuring dense architectures with ReLU, dropout, and Softmax to classify spectral features across 10 genres.
 - **Data Curation & Benchmarking:** Managed preprocessing (label encoding, scaling) and established rigorous benchmarks by evaluating model performance against KNN, Random Forest, XGBoost, and SVM baselines.

Visualizing High-Dimensional Data & Model Interpretability, University of Utah

- Focus:
- **Interactive Visual Analytics:** Developed the “*Projection Analysis Tool*” (PAT) utilizing interactive 2D/3D scatter plots to interpret high-dimensional embedding geometry and complex data structures.
 - **Dimensionality Reduction & Metrics:** Applied advanced dimensionality reduction techniques (e.g., PCA, t-SNE) to project high-dimensional feature spaces into interpretable 2D/3D visualizations, enabling the identification of clusters, outliers, and data quality issues in large-scale datasets.

Service

- Conference Reviewer **International Conference on Machine Learning, ICML 2026**
- Conference Reviewer **International Conference on Machine Learning, ICML 2025**
- Conference Reviewer **Knowledge Discovery and Data Mining, KDD 2024**
- Conference Reviewer **International Conference on Machine Learning, ICML 2024**
- Conference Reviewer **Conference on Neural Information Processing Systems, NeurIPS 2023**
- Conference Reviewer **International Conference on Learning Representations, ICLR 2023**