Samuel J. Wiseman

swiseman10@gmail.com | swiseman.github.io | github.com/swiseman

Work Experience

Research Scientist, Apple Foundation Models Team – New York, NY   Oct. 2023 – present
- Scaled a hybrid SSM/transformer model to hundreds of billions of parameters, outperforming an all-transformer baseline. Implemented custom Pallas kernels to speed up pretraining on TPU.
- Main contributor to Apple foundation models' context lengthening training phase: scaled to 131k token context length, tuned training recipe, improved data mixture and evaluation, ran context lengthening in production.
- Improved general pretraining data: investigated n-gram-LM- and semantic deduplication-based approaches to filtering pretraining data.

Assistant Professor, Duke University, Dept. of Computer Science – Durham, NC   July 2021 – Aug. 2023

Research Assistant Professor, TTIC – Chicago, IL   Sept. 2018 – Aug. 2021

Research Intern, Facebook AI Research – New York, NY   Summers 2016 & 2017
- Mentors: Marc'Aurelio Ranzato, Arthur Szlam, Mike Lewis, Sumit Chopra

Education

Harvard University, Ph.D., Computer Science   Sept. 2012 – May 2018
- Advisors: Alexander M. Rush, Stuart M. Shieber

Princeton University, A.B., Philosophy, Magna Cum Laude   Sept. 2006 – June 2010
- Certificate: Program in Applications of Computing

Selected Honors and Awards

- Outstanding Paper Award, ACL 2023
- Siebel Scholar, 2018
- Honorable Mention for Best Paper, EMNLP 2016
- Phi Beta Kappa, Princeton University, 2010

Selected Papers (Google Scholar: 5.7k citations, h-index 21)

AXLearn: Modular Large Model Training on Heterogeneous Infrastructure. Lee et al. July 2025.

Apple Intelligence Foundation Language Models. Gunter et al. July 2024.

Chess as a Testbed for Language Model State Tracking. Toshniwal, Wiseman, Livescu, Gimpel. In *AAAI* 2022.

Data-to-text Generation by Splicing Together Nearest Neighbors. Wiseman, Backurs, Stratos. In *EMNLP* 2021.

Learning Neural Templates for Text Generation. Wiseman, Shieber, Rush. In *EMNLP* 2018.

Challenges in Data-to-Document Generation. Wiseman, Shieber, Rush. In *EMNLP* 2017.

Sequence-to-Sequence Learning as Beam-Search Optimization. Wiseman, Rush. In *EMNLP* 2016.

Selected Service

- Publications Co-Chair, EMNLP 2022
- Area Chair at EMNLP 2021, ACL 2020, EMNLP 2020 (Outstanding AC award)

Teaching

- Duke CS 590.06 (now CS 574): Fundamentals of Deep Learning
- Duke CS 590.03: Neurosymbolic ML
- Duke CS 590.03: Advanced Natural Language Processing (co-taught with Bhuwan Dhingra)
- Duke CS 590.03: Introduction to NLP (co-taught with Bhuwan Dhingra)