

L2M2 2025

**The First Workshop on Large Language Model  
Memorization (L2M2)**

**Proceedings of the Workshop**

August 1, 2025

The L2M2 organizers gratefully acknowledge the support from the following sponsors.

**In collaboration with**



©2025 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](mailto:acl@aclweb.org)

ISBN 979-8-89176-278-7

## Introduction

The First Workshop on Large Language Model Memorization (L2M2), co-located with ACL 2025 in Vienna, brings together researchers studying the phenomenon of memorization in large language models from multiple perspectives.

Large language models (LLMs) are known to memorize their training data, and this phenomenon has inspired multiple distinct research directions. Some researchers focus on understanding LLM memorization, attempting to localize memorized knowledge or identify which examples are most likely to be memorized. Others aim to edit or remove information that an LLM has memorized. Still others study the downstream implications of LLM memorization, including legal concerns associated with memorizing copyrighted articles, privacy risks associated with LLMs leaking private information, and benchmarking concerns that LLMs are memorizing test data.

This workshop seeks to provide a central venue for researchers studying LLM memorization from these different angles, fostering collaboration and advancing our understanding of this important phenomenon.

# Organizing Committee

## **Workshop Chair**

Robin Jia, University of Southern California

## **Workshop Chair**

Eric Wallace, OpenAI and University of California Berkeley

## **Workshop Chair**

Yangsibo Huang, Google

## **Workshop Chair**

Tiago Pimentel, ETH Zürich

## **Workshop Chair**

Pratyush Maini, Carnegie Mellon University

## **Workshop Chair**

Verna Dankers, University of Edinburgh

## **Workshop Chair**

Johnny Wei, University of Southern California

## **Workshop Chair**

Pietro Lesci, University of Cambridge

# Program Committee

## Program Chairs

Verna Dankers, University of Edinburgh  
Yangsibo Huang, Google  
Robin Jia, University of Southern California  
Pietro Lesci, University of Cambridge  
Pratyush Maini, Carnegie Mellon University  
Tiago Pimentel, ETH Zürich  
Johnny Wei, University of Southern California

## Area Chairs

Verna Dankers, University of Edinburgh  
Yangsibo Huang, Google  
Robin Jia, University of Southern California  
Pietro Lesci, University of Cambridge  
Pratyush Maini, Carnegie Mellon University  
Tiago Pimentel, ETH Zürich  
Johnny Wei, University of Southern California

## Reviewers

Aryaman Arora, Stanford University  
Marco Bombieri, University of Verona  
Peter Carragher, Carnegie Mellon University  
Ting-Yun Chang, University of Southern California  
Bowen Chen, University of Tokyo  
Matthew Finlayson, University of Southern California  
James Flemings, University of Southern California  
Ameya Godbole, University of Southern California  
Patrick Haller, Humboldt Universität Berlin  
Skyler Hallinan, University of Southern California  
Kanyao Han, Walmart Global Tech and University of Illinois at Urbana-Champaign  
Yuzheng Hu, University of Illinois at Urbana-Champaign  
Jing Huang, Stanford University  
Shotaro Ishihara, Nikkei Inc.  
Masaru Isonuma, National Institute of Informatics and Tohoku University  
Matthew Jagielski, Google  
Dongjun Jang, Seoul National University  
Mohammad Aflah Khan  
Cristina Lopes, University of California, Irvine  
Lucie Charlotte Magister, University of Cambridge  
Matthieu Meeus, Imperial College London  
Marius Mosbach, McGill University and Mila - Quebec Artificial Intelligence Institute  
Max Ploner, Humboldt Universität Berlin  
Xiangyu Qi, Princeton University  
Nishat Raihan, George Mason University

Leonardo Ranaldi, University of Edinburgh  
Vikas Raunak, Google DeepMind  
Suchir Salhan, University of Cambridge  
Avi Schwarzschild, Carnegie Mellon University  
Igor Shilov, Imperial College London  
Anshuman Suri, Northeastern University  
Anvith Thudi, University of Toronto  
Martin Tutek, University of Zagreb  
Juraj Vladika, Technische Universität München  
Ryan Yixiang Wang, University of Southern California  
Boyi Wei, Princeton University  
Johnny Wei, University of Southern California  
Fan Wu, University of Illinois  
Chiyuan Zhang, Google  
Bihe Zhao, CISPA Helmholtz Center for Information Security

## Table of Contents

<i>Factual Knowledge in Language Models: Robustness and Anomalies under Simple Temporal Context Variations</i>	
Hichem Ammar Khodja, Frederic Bechet, Quentin Brabant, Alexis Nasr and Gwéno�� Lecorv��	1
<i>Memorization in Language Models through the Lens of Intrinsic Dimension</i>	
Stefan Arnold	23
<i>From Data to Knowledge: Evaluating How Efficiently Language Models Learn Facts</i>	
Daniel Christoph, Max Ploner, Patrick Haller and Alan Akbik	29
<i>Towards a Principled Evaluation of Knowledge Editors</i>	
Sebastian Pohl, Max Ploner and Alan Akbik	47
<i>On the Way to LLM Personalization: Learning to Remember User Conversations</i>	
Lucie Charlotte Magister, Katherine Metcalf, Yizhe Zhang and Maartje Ter Hoeve	61
<i>From Teacher to Student: Tracking Memorization Through Model Distillation</i>	
Simardeep Singh	78
<i>Understanding Verbatim Memorization in LLMs Through Circuit Discovery</i>	
Ilya Lasy, Peter Knees and Stefan Woltran	83
<i>Quantifying Memorization in Continual Pre-training with Japanese General or Industry-Specific Corpora</i>	
Hiromu Takahashi and Shotaro Ishihara	95
<i>Memorization is Language-Sensitive: Analyzing Memorization and Inference Risks of LLMs in a Multilingual Setting</i>	
Ali Satvaty, Anna Visman, Dan Seidel, Suzan Verberne and Fatih Turkmen	106
<i>Quantifying Memorization and Parametric Response Rates in Retrieval-Augmented Vision-Language Models</i>	
Peter Carragher, Abhinand Jha, Raghav R and Kathleen M. Carley	127
<i>Empirical Evaluation of Loss Masking to Selectively Prevent Memorization</i>	
Tagore Rao Kosireddy and Evan Lucas	142
<i>Bring Your Own Knowledge: A Survey of Methods for LLM Knowledge Expansion</i>	
Mingyang Wang, Alisa Stoll, Lukas Lange, Heike Adel, Hinrich Schuetze and Jannik Str��tgen	150
<i>Memorization: A Close Look at Books</i>	
Iris Ma, Ian Domingo, Alberto Krone-Martins, Pierre Baldi and Cristina Lopes	169
<i>Memory Tokens: Large Language Models Can Generate Reversible Sentence Embeddings</i>	
Ignacio Sastre and Aiala Ros��	183
<i>Robust Data Watermarking in Language Models by Injecting Fictitious Knowledge</i>	
Xinyue Cui, Johnny Wei, Swabha Swayamdipta and Robin Jia	190
<i>Better Aligned with Survey Respondents or Training Data? Unveiling Political Leanings of LLMs on U.S. Supreme Court Cases</i>	
Shanshan Xu, Santosh T.y.s.s, Yanai Elazar, Quirin Vogel, Barbara Plank and Matthias Grabmair	205
<i>Capacity Matters: a Proof-of-Concept for Transformer Memorization on Real-World Data</i>	
Anton Changelidis and Aki H��rm��	227