Ayush Bharadwaj

Research Scientist / Engineer - AI Safety

ayusbhar2.github.io

ayush.bharadwaj@gmail.com

EDUCATION

M.A. Mathematics | Convex Optimization, Algebraic Geometry   01/2020 – 01/2023
San Francisco State University

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

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

PROJECTS

Robustness Verification of Neural Networks using Optimization Methods
  Personal project implemented in Python

Complex Critical Points of Deep Linear Neural Networks
  Thesis exploring the loss landscape using computational algebraic geometry

EXPERIENCE

AI Safety Researcher   05/2024 – present
Supervised Program for Alignment Research (SPAR)   *San Jose, CA*

• Research work (in-progress) investigating bounds on the local learning coefficient (LLC) via second (and higher) order Taylor series expansion of the behavioral loss.

• Implemented a common python module (using JAX) for identifying and ablating the degenerate directions of the loss function for various model architectures across the team.

Applied Machine Learning Scientist   05/2016 – 10/2019
Postmates (acquired by Uber)   *San Francisco, CA*

• Bootstrapped the company's machine learning system for automated payment fraud detection.

• Trained tree ensembles to mimic human decisions in the absence of ground truth.

• Promoted the use of cross entropy for diagnosing model performance under high class skew.

• Developed data pipelines for online and offline feature generation and storage.

• Built an experiment design framework to measure model performance in production.

Risk Analyst   05/2014 – 05/2016
PayPal   *San Jose, CA*

• Decision Analysis • Fraud Detection • Revenue Optimization • Machine Learning

Software Engineer   06/2004 – 07/2012
Infosys   *Bangalore, India*

• Software development • Query optimization • Systems Integration • Incident management