

KnowFM 2026

**The 4th Workshop on Towards Knowledgeable Foundation
Models**

Proceedings of the Workshop

July 3, 2026

©2026 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

ISBN 979-8-89176-403-3

Preface

Welcome to the 4th Workshop on Towards Knowledgeable Foundation Models (KnowFM 2026), co-located with ACL 2026 and held in San Diego, California, United States, on July 3, 2026.

Foundation models have become vast repositories of world and task knowledge, yet how that knowledge is acquired, stored, retrieved, and faithfully used remains only partially understood. KnowFM is organized around the lifecycle of knowledge in foundation models, spanning four stages: how knowledge *emerges* during pre-training, how it is *injected* from external sources, how it is *updated* or edited over time, and how it is *probed and generated* at inference. The workshop brings together research on the analysis of knowledge in foundation models (how much they know, where that knowledge comes from, and how it is represented), knowledge injection during training and inference, retrieval-augmented generation, knowledge updating and editing, knowledge extraction and distillation, hallucination mitigation and the evaluation of faithfulness, knowledge conflicts between parametric memory and retrieved or long-context information, and multimodal and agentic knowledge systems.

This fourth edition of the workshop accepted 38 papers. The present volume collects the 14 accepted papers whose authors elected for archival inclusion in the ACL Anthology proceedings; the remaining accepted papers were presented at the workshop on a non-archival basis. We are also delighted to host invited talks by Sewon Min (University of California, Berkeley), Eunsol Choi (New York University), Mohit Iyyer (University of Maryland), Danqi Chen (Princeton University), Yulia Tsvetkov (University of Washington), and Luna Dong (Meta).

We are grateful to the 72 members of our Program Committee, whose careful and timely reviews made the program possible, and to our Advising Committee, Heng Ji, Mohit Bansal, Isabelle Augenstein, and Jiawei Han, for their guidance. We thank the authors for their contributions, the invited speakers for sharing their perspectives, and the ACL 2026 organizers and publication chairs for their support. We hope you find the workshop stimulating.

The KnowFM 2026 Organizing Committee

Organizing Committee

Organizing Committee

Canyu Chen, Northwestern University
Yuji Zhang, University of Illinois Urbana-Champaign
Zoey Sha Li, Amazon
Zihan Wang, Northwestern University
Qineng Wang, Northwestern University
Jinyan Su, Cornell University
Priyanka Kargupta, University of Illinois Urbana-Champaign
Sara Vera Marjanović, University of Copenhagen
Jeff Z. Pan, University of Edinburgh
Manling Li, Northwestern University

Publication Chair

Canyu Chen, Northwestern University

Advising Committee

Heng Ji, University of Illinois Urbana-Champaign
Mohit Bansal, University of North Carolina at Chapel Hill
Isabelle Augenstein, University of Copenhagen
Jiawei Han, University of Illinois Urbana-Champaign

Program Committee

Program Committee

Shivam Adarsh, Copenhagen University
Atul Anand, curiouskalp
Ashwin Aravind, Indian Institute of Technology, Bombay
Nikolay Banar, Universiteit Antwerpen
Prakhar Bansal, OpenAI
Mina Basirat, University of Central Florida
Aprameya Bharadwaj, Adobe Systems
Zacharie Bugaud, Astera Institute
Yisu Chen, Wilfrid Laurier University
Gaurav Chodwadia, Walmart
Lu Dai, Hong Kong University of Science and Technology
Xiaoyu Deng, University of Pennsylvania
Pranamy Nilesh Deshpande, Gokhale Education Society's R. H. Sapat College of Engineering,
Management Studies and Research
Sahil Rajesh Dhayalkar, Brain Corporation
Liang Ding, Alibaba Group
Yixuan Du, Georgetown University
Ritvik Garimella, University of South Carolina
Dipanwita Guhathakurta, International Business Machines
Akshat Gupta, University of California, Berkeley
Joanna Hao, University of Alberta
Mohd Ariful Haque, Clark Atlanta University
Tianyi Huang, Cleanlab
Shabbir Hussain, Amazon
Parin Rajesh Jhaveri, J.P. Morgan Chase
Adam Kamel, University of Waterloo
Gael Kengmegni, Independent
Yoav Kor, Tel Aviv University
Ashish Balkishan Lathkar, Hemut Co
Qi Li, Tsinghua University
Zichao Li, University of Waterloo
Duoduo Liao, George Mason University
Shiyang Liu, Huazhong University of Science and Technology
Devarshi Mahajan, Northeastern University
Zhaohan Meng, University of Glasgow
Noor Islam S. Mohammad, Istanbul Technical University
Yi Nian, University of Southern California
Michael-Andrei Panaitescu-Liess, University of Maryland, College Park
Rahul Pandya, DePaul University
Hari Charan Panjwani, Facebook
Maitrik Patel, Apple
Pankayaraj Pathmanathan, University of Maryland, College Park
Viswanathan Ranganathan, IEEE
Ravi Ranjan, Florida International University
Shivam Ratnakar, University of Southern California
Pravish Sainath, Montreal Institute for Learning Algorithms, University of Montreal, University of

