Ishan Garg

Website | Google Scholar | GitHub | LinkedIn

Experience

Zenon AI | July 2023 - present

Lead AI Forward Deployed Engineer. Built and deployed AI systems for financial document review and legacy code modernisation, and built an NLP system for matching candidates to roles.

Research training

ARENA 9.0 | October - November 2026

Mechanistic interpretability, reinforcement learning, language model evaluations, and AI control. Continued work on evaluation tooling through the extension with Generality Labs.

BlueDot Impact | June - July 2026

Technical AI Safety course

Education

BITS Pilani | July 2019 - June 2023

B.E. (Hons), Computer Science

Publications

Mechanistic Analysis of Alignment Algorithms in Language Models (arXiv, 2026)

On the Integration of Cross-Conformal Prediction, Ensembles, and Sampling for Uncertainty Quantification in One-Class Anomaly Detection (COPA, 2025)

Research projects