

# Alon Albalak

PhD, Computer Science  
University of California, Santa Barbara  
✉ [alonalbalak@gmail.com](mailto:alonalbalak@gmail.com)  
🌐 [Personal Webpage](#)  
🐙 [GitHub](#) [in LinkedIn](#)  
🐦 [Twitter](#) [Scholar](#)



## About Me

I am the Data Team Lead at SynthLabs, where we focus on research for post-training large foundation models. I received my Ph.D from the Computer Science Department at the University of California, Santa Barbara, while I was a member of the NLP Group, co-advised by William Yang Wang and Xifeng Yan.

**The primary research focus of my research has been applying ML methods to NLP to improve data quality and model performance.** In my research I have explored the use of methods including multi-armed bandits, data selection, multitask learning, transfer learning, reinforcement learning, and neuro-symbolic methods. Additionally, I have a wide array of interests in other topics including model efficiency, logic and reasoning, conversational AI, information retrieval, and multilingual models.

**In the future, I am most interested in 2 main directions of work.** First, I would like to continue my pursuit of data-centric research by understanding models from a data perspective, developing methods that improve data quality and data efficiency. Additionally, I am also very excited to apply my data-centric background to many areas of ML research, including model pretraining, alignment, fine-tuning, tool use, and neuro-symbolic systems.

## Education

- 2018–2024 *Ph.D, Computer Science, University of California, Santa Barbara.*  
[UCSB NLP Group](#)  
Dissertation: [Understanding and Improving Language Models Through a Data-Centric Lens](#)  
Advisors: [William Yang Wang](#) and [Xifeng Yan](#)
- 2016–2018 *B.S., Mathematics, Wayne State University.*

## Selected Publications ([Full publication list](#))

- 2025 *Towards System 2 Reasoning in LLMs: Learning How to Think With Meta Chain-of-Thought.*  
Violet Xiang, Charlie Snell, Kanishk Gandhi, [Alon Albalak](#), Anikait Singh, Chase Blagden, Duy Phung, Rafael Rafailov, Nathan Lile, Dakota Mahan, Louis Castricato, Jan-Philipp Franken, Nick Haber, Chelsea Finn. [Preprint](#)
- 2024 *A Survey on Data Selection for Language Models.*  
[Alon Albalak](#), Yanai Elazar, Sang Michael Xie, Shayne Longpre, Nathan Lambert, Xinyi Wang, Niklas Muennighoff, Bairu Hou, Liangming Pan, Haewon Jeong, Colin Raffel, Shiyu Chang, Tatsunori Hashimoto, William Yang Wang  
**TMLR**, Transactions on Machine Learning Research, [Paper](#) [[Github](#)]
- 2024 *Generative Reward Models.*  
Dakota Mahan\*, Duy Van Phung\*, Rafael Rafailov\*, Chase Blagden, Nathan Lile, Louis Castricato, Jan-Philipp Franken, Chelsea Finn, [Alon Albalak\\*](#). [Preprint](#)
- 2024 *DataComp-LM: In search of the next generation of training sets for language models.*  
Jeffrey Li\*, Alex Fang\*, Georgios Smyrnis\*, Maor Ivgi\*, ... [Alon Albalak](#), ..., Achal Dave\*, Ludwig Schmidt\*, Vaishal Shankar\*  
**NeurIPS**, Datasets and Benchmarks Track, [Paper](#) [[Website](#)] [[Code](#)]
- 2024 *Surveying the Effects of Quality, Diversity, and Complexity in Synthetic Data From Large Language Models.*  
Alex Havrilla, Andrew Dai, Laura O'Mahony, Koen Oostermeijer, Vera Zisler, [Alon Albalak](#), ... [Preprint](#)

