Daniel Israel

Website: danielmisrael.github.io Cell: (408)-966-7131 Email: disrael@ucla.edu

Education

UCLA, Los Angeles, CA
PhD in Computer Science, September 2022 - Present

- Advisors: Aditya Grover, Guy Van den Broeck
- Computer Science Department Scholar Award
- Master of Science Degree Completed

CALIFORNIA INSTITUTE OF TECHNOLOGY, Pasadena, CA
B.S. in Computer Science, Minor in Information and Data Science

- Advisors: Anima Anandkumar, Pietro Perona, Frederick Eberhardt
- GPA: 4.0

Research Experience

UCLA
Graduate Research Student, September 2022 – Now

- Research in Efficient LLM inference algorithms
- Adaptive Parallel Decoding [1]
  - Accelerating diffusion LLM parallel generation
  - NeurIPS 2025 Spotlight (3.1% Acceptance Rate)
  - First author and sole code contributor
- Prepacking [2]
  - Efficient and exact speedup for LLM prefilling
  - AISTATS 2025
- Planned Diffusion [3]
  - Incorporates semantic parallelism into a hybrid AR/diffusion LLM
  - Under Review at ICLR 2026
- Pitfalls of KV Compression [4]
  - Empirical analysis of KV compression techniques
  - Senior author and undergraduate mentoring role
  - Under Review at ICLR 2026
- MARIA [5]:
  - Autoregressive text infilling made fast and scalable

CALTECH
Research Assistant, March 2021 – June 2022

- Project with Professor Anima Anandkumar implementing model based reinforcement learning to control Cassie bipedal walking

- Worked under Professor Pietro Perona to model the development of number sense in the human brain using computer vision models

Other Experience

UCLA CS DEPARTMENT / CALTECH CMS DEPARTMENT
Teaching Assistant

- TA for CS81: Introduction to Computer Science by Professor Stahl
- TA for CS161: Fundamentals of Artificial Intelligence by Professor Guy Van den Broeck
- TA for CS M146: Introduction to Machine Learning by Professor Aditya Grover
- TA for ACM 116: Introduction to Probability Theory taught by Professor Kostia Zuev
- TA for CS 155: Machine Learning and Data Mining taught by Professor Yisong Yue
- Hosted weekly office hours and helped grade problem sets

SPLUNK, INC.
Intern, June 2020 – August 2020

- Worked with Splunk Security Analytics team and updated architecture to support asynchronous processing using Apache Pulsar
- After conducting performance tests, established 4x improvement in throughput

Skills

- Ability to devise and implement novel algorithms to accelerate LLM inference
- Experience with modern ML frameworks: PyTorch, Tensorflow, Huggingface, Wandb
- Strong understanding of probabilistic inference algorithms in graphical models
- Analytical skills and ability to visualize and present research
- Strong mathematical background in probability theory, linear algebra, and discrete math
- Thorough understanding of modern deep generative model landscape

Selected Works

[1] Israel, Daniel, Guy Van den Broeck, and Aditya Grover. "Accelerating Diffusion LLMs via Adaptive Parallel Decoding." Advances in Neural Information Processing Systems 39 (2025).

[2] Siyan Zhao, Daniel Israel, Guy Van den Broeck and Aditya Grover. Prepacking: A Simple Method for Fast Prefilling and Increased Throughput in Large Language Models, In Proceedings of the 28th International Conference on Artificial Intelligence and Statistics (AISTATS), 2025

[3] Israel, Daniel, Tian Jin, Ellie Cheng, Guy Van den Broeck, Aditya Grover, Suvinay Subramanian, Michael Carbin. "Planned Diffusion." arXiv:2510.18087, 2025.

[4] Chen, Alex, et al. "The Pitfalls of KV Cache Compression." arXiv preprint arXiv:2510.00231 (2025).

[5] Israel, Daniel, Aditya Grover, and Guy Van den Broeck. "Enabling Autoregressive Models to Fill In Masked Tokens." arXiv preprint arXiv:2502.06901 (2025).