Zach Nussbaum

773-251-5601 | zanussbaum@gmail.com | https://www.linkedin.com/in/zach-nussbaum/ | https://github.com/zanussbaum

EXPERIENCE

Principal Machine Learning Engineer   March 2023 - Present
Nomic AI   *New York, NY*
- Led Embedding Team; implemented Nomic Embed, powering Atlas embedding space visualization for text and multimodal applications
- Trained and deployed Nomic Embed Text, a state of the art long-context embedding model that outperforms OpenAI and has 15M+ montly downloads
- Trained Nomic Embed Vision, the first multimodal embedding space that performs highly on both multimodal and text embedding benchmarks
- Deployed Nomic Embed models serving as backbone to Atlas, handling billions of tokens monthly

Machine Learning Engineer   January 2022- March 2023
Deep Genomics   *Remote*
- Reimplemented and improved DeepMind gene expression model in Tensorflow by 15-30% on TPUs
- Lead research on representation learning for chemical modifications of RNA molecules using Transformer based architectures.
- Helped standardize and run highly distributed training of attention-based models across many GPUs and TPUs
- Reduce latency of large Deep Learning models by ∼16x by utilizing model distillation and TensorRT
- Ported Maximal Update Parameterization to Tensorflow to allow for zero-shot hyperparameter transfer across model width and depth.

Machine Learning Engineer   September 2020 - December 2021
Amazon   *Seattle, Washington*
- Maintain and improve evaluation pipelines for Alexa Speech Speaker ID model for 25M+ global consumers
- Improved runtime of large PySpark processing jobs by 25% and reduced job failures by 50%
- Drove redesign and engineering for Alexa SpeakerID weekly evaluation pipeline, reducing manual work by 75%
- Optimized AWS EMR clusters for ML Ad Ranking Team, reducing operational costs by 50% leveraging AWS Lambda and AWS Event Bridge

EDUCATION

Davidson College   Davidson, NC
Bachelor of Science in Computer Science, Division 1 Baseball Player   *Aug. 2016 – May 2020*

PUBLICATIONS

1. **Nomic Embed: Training a Reproducible Long Context Text Embedder**. Z. Nussbaum, JX. Morris, B. Duderstadt, A. Mulyar. Arxiv Preprint, 2024

2. **Nomic Embed Vision: Expanding the Latent Space** . Z Nussbaum, B Duderstadt, A Mulyar. Arxiv Preprint, 2024

3. **A Tale of Two Long Tails**. D. D'souza, **Z. Nussbaum**, C. Agarwal, S. Hooker. UDL Workshop ICML, 2021

4. **GPT4All: An Ecosystem of Open Source Compressed Language Models**. Y. Anand, Z. Nussbaum et. al. NLP-OSS EMNLP 2023

5. **DNA-Diffusion: Leveraging Generative Models for Controlling Chromatin Accessiblity And Gene Expression Via Synthetic Regulatory Elements**. S. Senan, A. Janardhan Reddy, Z. Nussbaum, A. Wenteler, M. Bejan, MI. Love, W. Meuleman, L. Pinello. ICLR 2024

6. **An RNA foundation model enables discovery of disease mechanisms and candidate therapeutics** A. Celaj, AJ. Gao, T. Lau, E. Holgersen, A. Lo, V. Lodaya, C. Cole, R. Denroche, C. Spickett, O. Wagih, P. Pinheiro, P. Vora, P. Mohammadi-Shemirani, S. Chan, Z. Nussbaum, X. Zhang, H. Zhu, E. Ramamurthy, B. Kanuparthi, M. Iacocca, D. Ly, K. Kron, M. Verby, K. Cheung-Ong, Z. Shalev, B. Vaz, S. Bhargava, F. Yusuf, S. Samuel, S. Alibai, Z. Baghestani, X. He, K. Krastel, O. Oladapo, A. Mohan, A. Shanavas, M. Bugno, J. Bogojeski, F. Schmitges, C. Kim, S. Grant, R. Jayaraman, T. Masud, A. Deshwar, S. Gandhi, B. Frey