ZACHARY ANKNER

ankner@mit.edu 617-939-3564 www.zackankner.com zankner

PUBLICATIONS (* indicates primary author)

Conference / Workshop Works

- **Z. Ankner***, R. Parthasarathy, A. Nrusimha, C. Rinard, J. Ragan-Kelley, and W. Brandon, "Hydra: Sequentially-dependent draft heads for medusa decoding," in *COLM*, 2024.
- **Z. Ankner***, N. Saphra, D. Blalock, J. Frankle, and M. Leavitt, "Dynamic masking rate schedules for MLM pretraining," in *EMNLP*, 2024.
- **Z. Ankner***, A. Renda*, and M. Carbin, "Renamer: A transformer architecture in-variant to variable renaming," in *MLSys Workshop Neurips*, 2023.
- J. R. Shue*, E. R. Chan*, R. Po*, **Z. Ankner***, J. Wu, and G. Wetzstein, "3d neural field generation using triplane diffusion," in *CVPR*, 2023.
- **Z. Ankner***, A. Renda, G. K. Dziugaite, J. Frankle, and T. Jin, "The effect of data dimensionality on neural network prunability," in *ICNB Workshop Neurips*, 2022.

Pre-print / In-review

- **Z. Ankner***, C. Blakeney, K. Sreenivasan, M. Marion, M. L. Leavitt, and M. Paul, "Perplexed by perplexity: Perplexity-based data pruning with small reference models," 2024. arXiv: 2405.20541.
- **Z. Ankner***, M. Paul, B. Cui, J. D. Chang, and P. Ammanabrolu, "Critique-out-loud reward models," 2024. arXiv: 2408.11791.
- R. Parthasarathy, **Z. Ankner***, and A. Gokaslan, "Vid3d: Synthesis of dynamic 3d scenes using 2d video diffusion," 2024. arXiv: 2406.11196.
- W. Brandon, A. Nrusimha, K. Qian, *et al.*, "Striped attention: Faster ring attention for causal transformers," 2023. arXiv: 2311.09431.

Journal Articles

- **Z. Ankner***, P. Balaji, Y. Zhu, C. K. Hiew, P. Wang, and A. Gupta, "Entailsum: An entailment-based approach to aspect-based text summarization with automated aspect adaptation," *IJPRAI*, vol. 36, 2022.

RELEVANT EXPERIENCE

Research Scientist Intern

MosaicML

June 2022 – Ongoing

- Developed the first reward model trained to produce critiques in natural language before predicting the reward. Improved pairwise preference modeling by up to 5.84%.
- Developed neural filtering technique for LLM pre-training based on hard-example-mining that improved 1B parameter model's average downstream performance by 2%.

EDUCATION

Junior

MIT

Sept 2001 – June 2025

GPA: 5.0/5.0
Activities:

- Co-President AI@MIT (2022-2024)
- Co-Lead MIT SIPB Deep Learning Reading Group (2021-current)
- Co-Lead AI@MIT Reading Group (2021-2022, 2024-current)
- Member of MIT MLSys Reading Group (2023-current)
- Member of MIT AI Alignment (2022-current)

Relevant coursework:

- Linear Algebra and Optimization
- Stochastic Processes
- Equivariant Neural Networks
- Quantative Methods for NLP
- Advances in Computer Vision

SKILLS

Python | PyTorch | Tensorflow | Pre-training | Transformers | Spark | Research | LaTeX | Java | Javascript | Node.js | Next.js | React

REFERENCES

Jonathan Frankle
Chief Scientist, MosaicML
jonathan.frankle@databricks.com

Michael Carbin
Professor, MIT
mcarbin@mit.edu

Jonathan Ragan-Kelley
Professor, MIT
jrk@mit.edu

Prithviraj Ammanabrolu
Professor, UCSD
prithvi@ucsd.edu

- Led scaling experiments to profile best transformer architecture on H100s.
- Determined evaluation procedure for data-constrained LLM pre-training and determined optimal data mixture which improved 3B parameter model's average downstream performance by 3.2%. Findings used for pretraining DBRX.
- Re-implemented the DoReMi domain weighting algorithm and evaluated corresponding performance lift.
- Worked on retrieval-based pre-training approaches to improve LLMs on knowledge-intensive tasks. Implemented large-scale pre-training and efficient approximate KNN search.
- Worked on masking rate schedulers for improving masked language model pre-training which led to 1.89x speedup.

Undergraduate Researcher

Programming Systems Group - MIT CSAIL

October 2021 – Ongoing

- Worked on learning neural surrogates of classical programs, specifically learning a transformer-based surrogate of a CPU simulator. Designed and implemented an attention mechanism that makes transformers invariant to semantics preserving variable renamings, setting a new state of the art on the BHIVE dataset.
- Empirically investigated the effect the redundancy in the input distribution being learned has on neural network prunability.

Researcher

Amar Gupta's Lab - MIT CSAIL

August 2020 – August 2021

- Developed entailment module that can integrate with any summarization model to generate zero-shot topic-oriented summaries. Achieved new state-of-the-art performance on the MulitAspect-News dataset.
- Authored research proposal to CSAIL FinTech alliance that was granted.

ML Engineer Intern

Brain Power LLC

June 2019 – August 2019

- Trained and implemented neural networks for facial recognition, facial emotion classification, and body pose estimation.
- Developed a data processing pipeline to retrieve streams of video data from classrooms and apply the aforementioned neural networks.