ALEX STEIN

Github: alexstein0 ◇ Google Scholar ◇ Twitter: alex_stein0

(561) · 779 · 7949 ◇ astein0@umd.edu

RESEARCH OVERVIEW

My research interests cover a range of machine learning topics, including but not limited to: LLM reasoning, adversarial attacks, and efficient implementation of transformer architectures. I am especially interested in exploring the extent to which LLMs understand context, reason strategically, and generalize beyond their training data.

EDUCATION

University of Maryland **College Park, MD**
PhD Candidate in Computer Science August 2021 - Present
· Advised by Prof. Tom Goldstein and Prof. John Dickerson

Columbia University **New York City, NY**
BS Computer Science Graduated: May 2016
· Focus: Machine learning
· Minor: Operations Research

EMPLOYMENT EXPERIENCE

Capital One **McLean, VA**
Applied Research PhD Intern, Behavioral Modeling Team June 2024-Sept 2024
· Applied Researcher focused on using transformers to improve performance on temporal tabular data tasks such as event prediction and time series

University of Maryland **College Park, MD**
Research Assistant May 2022-Present
· Research assistant in Tom Goldstein's lab exploring the ways in which LLMs reason in long context settings, as well other generalization tasks

RBC Capital Markets **New York City, NY**
VP, Quantitative Researcher and Development, Global Equities Aug 2016-Dec 2021
· Created alpha signals used to predict market direction
· Trained predictive models used in production systems for scheduling, liquidity seeking, and routing algorithms
· Co-lead rewrite of entire trading system, personally contributing over 10,000 lines of code

TEACHING EXPERIENCE

University of Maryland **College Park, MD**
Teacher's Assistant
· Analysis of Algorithms (Fall 2021)
· Introduction to Machine Learning (Spring 2022)

TECHNICAL EXPERIENCE

Computer Languages	Python, Java, Matlab, LaTeX, Q/kDB+
ML Frameworks	Pytorch, Huggingface, Lightning, Pandas, Numpy
Github	https://github.com/alexstein0

PAPERS AND PREPUBLICATIONS

1. **Alex Stein**, Samuel Sharpe, Doron Bergman, Senthil Kumar, C. Bayan Bruss, John Dickerson, Tom Goldstein, and Micah Goldblum. A simple baseline for predicting events with auto-regressive tabular transformers, 2024. Submitted to The Thirteenth International Conference on Learning Representations (ICLR) (Under Review)

2. Jonas Geiping, **Alex Stein**, Manli Shu, Khalid Saifullah, Yuxin Wen, and Tom Goldstein. Coercing llms to do and reveal (almost) anything, 2024. Submitted to IEEE Secure and Trustworthy Machine Learning (SaTML) (Under Review)

3. Sean McLeish, Arpit Bansal, **Alex Stein**, Neel Jain, John Kirchenbauer, Brian R. Bartoldson, Bhavya Kailkhura, Abhinav Bhatele, Jonas Geiping, Avi Schwarzschild, and Tom Goldstein. Transformers can do arithmetic with the right embeddings. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024

4. Monte Hoover, Ryan Synk, Neel Jain, John Kirchenbauer, **Alex Stein**, Manli Shu, Ramani Duraiswami, and Tom Goldstein. Running huge context windows on tiny GPUs, 2024. Submitted to The Thirteenth International Conference on Learning Representations (ICLR) (Under Review)

5. Avi Schwarzschild, Sean Michael McLeish, Arpit Bansal, Gabriel Diaz, **Alex Stein**, Aakash Chandnani, Aniruddha Saha, Richard Baraniuk, Long Tran-Thanh, Jonas Geiping, and Tom Goldstein. Algorithm design for learned algorithms. 2024

6. Avi Schwarzschild, Arpit Bansal, Sean McLeish, **Alex Stein**, and Tom Goldstein. What algorithms do machines learn to solve mazes? 2023. arXiv pending

7. **Alex Stein**, Avi Schwarzschild, Michael Curry, Tom Goldstein, and John Dickerson. Neural auctions compromise bidder information. 2023

8. **Alex Stein**, Kamil Doruk Gur, and Ian Miers. Deadpool: a deeper, darker, dark pool from horizontally scalable parallel mpc for databases. *arXiv pending*, 2023

9. Gowthami Somepalli, Khalid Saifullah, Hamid Kazemi, **Alex Stein**, Arpit Bansal, David Miller, Micah Goldblum, and Tom Goldstein. Edm: Extracted descriptions of music from generative models. 2023