DARSHAN ASHOK FOFADIYA

Senior Applied Scientist – Amazon

253-368-9607 | fofadiyadarshan@gmail.com | linkedin.com/in/darshan-fofadiya | Seattle, WA

RESEARCH SUMMARY

Senior Applied Scientist with **11+ years of experience** architecting large-scale ML systems. Specializing in **High-Performance Computing (HPC)**, **GPU optimization**, and **non-convex optimization landscapes**. Bridging the gap between theoretical guarantees and empirical scale by designing custom optimizers (GatedAdam) and novel architectures (Idea-Gated Transformers). Combining industrial rigor with independent research to solve stability and trainability in deep networks.

INDEPENDENT RESEARCH & NOVEL ARCHITECTURES

Idea-Gated Transformers: Latent Semantic Planning | *Under Review, ICML 2026*   2025 – Present
  - **Novel Architecture:** Proposed a Transformer variant that separates "concept planning" from "token generation" using differentiable soft-gating. This mitigates **associative drift** in long-horizon generation.
  - **Relevance to Physics:** The separation of "semantic state" from "syntactic generation" mirrors the need to maintain quantum state coherence against decoherence drift.
  - **Implementation:** Built custom PyTorch kernels to handle the gating mechanism efficiently on GPU.

GatedAdam: Biologically Inspired Sparse Optimization | *Preprint*   2025 – Present
  - **Solving Gradient Vanishing:** Developed a custom optimizer to tackle the **vanishing gradient problem** (analogous to Barren Plateaus) by using activation saturation as a constructive signal.
  - **Mechanism:** Implemented a "Benefit Score" gating mechanism ($O(1)$ overhead) to enforce **93.43% structural sparsity**, replacing standard regularization with a physical "braking system."
  - **Stability:** Stabilized training in non-convex regimes where standard Adam diverged.

PROFESSIONAL EXPERIENCE

Amazon   Seattle, WA
Senior Applied Scientist (L6)   *Oct 2023 – Present*
  - **HPC & Distributed Systems:** Architected a custom density-based clustering engine for **450k high-dimensional entities** across 20+ languages. Solved **topological collapse** in embedding spaces by engineering an "exponential growth trigger" with binary search, stabilizing graph construction where standard DBSCAN failed.
  - **GPU Optimization:** Optimized GPU memory utilization for distributed training runs, reducing compute cost by 40% while increasing batch size throughput.
  - **Causal Inference:** Applied Double Machine Learning (DML) to assess causal uplift, influencing strategic resource allocation for Amazon Books.

Amazon   Seattle, WA
Applied Scientist II   *June 2018 – Sep 2023*
  - **Robustness & Noise Tolerance:** Owned the architecture for **False Reject Detection** in Alexa. Developed semi-supervised frameworks to robustly learn from noisy, high-variance data in low-resource domains.
  - **Scale:** Deployed production-grade NER and Intent Classification models serving millions of daily requests with strict latency constraints.
  - Published research in **Findings of the ACL (2021)**.

ZS Associates   Pune, India / Evanston, IL
Data Scientist   *May 2015 – May 2018*
  - Built **embedding representations** for all US physicians to improve targeting and segmentation, preceding the widespread adoption of transformer-based embeddings in the pharma domain.

Mu Sigma   Bangalore, India
Decision Scientist   *Aug 2014 – May 2015*
  - Worked on customer segmentation and targeting problems for major US retail and pharma clients.

PUBLICATIONS & PATENTS

Preprint: *Idea Gated Transformers: Latent Semantic Planning.* arXiv:2512.03343 (Submitted ICML 2026).

Paper: *False Reject Detection for Low-Resource Domains in Neural Spoken Language Understanding.* Findings of the ACL-IJCNLP 2021.

Patent US11604925: *Architecture for gazetteer-augmented named entity recognition.* Issued Mar 2023.

Patent US11823671: *Architecture for context-augmented word embedding.* Issued Nov 2023.

EDUCATION

University of Mumbai   Mumbai, India
Bachelor of Engineering in Electrical Engineering   *Aug 2010 – May 2014*
– Focus on Complex Analysis, Advanced Mathematics, and Control Systems.

TECHNICAL SKILLS

HPC & Frameworks: PyTorch (Advanced), CUDA Optimization, Distributed Data Parallel (DDP), JAX
Optimization: Stochastic Gradient Descent Dynamics, Loss Landscape Analysis, Custom Optimizer Implementation
Algorithms: Graph Theory, Transformer Architectures, Reinforcement Learning (PPO), Density Clustering
Languages: Python, C++, SQL