Montreal
Kanchan Sarkar, Tiktok
Krishna Senthilkumar, Algoverse AI Research
Eugenie Shi, Stanford University
Jiazhao Shi, Amazon
Shuzheng Si, Tsinghua University
Malaiappan Srinivasan, University of Zurich
Elena Stringli, National Technical University of Athens
Hexiang Tan, Chinese Academy of Sciences
Wout Teillers, University of Groningen
Rohith Uppala, LinkedIn
Matias Valdenegro-Toro, University of Groningen
Tomás Vergara Browne, McGill University
Yike Wang, Department of Computer Science, University of Washington
Zhaohui Geoffrey Wang, University of Southern California
Juesi Xiao, Tianjin University
Zhichao Xu, Amazon
Shuo Yang, University of Melbourne
Yuqing Yang, University of Southern California
Haeun Yu, University of Copenhagen
Wenhao Yuan, The University of Hong Kong
Wajdi Zaghouani, Northwestern University
Pardis Sadat Zahraei, University of Illinois at Urbana-Champaign
Fengwen ZHAI, Lanzhou Jiaotong University
Wenjing Zhai, The High School Affiliated to Beijing Normal University
Weizhi Zhang, University of Illinois at Chicago
Jianing Zhao, Tsinghua University
Yuanjie Zhu, University of Illinois at Chicago

Table of Contents

<i>Annotation Frameworks Shape Model Knowledge: Safety Alignment in Large Language Models</i> Wajdi Zaghouani	1
<i>Blind Single-Layer Activation Edits Show a Break/Fix Asymmetry in Factual Recall</i> Zacharie Bugaud	13
<i>What Does Alignment Cost? The Structural Brittleness of Chain-of-Thought Reasoning</i> Joanna Hao, Shanduojiang Jiang and Sai Asish Nakka	25
<i>bLLeQA: Benchmarking LLMs for Grounded Legal Question-Answering in French and Dutch</i> Nikolay Banar, Ehsan Lotfi, Jens Van Nooten, Marija Kliocaitė and Walter Daelemans	34
<i>VLA-Forget: Vision-Language-Action Unlearning for Embodied Foundation Models</i> Ravi Ranjan and Agoritsa Polyzou	60
<i>Overcoming the Impedance Mismatch: A Theoretical Roadmap for Fusing Foundation Models and Knowledge Graphs</i> Sahil Rajesh Dhayalkar	78
<i>LLM-MemCluster: Empowering Large Language Models with Dynamic Memory for Text Clustering</i> Yuanjie Zhu, Liangwei Yang, Ke Xu, Weizhi Zhang, Zihong Song, Jindong Wang and Philip S. Yu	90
<i>Test-Time Training for Zero-Resource Dense Retrieval Reranking</i> Shiyuan Liu and Yichen Li	105
<i>Multimodal Generative Engine Optimization: Rank Manipulation for Vision–Language Model Rankers</i> Yixuan Du, Chenxiao Yu, Haoyan Xu, Ziyi Wang, Yue Zhao and Xiyang Hu	115
<i>Beyond Retrieval: Bi-Temporal State Arbitration for Longitudinal Healthcare Agents</i> Jianing Zhao, Xiaoquan Zhi and Xinqiang Yu	129
<i>RSCE: Training-Free Residual Stream Encoding for Persistent Context Amortization</i> Adam Kamel and Eric Xu	138
<i>Tricking Open-World Object Recognition Models: Uncertainty in Out-of-Distribution Detection</i> Wout Teillers and Matias Valdenegro-Toro	147
<i>Knowledge Localization and Editability in Small Language Models: A Multi-Stage Experimental Study</i> Pranamy Niles Deshpande, Aiswarya Konavoor and Sreedath Panat	165
<i>One Retrieval to Cover Them All: Co-occurrence-Aware Knowledge Base Reorganization for Session-Level RAG</i> Shivam Ratnakar, Yixuan Zhu, Cecilia Cheng and Chaya Vijayakumar	173