Ayush Bharadwaj

AI Alignment Researcher / Engineer

linkedin.com/in/ayushbharadwaj
ayush.bharadwaj@gmail.com

Education

M.A. Mathematics | Algebraic Geometry, Optimization — 01/2020 – 01/2023
San Francisco State University
Advisor: Prof. Serkan Hosten
Thesis: Complex Critical Points of Deep Linear Networks
Coursework: Abstract Algebra, Measure and Integration, Complex Analysis, Functional Analysis

M.S. Operations Research — 08/2012 – 01/2014
Case Western Reserve University

B.Tech. Civil Engineering — 06/2000 – 07/2004
Indian Institute of Technology

Research Experience

Independent — 05/2024 – present
Supervised Program for Alignment Research (SPAR) — *San Jose, CA*
Research work (in-progress) on parametrization-invariant mechanistic interpretability of neural networks.
• Approximation of RLCT via Taylor expansion of the loss function
• Quantification of functional degeneracy in overparametrized models via Hessian rank

Publications & Preprints

• Complex Critical Points of Deep Linear Networks [arxiv:2301.12651]
• Functional Degeneracy in Neural Networks: Measurement and Pruning [NeurIPS 2026 Workshop]

Selected Writing

• Hessian analysis with JAX: a platform-agnostic, high-performance approach [lesswrong]
• Verifying Neural Network Robustness with Linear Programming [blog]

Software Projects

• verifiNN | Python library for Robustness Verification of Neural Networks using Optimization methods
• homotopy_continuatioNN | Julia application for generating and solving polynomial systems associated with identity activated neural networks

Employment

Data Scientist, Risk / Fraud Detection — 05/2016 – 10/2019
Postmates (acquired by Uber) — *San Francisco, CA*
• Statistical Learning • Imbalanced Classification • Ensemble Methods • Anomaly Detection

Risk Analyst, Consumer Risk — 05/2014 – 05/2016
PayPal — *San Jose, CA*
• Fraud detection • Revenue optimization • Machine learning • Statistical analysis

Software Engineer / Consultant — 06/2004 – 07/2012
Infosys Limited — *Bangalore, India*
• Software development • Query optimization • Systems integration • Incident management