# Abhishek Divekar

Sr. Applied Scientist @Amazon · MSCS UT Austin

adivekar-utexas | Abhishek Divekar | Abhishek Divekar

## Research Interests

Large Language Models (post-training, evaluation, red-teaming, guardrailing), Synthetic Dataset Generation, Retrieval-Augmented Generation, LLM-as-a-Judge, Multimodal Agentic Systems, AutoML, Anomaly Detection, NLP.

## Education

### The University of Texas at Austin — Austin, TX

Master of Science (Thesis track), Computer Science — August 2020 - May 2024

- Thesis advisor: Prof. Greg Durrett.
- Thesis work on diverse synthetic dataset generation with retrieval augmentation, published at EMNLP 2024.

### Veermata Jijabai Technological Institute (VJTI) — Mumbai, India

B. Tech in Information Technology, Computer Science — 2013 - 2017

- Research Assistant in Prof. Mahesh Shirole's lab (Anomaly-based Network Intrusion Detection).

## Publications

### External Conferences

Google Scholar: https://scholar.google.com/citations?user=1Sf2oM4AAAAJ

- Abhishek Divekar and Anirban Majumder. *"PRECISE: Reducing the Bias of LLM Evaluations using Prediction-Powered Ranking Estimation"*, Association for the Advancement of Artificial Intelligence **(AAAI 2026 - Innovative Applications, Oral)** [*Acceptance Rate: 19.2%*] (arXiv:2601.18777)
- Abhishek Divekar and Greg Durrett. *"SynthesizRR: Generating Diverse Datasets with Retrieval Augmentation"*, Conference on Empirical Methods in Natural Language Processing **(EMNLP 2024 - Main)** [*Acceptance Rate: 20.8%*] (arXiv:2405.10040)
- Suhas S Kowshik*, Abhishek Divekar*, and Vijit Malik. *"CorrSynth: A Correlated Sampling Method for Diverse Dataset Generation from LLMs"*, Conference on Empirical Methods in Natural Language Processing **(EMNLP 2024 - Main)** [*Acceptance Rate: 20.8%*] (arXiv:2411.08553)
- Abhishek Divekar, Meet Parekh, Vaibhav Savla, Rudra Mishra, and Mahesh Shirole. *"Benchmarking datasets for Anomaly-based Network Intrusion Detection: KDD CUP 99 alternatives"*, IEEE International Conference on Computing, Communication and Security **(IEEE ICCCS 2018, Oral)** (arXiv:1811.05372)

### Amazon Machine Learning Conference (AMLC)

AMLC is Amazon's premier internal AI conference (13th year), featuring ~1000 submissions and keynotes by Turing Award laureates and leading academics (Public Announcement)

- Abhishek Divekar* and Vijit Malik*. *"TrojanShot: One Learned Example Can Red-Team and Guard Large Language Models"*, **(AMLC 2025 - Main, Oral)** [*Acceptance Rate: ~10%*] (Abstract)
- Abhishek Divekar and Greg Durrett. *"SynthesizRR: Generating Diverse Datasets with Retrieval Augmentation"*, **(AMLC 2024 - NLP, Oral)** [*Acceptance Rate: ~10%*] (Abstract)
- Abhishek Divekar et al.. *"Squeezing the last DRiP: AutoML for Cost-constrained Product Classification"*, **(AMLC 2021 - ML & Algorithms, Poster)** [*Acceptance Rate: ~30%*] (Abstract)
- Abhishek Divekar et al.. *"Entity Prediction Service — a configurable, end-to-end AutoML system for Product Classification"*, **(AMLC 2021, Poster)** [*Acceptance Rate: ~30%*] (Abstract)

