Abhimanyu Hans

College Park, MD · ahans1@umiacs.umd.edu · (929) 832-1886 · https://ahans30.github.io/

Research Interests

My research interests revolve around ensuring the security, efficiency and robustness of generative models. Particularly, I care about memorization and privacy related problems around language models. I'm also keen to "smart" scale LLMs models where we can maximise performance per flop (for eg. recurrence).

Education

University of Maryland – College Park
PhD Computer Science (Advised by **Prof. Tom Goldstein**) *Aug 2022 – May 2026 (Expected)*

University of Delhi
BS Mathematics (Honours) *2014 – 2017*

Publication

- Spotting LLMs With Binoculars: Zero-Shot Detection of Machine-Generated Text *[ICML] [code] [demo]*
  A. Hans, A. Schwarzschild, V. Cherepanova, H. Kazemi, A. Saha, M. Goldblum, J. Geiping, T. Goldstein
- Be like a Goldfish, Don't Memorize! Mitigating Memorization in Generative LLMs *[NeurIPS] [code]*
  A. Hans, Y. Wen, J. Kirchenbauer, H. Kazemi, P. Singhania, S. Singh, G. Somepalli, J. Geiping, A. Bhatele, T. Goldstein

Awards

- Invention of the Year 2024 [**Maryland Innovate**, 2024]

Relevant Work Experience

University of Maryland Institute for Advanced Computer Studies *College Park, MD*
Graduate Research Assistant *May 2023 - Present*

- Working on language models from security and alignment perspective.

PayPal *Remote Work (India)*
Machine Learning Engineer *Oct 2021 - Aug 2022*

- Automated model monitoring framework, saving 70+ FTE weeks, and deploying interactive dashboards for multiple model-specific teams (spanning 100+ model use cases)
- Employed quantitative metrics (AUC, Accuracy, KS-statistic, population stability index) and qualitative analysis (LIME) to enhance model performance and feature interpretability.

HDFC Bank *Mumbai, India*
Machine Learning Engineer *Jan 2020 - Oct 2021*

- Created an anomalous transaction detection system using AutoEncoder and Isolation Forest algorithms, enhancing fraud detection capabilities to achieve 25% recall at 0.1% FPR
- Developed a user-friendly model repository CRUD application using Flask, Jinja Templates, and PostgreSQL, streamlining internal model management processes.

Think360.ai *Mumbai, India*
Associate Data Scientist *Oct 2018 - Jan 2020*

- Developed a deep tree models for predicting repossession values of construction equipment.
- Developed an automated business intelligence pipeline using Tableau, R, and Bash, serving over 5000 FTE with dashboards and email reports.

Skills

- **Languages & Tools:** Python, Java, JavaScript, SQL, Git, Docker, Linux, AWS, Shell
- **Frameworks:** PyTorch, HuggingFace suite, Flask, React, Tableau, Pandas, NumPy, Sklearn, SpringMVC

Patents

Method for Detecting Text Generated by Large Language Models.

Projects (Machine Learning & Software Engineering)

Multi-Armed Bandit Algorithms *(**code**)*
A *NumPy*-based implementation of Greedy, Epsilon-Greedy, and Upper Confidence Bound algorithms along with their empirical performance analysis.

Active Learning *(**code**) (**slides**)*
An approach to ML wherein the learner queries the user with best *instance for a label.*

Model Repository *(**code**)*
Full-stack Flask app serving as a repository for models and their performance metadata.