

Yapei Chang

📍 College Park, MD ✉ yapeic@umd.edu 🔗 <https://lilakk.github.io/> 🌐 lilakk

Technical Skills

Python, PyTorch, deep learning, large language models

Work Experience

PhD Research Intern *Allen Institute for AI, Olmo team (Seattle, WA)* *Jun 2025 – Dec 2025*

- Contributed to decisions around special token inclusion during the Olmo 3 midtraining phase.
- Building scalable pipelines to produce synthetic training and evaluation data for procedural knowledge.

Software Developer Engineer Intern *Amazon (Seattle, WA)* *May 2021 – Aug 2021*

Education

University of Maryland, College Park *Ongoing*

PhD in Computer Science

- Advised by [Prof. Mohit Iyyer](#)

University of Massachusetts Amherst

MS in Computer Science

- Advised by [Prof. Mohit Iyyer](#), obtained the MS degree in the MS/PhD program here before transferring

Smith College

BA in Computer Science; Minor in Mathematics

Selected Publications

For a full, most up-to-date list, see my [Google Scholar](#) page.

How2Everything: Mining the Web for How-To Procedures to Evaluate and Improve LLMs Preprint

Yapei Chang, Kyle Lo, Mohit Iyyer, Luca Soldaini

[Paper](#) • [Code](#)

Olmo 3 Technical Report

Olmo Team, including *Yapei Chang*

[Paper](#)

Semantically-Aware Rewards for Open-Ended R1 Training in Free-Form Generation Preprint

Zongxia Li, *Yapei Chang*, Yuhang Zhou, Xiyang Wu, Zichao Liang, Yoo Yeon Sung, Jordan Lee Boyd-Graber

[Paper](#) • [Code](#)

BLEUBERI: BLEU is a surprisingly effective reward for instruction following NeurIPS 2025

Yapei Chang, Yekyung Kim, Michael Krumdick, Amir Zadeh, Chuan Li, Chris Tanner, Mohit Iyyer

[Paper](#) • [Code](#)

CLIPPER: Compression enables long-context synthetic data generation COLM 2025

Chau Minh Pham, *Yapei Chang*, Mohit Iyyer

[Paper](#) • [Code](#)

PostMark: A Robust Blackbox Watermark for Large Language Models EMNLP 2024

Yapei Chang, Kalpesh Krishna, Amir Houmansadr, John Wieting, Mohit Iyyer

[Paper](#) [↗](#) • [Code](#) [↗](#)

FABLES: Evaluating faithfulness and content selection in book-length summarization

COLM 2024

Yekyung Kim, **Yapei Chang**, Marzena Karpinska, Aparna Garimella, , Varun Manjunatha, Kyle Lo, Tanya Goyal, Mohit Iyyer

[Paper](#) [↗](#) • [Code](#) [↗](#)

BooookScore: A Systematic Exploration of Book-length Summarization in the Era of LLMs

ICLR 2024

Oral Presentation
(Top 1.2%)

Yapei Chang, Kyle Lo, Tanya Goyal, Mohit Iyyer

[Paper](#) [↗](#) • [Code](#) [↗](#)

Teaching

PhD Mentor for the UMass Industry Mentor Program (with Microsoft)

Jan 2025 – Present

PhD Mentor for the UMass Industry Mentor Program (with Cohere)

Feb 2024 – May 2024

TA for CS685 Advanced Natural Language Processing

Feb 2023 – May 2023

Community Service

ARR Reviewer

- May 2025, February 2025, October 2024, August 2024, June 2024, a few cycles in 2023

ICLR Reviewer

- 2025, 2026

NeurIPS Reviewer

- 2025