## Invited Talks & Presentations

### PRECISE: Reducing the Bias of LLM Evaluations using Prediction-Powered Ranking Estimation

AAAI 2026 — Oral presentation, Philadelphia, USA — 2026

### Squeezing the last DRiP: AutoML for Cost-constrained Product Classification

AI/MLSys 2022 — Amazon internal conference — 2022

### Machine Learning at Amazon

Amazon ML Summer School — Invited lecture 2022

### Applied Machine Learning in Industry

Chapman University — Invited talk

### ML in Production

AnalyticsVidhya Datahack — Invited talk

## Teaching & Mentoring

- **Team India AI Olympiad Coach 2025**: Coaching the Indian national team for the International AI Olympiad.
- **Amazon ML Summer School**: Invited lecturer on applied machine learning topics.
- **Mentoring**: Actively mentor junior Applied Scientists at Amazon.
- **Interviewing**: Conducted 175+ interviews for Applied Scientist roles at Amazon.

## Academic Service

- **Area Chair**: Amazon Machine Learning Conference (AMLC) 2025.
- **Reviewer / Program Committee Member (2025)**: ACL, ICLR, AAAI, CoLM, WSDM, SIGIR, WWW.

## Open-Source Contributions

- Led open-sourcing of 3 Python libraries on Amazon Science:
  - **FMCore**: Foundation model training utilities.
  - **Bears**: Batch inference framework.
  - **Concurry**: Async orchestration library.

## Professional Experience

### Senior Applied Scientist (L6)

Amazon — Central Machine Learning December 2021 - Present

- Lead efforts on both research and applied science projects, advising teams across Amazon businesses and developing scientific charters for projects tracked at the VP and SVP level.
- Launched Amazon Rufus India (LLM-powered shopping assistant), used by 100M+ customers. Responsible for red-teaming on India-specific compliance policies.
- Published 3 first-author papers at Tier-1/A* venues (AAAI, EMNLP) and 3 papers at Amazon Machine Learning Conference.

### Applied Scientist 2 (L5)

Amazon — India Machine Learning October 2019 - November 2021

- Designed and implemented data-processing nodes for an internal AutoML platform (160K lines of Python). Platform deployed 1,000+ ML models across 14 Amazon websites.
- BERT-based product classification for 10,000+ categories; +7% accuracy over FastText baseline.
- Batch prediction system for ~500MM products; depth-first preprocessing improved latency by 30-50%.

### Software Development Engineer

Amazon — Kindle & Payments July 2017 - September 2019

- Multi-device purchase flow for Kindle Europe. REST APIs for Tier-1 payments service (Kindle, Alexa, Prime). Reduced API latency by 25%.

### Research Collaborator

United States

NSF-Simons AI Institute for Cosmic Origins (CosmicAI) February 2025 - Present

- Developing novel LLM frameworks for astronomical applications within the “Explorable Universe” project, with researchers at UT Austin, NYU and UCLA.

### Research Assistant

Mumbai, India

Veermata Jijabai Technological Institute (VJTI) April 2017 - June 2017

- First author on full-length paper at IEEE ICCCS-2018 on benchmarking intrusion detection datasets.

### Software Development Intern

Pune, India

Barclays May 2016 - July 2016

- Prototype reducing FX transaction time from T+2 days to T+120 seconds using Ripple blockchain and Node.js.

## Certifications

- Online Learning and Optimization
- Advanced Linear Algebra for Computing
- Reinforcement Learning
- Machine Learning
- Case Studies in Machine Learning

## Honors & Awards

- Winner (out of 300 participants), Amazon Machine Learning Challenge, Chennai 2017.
- Invited talk: “Squeezing the last DRiP: AutoML for Cost-constrained Product Classification”.

## Skills

| | |
|---|---|
| **Core Expertise** | **GenAI / LLMs**, **Applied Machine Learning**, Post-training LLMs, LLM-as-a-Judge, Synthetic Data Generation, Red-Teaming LLMs, AutoML |
| **Programming** | **Python**, Java, C++, Node.js, Docker |
| **Research Areas** | Multimodal Agentic Systems, Retrieval Augmentation, Anomaly Detection, NLP |
| **Languages** | **English**, **Hindi** |