- 2024 *Generalization vs. Memorization: Tracing Language Models' Capabilities Back to Pretraining Data.*  
 Antonis Antoniadis, Xinyi Wang, Yanai Elazar, Alfonso Amayuelas, [Alon Albalak](#), Kexun Zhang, William Yang Wang  
**NeurIPS**, Workshop on Attributing Model Behavior at Scale, [Paper](#)
- 2024 *The Responsible Foundation Model Development Cheatsheet: A Review of Tools & Resources.*  
 Shayne Longpre, Stella Biderman, [Alon Albalak](#), Gabriel Ilharco, Sayash Kapoor, Kevin Klyman, ...  
**TMLR**, Transactions on Machine Learning Research, [Paper](#) [[Website](#)]
- 2024 *Eagle and Finch: RWKV with Matrix-Valued States and Dynamic Recurrence.*  
 Bo Peng\*, Daniel Goldstein\*, Quentin Anthony\*, [Alon Albalak](#), ...  
**COLM**, Conference on Language Modeling, [Paper](#)
- 2024 *A Mathematical Framework, a Taxonomy of Modeling Paradigms, and a Suite of Learning Techniques for Neural-Symbolic Systems.*  
 Charles Dickens, Connor Pryor, Changyu Gao, [Alon Albalak](#), Eriq Augustine, William Wang, Stephen Wright, Lise Getoor. [Preprint](#)
- 2023 *Improving Few-Shot Generalization by Exploring and Exploiting Auxiliary Data.*  
[Alon Albalak](#), Colin Raffel, William Yang Wang  
**NeurIPS**, Main Conference, [Paper](#) [[code](#)] [[presentation](#)]
- 2023 *Efficient Online Data Mixing For Language Model Pre-Training.*  
[Alon Albalak](#), Liangming Pan, Colin Raffel, William Yang Wang  
**NeurIPS**, Workshop on Robustness of Few-shot and Zero-shot Learning in Foundation Models, [Preprint](#)
- 2023 *RWKV: Reinventing RNNs for the Transformer Era.*  
 Bo Peng\*, Eric Alcaide\*, Quentin Anthony\*, [Alon Albalak](#), ...  
**EMNLP**, Findings, [Paper](#) [[code](#)]
- 2023 *Logic-LM: Empowering Large Language Models with Symbolic Solvers for Faithful Logical Reasoning.*  
 Liangming Pan, [Alon Albalak](#), Xinyi Wang, William Yang Wang  
**EMNLP**, Findings, [Paper](#) [[code](#)]
- 2023 *CausalDialogue: Modeling Utterance-level Causality in Conversations.*  
 Yi-Lin Tuan, [Alon Albalak](#), Wenda Xu, Michael Saxon, Connor Pryor, Lise Getoor, William Yang Wang  
**ACL**, Findings, [Paper](#) [[code](#)]
- 2023 *Addressing Issues of Cross-Linguality in Open-Retrieval Question Answering Systems For Emergent Domains.*  
[Alon Albalak](#), Sharon Levy, William Yang Wang.  
**EACL**, Demonstration Track. [Paper](#) [[code](#)]
- 2023 *NeuPSL: Neural Probabilistic Soft Logic.*  
 Connor Pryor, Charles Dickens, Eriq Augustine, [Alon Albalak](#), William Wang, L. Getoor  
**IJCAI**, Main Conference, [Paper](#) [[code](#)]
- 2022 *FETA: A Benchmark for Few-Sample Task Transfer in Open-Domain Dialogue.*  
[Alon Albalak](#), Yi-Lin Tuan, Pegah Jandaghi, Connor Pryor, Luke Yoffe, Deepak Ramachandran, Lise Getoor, Jay Pujara, William Yang Wang.  
**EMNLP**, Main Conference. [Paper](#) [[code](#)]
- 2022 *Making Something out of Nothing: Building Robust Task-oriented Dialogue Systems from Scratch.*  
 Zekun Li, Hong Wang, [Alon Albalak](#), Yingrui Yang, Jing Qian, Shiyang Li, Xifeng Yan  
**Alexa Prize Taskbot Challenge 2022**. [Paper](#)
- 2022 *D-REX: Dialogue Relation Extraction with Explanations.*  
[Alon Albalak](#), Varun Embar, Yi-Lin Tuan, Lise Getoor, William Yang Wang.  
**ACL**, NLP for Conversational AI Workshop. [Paper](#) [[code](#)]
- 2021 *Systems and methods for determining and using semantic relatedness to classify segments of text.*  
 Rohit Jain, Devin H. Redmond, Richard B. Sutton, [Alon Albalak](#), Sharon Huffner.  
**US Patent 11914963**, [Patent](#)

- 2021 *Modeling Disclosive Transparency in NLP Application Descriptions*.  
Michael Saxon, Sharon Levy, [Alon Albalak](#), Xinyi Wang, William Yang Wang  
**EMNLP**, Main Conference. [Paper](#)

## Professional Experience

- April 2024 – present *Data Team Lead, SynthLabs*.
- Directed the internal research agenda on synthetic data generation, data filtering, and reward models
  - Developed and led open-science collaborations with the broader research community
  - Resulting Publications:** [Generative Reward Models](#), [Towards System 2 Reasoning in LLMs: Learning How to Think With Meta Chain-of-Thought](#)
- June 2022 – September 2022 *Research Science Intern, Meta AI*.
- Directed and executed on 2 projects in collaboration with researchers across the company
  - Explored data-efficiency through the use of multi-task learning and various prompting methods for small language models
  - Explored the use of parameter-efficient methods for zero-shot generalization
  - Resulting Publications:** [Data-Efficiency with a Single GPU](#)
- June 2019 – September 2020 *Research Associate, Theta Lake*.
- Built classifiers for automated risk detection in regulated industries through the use of natural language processing and other machine learning techniques
  - Took multiple projects from inception to production, and developed 2 patent pending methods along the way
  - Resulting Patent:** US Patent 11914963

## Fellowships & Awards

- 2023 *Neurips Scholar Award, 37th Conference on Neural Information Processing Systems*.
- 2018 *Integrative Graduate Education and Research Traineeship (IGERT) Fellow, University of California, Santa Barbara*.
- 2018 *Academic Excellence Fellowship, University of California, Santa Barbara*.
- 2018 *Chia Kuei Tsao Award, Wayne State University*.  
For outstanding academic achievement in the undergraduate mathematics program

## Service & Outreach

- ACL 2023-24 Workshop Organizer - NLP For Conversational AI ([NLP4ConvAI](#))
- ACL 2023 Social Organizer - Mindfulness meditation in a time of NLP hyperactivity
- NeurIPS 2022 Workshop Organizer - Transfer Learning for NLP ([TL4NLP](#)): Insights and Advances on Positive and Negative Transfer. [Proceedings](#).
- 2022-2024 Program Committee: ACL, NAACL, EMNLP, AAAI

## Technical skills

- Tools Python, C++, Shell, AWS, Azure
- Packages PyTorch, TensorFlow, HuggingFace, NumPy, SciPy
- Machine Learning Natural language processing (NLP), computer vision (CV), transformers, generative AI, statistical analysis, regression, clustering

## Military Experience

- 2012 – 2015 *Reconnaissance Sabotage Unit, Israel Defense Forces*.
- Engineering, demolitions, and reconnaissance specialty training
  - Battalion lead navigator