

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

Publications

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](#)

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

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

[Paper](#) • [Code](#)

BEARCUBS: A benchmark for computer-using web agents COLM 2025

Yixiao Song, Katherine Thai, Chau Minh Pham, *Yapei Chang*, Mazin Nadaf, Mohit Iyyer

[Paper](#) • [Website](#)

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, <i>Yapei Chang</i> , 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
<i>Yapei Chang</i> , Kyle Lo, Tanya Goyal, Mohit Iyyer	Oral Presentation (Top 1.2%)
Paper • Code	
RankGen: Improving Text Generation with Large Ranking Models	EMNLP 2022
Kalpesh Krishna, <i>Yapei Chang</i> , John Wieting, Mohit Iyyer	Oral Presentation (Top 21.1%)
Paper • Code	
RELiC: Retrieving Evidence for Literary Claims	ACL 2022
Katherine Thai, <i>Yapei Chang</i> , Kalpesh Krishna, Mohit Iyyer	
Paper • Website	
A Broader Range for 'Meaning the Same Thing': Human Against Machine on Hard Paraphrase Detection Tasks	ACS 2020
Jamie C. Macbeth, <i>Yapei Chang</i> , Jingyu Gin Chen, Yining Hua, Sandra Grandic, Winnie X. Zheng	
Paper	

Teaching

PhD Mentor for the UMass Industry Mentor Program (with Microsoft)	<i>Jan 2025 – Present</i>
PhD Mentor for the UMass Industry Mentor Program (with Cohere)	<i>Feb 2024 – May 2024</i>
TA for CS685 Advanced Natural Language Processing	<i>Feb 2023 – May 2023</i>

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