

Matthew Finlayson

mattf1n.github.io mfinlays@usc.edu

EDUCATION

- University of Southern California (usc)** 2023–
Viterbi School of Engineering • Ph.D. candidate in computer science • Advised by Swabha Swayamdipta and Xiang Ren.
- Harvard University** 2015–2021
John A. Paulson School of Engineering and Applied Sciences • A.B. Cum Laude in Field/Highest Honors in Computer Science and Linguistics (Joint) • GPA 3.9 out of 4.0. • Advised by Stuart Shieber and Yonatan Belinkov.

EXPERIENCE

- UC Berkeley, Simons Institute for the Theory of Computing** 2025
Special Year on Large Language Models and Transformers.
Visiting student researcher.
- Meta, Generative AI (GenAI)** 2024
Research intern, advised by Aasish Pappu.
- The Allen Institute for AI (AI2), Aristo** 2021–2023
Pre-doctoral researcher advised by Peter Clark and Ashish Sabharwal.
- Microsoft, Natural Language Experiences** 2020
Software engineering intern.

PUBLICATIONS & PREPRINTS

- [1] **Better Language Model Inversion by Compactly Representing Next-Token Distributions**
Murtaza Nazir, Matthew Finlayson, John X. Morris, Xiang Ren, and Swabha Swayamdipta
NeurIPS 2025.
- [2] **Teaching Models to Understand (but not Generate) High-risk Data**
Ryan Wang, Matthew Finlayson, Luca Soldaini, Swabha Swayamdipta, and Robin Jia
COLM 2025.
- [3] **Post-training an LLM for RAG? Train on Self-Generated Demonstrations**
Matthew Finlayson, Ilia Kulikov, Daniel M. Bikel, Barlas Oguz, Xilun Chen, and Aasish Pappu
ArXiv 2025.
- [4] **From Decoding to Meta-Generation: Inference-time Algorithms for Large Language Models**
Sean Welleck, Amanda Bertsch, Matthew Finlayson, Hailey Schoelkopf, Alex Xie, Graham Neubig, Ilia Kulikov, Zaid Harchaoui.
TMLR 2024.
- [5] **Logits of API-Protected LLMs Leak Proprietary Information**
Matthew Finlayson, Xiang Ren, and Swabha Swayamdipta.
COLM 2024 main conference.
- [6] **Closing the Curious Case of Neural Text Degeneration.**
Matthew Finlayson, John Hewitt, Alexander Koller, Swabha Swayamdipta, and Ashish Sabharwal.
ICLR 2024 main conference.

- [7] **Attentiveness to Answer Choices Doesn't Always Entail High QA Accuracy.**
 Sarah Wiegreffe, Matthew Finlayson, Oyvind Tafjord, Peter Clark, and Ashish Sabharwal.
 EMNLP 2023 main conference.
- [8] **Decomposed Prompting: A Modular Approach for Solving Complex Tasks.**
 Tushar Khot, Harsh Trivedi, Matthew Finlayson, Yao Fu, Kyle Richardson, Peter Clark, and Ashish Sabharwal.
 ICLR 2023 main conference.
- [9] **Lila: A Unified Benchmark for Mathematical Reasoning.**
 Matthew Finlayson, Swaroop Mishra, Pan Lu, Leonard Tang, Sean Welleck, Chitta Baral, Tanmay Rajpurohit, Oyvind Tafjord, Ashish Sabharwal, Peter Clark, and Ashwin Kalyan.
 EMNLP 2022 main conference.
- [10] **What Makes Instruction Learning Hard? An Investigation and a New Challenge in a Synthetic Environment.**
 Matthew Finlayson, Kyle Richardson, Ashish Sabharwal, and Peter Clark.
 EMNLP 2022 main conference.
- [11] **Causal Analysis of Syntactic Agreement Mechanisms in Neural Language Models.**
 Matthew Finlayson, Aaron Mueller, Sebastian Gehrmann, Stuart Shieber, Tal Linzen, and Yonatan Belinkov.
 ACL 2021 main conference.

| | | |
|---------------|--|-------|
| HONORS | National Science Foundation | 2023 |
| | Graduate Research Fellowship Program (GRFP) honorable mention. | |
| INVITED TALKS | Meta Fundamental AI Research (FAIR) | 2024 |
| | “The state of (meta-)decoding” | |
| | FAIR & USC Information Sciences Institute (ISI) | 2024 |
| | “How to find ChatGPT’s hidden size, and other low-rank logit tricks” | |
| | Carnegie Mellon University Language Technologies Institute | 2024 |
| | “What top-p sampling has to do with the softmax bottleneck.” | |
| | Instituto Superior Técnico (IST) & Unbabel Seminar | 2023 |
| | “Comprehensively evaluating LMs as general-purpose math reasoners” | |
| | Seminar on Formal Languages and Neural Networks (FLaNN) | 2022 |
| | “What can formal languages tell us about instruction learning?” | |
| | Allen Institute for AI (AI2) | 2022 |
| | “A Unified Benchmark for Mathematical Reasoning” | |
| SERVICE | Conference on Neural Information Processing Systems (NeurIPS) | 2024 |
| | Tutorial co-instructor on decoding algorithms for LLMs. | |
| | Reviewer | 2022– |
| | ARR, ACL, EMNLP, NeurIPS, ICLR, MathNLP, MATH-AI, CONLL, COLM | |
| | Mentor | 2023– |
| | Masters students: Shahzaib Saqib Warraich | |
| | Undergraduates: Jacky Mo, Ryan Wang, Murtaza Nazir | |

TEACHING

USC CSCI-544: Applied Natural Language Processing 2024
Teaching Assistant

Harvard CS-51: Abstraction and Design in Computation 2020–2021
Head Teaching Fellow

Harvard CS-187: Computational Linguistics and NLP 2019–2020
Curriculum developer, Teaching Fellow