Abhimanyu Hans

Research Interests

Security & Privacy, Post Training Large Language Models

Education

- 2022 2026 **Ph.D. and Masters**, Computer Sciences, University of Maryland, College Park O Advised by Prof. Tom Goldstein
- 2014 2017 Bachelors (Hons.), Mathematics, University of Delhi

Publications

(* indicates joint first authorship)

Abhimanyu Hans*, Avi Schwarzschild*, Valeriia Cherepanova, Hamid Kazemi, Aniruddha Saha, Micah Goldblum, Jonas Geiping, Tom Goldstein

41st International Conference on Machine Learning [Paper] [Code] [Demo]

NeurIPS 2024 Be like a Goldfish, Don't Memorize! Mitigating Memorization in Generative LLMs

Abhimanyu Hans, Yuxin Wen, Neel Jain, John Kirchenbauer, Hamid Kazemi, Prajwal Singhania, Siddharth Singh, Gowthami Somepalli, Jonas Geiping, Abhinav Bhatele, Tom Goldstein 38th Annual Conference on Neural Information Processing Systems [Paper] [Code]

SC 2024 Democratizing AI: Open-source Scalable LLM Training on GPU-based Supercomputers

Siddharth Singh, Prajwal Singhania, Aditya Ranjan, John Kirchenbauer, Jonas Geiping, Yuxin Wen, Neel Jain, Abhimanyu Hans, Manli Shu, Aditya Tomar, Tom Goldstein, Abhinav Bhatele International Conference for HPC, Networking, Storage and Analysis [Paper]

Preprint Has My System Prompt Been Used? Large Language Model Prompt Membership Inference

Roman Levin*, Valeriia Cherepanova*, <u>Abhimanyu Hans</u>*, Avi Schwarzschild, Tom Goldstein [Paper]

Work Experience

2021 – 2022 Machine Learning Engineer, PayPal India

Built an automated model monitoring framework with interactive dashboards across 100+ model use cases, using metrics (AUC, Accuracy, KS, PSI) and LIME to improve performance and interpretability.

2020 – 2021 Machine Learning Engineer, HDFC Bank

Developed an anomalous transaction detection system (AutoEncoder, Isolation Forest) achieving 25% recall at 0.1% FPR, and built a Flask-based CRUD app with PostgreSQL to streamline model management.

2018 - 2020 Data Scientist, Think360.AI

Built deep tree models to predict construction equipment repossession values and automated a BI pipeline (Tableau, R, Bash) delivering dashboards and reports to 5000+ users.

Selected Awards, Research in News and Services

- o Golfish Loss in Atlantic Magazine [link]
- o Binoculars on IEEE Spectrum [link]
- o Binoculars on Hacker News [link]
- o Binoculars on Business Insider [link]
- o Binoculars on Freethink [link]
- o Invention of the Year for Binoculars Patent in Maryland Innovate 2024 [link]
- o Best Performer Award, IIIT Bangalore 2023
- \circ Reviewer for ICML, ICLR, & NeurIPS