

CustomNLP4U 2026

**The Second Workshop on Customizable NLP: Progress and
Challenges in Customizing NLP for a Domain, Application,
Group, or Individual**

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 979-8-89176-396-8

Introduction

Welcome to the Second Workshop on Customizable NLP (CustomNLP4U), co-located with the 64th Annual Meeting of the Association for Computational Linguistics (ACL 2026) in San Diego, California. CustomNLP4U brings together researchers and practitioners exploring how language models can be adapted to the specific needs of individual users, groups or organizations.

Most language modeling research today focuses on building generalist models through pretraining at scale and reinforcement learning-based post-training. The increased capabilities of LLMs have promised greater productivity and innovation, seeing broad consumption for personal and commercial use. However, users' expectations, values, and workflows vary significantly across domains, organizations, cultures, and individuals — factors that are often under-considered in existing pipelines. Generalist models have non-uniform performance, particularly for consumers in sensitive and specialized domains such as law, finance, or health, and for individuals or cultures less represented online. For language models to deliver on their promise, there is a need to develop models that can be tailored to different consumers, easily controlled by them, and able to reason about users' private knowledge and context to provide personalized responses. Alongside these methodological questions, model customization raises ethical and security concerns around copyrighted data, user privacy, and bias — especially in sensitive domains where customization can bring large benefits but also higher risks. To discuss the progress and challenges on these interdisciplinary questions, CustomNLP brings together researchers from diverse fields spanning Natural Language Processing (NLP), Machine Learning (ML), Human-Computer Interaction (HCI), and Fairness and AI Policy.

We received 47 submissions and accepted 22 papers for presentation at the workshop — an acceptance rate of 46.8%. Of these, 17 are archival contributions and appear in this volume; the remaining 5 are non-archival and were presented at the workshop only. Every paper received careful evaluation through a double-blind review process coordinated through OpenReview, with each submission seen by multiple members of our program committee.

We thank the authors for entrusting their work to CustomNLP4U, and we are deeply grateful to our program committee and area chairs for their thoughtful reviews under a tight schedule. We thank the members of our steering committee — Hamed Zamani, Dongyeop Kang, and Yulia Tsvetkov — for their guidance throughout the planning of the workshop. We also thank the ACL 2026 workshop chairs and publication chairs for their support, and the broader *CL community for continuing to make space for focused workshops alongside the main conference.

We hope the program sparks new collaborations and ideas, and that you enjoy the workshop.

The CustomNLP4U 2026 Organizers

Sheshera Mysore, Sachin Kumar, Vidhisha Balachandran, Shirley Anugrah Hayati, Faeze Brahman, Hanane Nour Moussa, and Alireza Salemi

Organizing Committee

Organizers

Sheshera Mysore, Microsoft Office of Applied Research, USA
Sachin Kumar, The Ohio State University and Allen Institute for AI, USA
Vidhisha Balachandran, Microsoft Research, USA
Shirley Anugrah Hayati, University of Minnesota, USA
Faeze Brahman, Allen Institute for AI, USA
Hanane Nour Moussa, The Ohio State University, USA
Alireza Salemi, University of Massachusetts Amherst, USA

Steering Committee

Hamed Zamani, University of Massachusetts Amherst, USA
Dongyeop Kang, University of Minnesota, USA
Yulia Tsvetkov, University of Washington, USA

Program Committee

Reviewers

Shardul Aggarwal, Arnav Arora, Anirudh Atmakuru, Katherine Atwell

Xiaoyan Bai, Ahsaas Bajaj, Nishant Balepur

Isabel Cachola, Yapei Chang, Daiwei Chen, Harshita Chopra

Debarati Das, Jane Arleth dela Cruz

Cristina Garbacea, Sayan Ghosh, Anmol Goel, Taicheng Guo, Yaoqi Guo

Jawad Hossain, Tiancheng Hu

Brihi Joshi

To Eun Kim, Yekyung Kim, Kemal Kirtac

Chen Cecilia Liu, Haokun Liu, Yuhan Liu

Stephen Meisenbacher

Krithika Ramesh, Jenna Russell

Amrita Singh

Yu-Min Tseng

Anvesh Rao Vijjini

Manya Wadhwa, Jiajing Wan

Wentao Zhang, Wangjie Zheng, Karen Zhou, Shengqi Zhu

Table of Contents

<i>BAID: A Benchmark for Bias Assessment of AI Detectors</i> Priyam Basu, Yunfeng Zhang and Vipul Raheja	1
<i>Small Language Models for the Democratization of Financial Literacy: Challenges and Opportunities</i> Tagore Rao Kosireddy, Jeffrey David Wall and Evan Lucas	11
<i>From Understanding to Engagement: Personalized pharmacy Video Clips via Vision Language Models (VLMs)</i> Suyash Mishra, Qiang Li, Anubhav Girdhar and Srikanth Patil	24
<i>Evaluating Customized vs. Generalist Transformer-based Models for Legal Contract Classification</i> Amrita Singh, H. Suhan Karaca, Aditya Joshi, Hye-young Paik and Jiaojiao Jiang	44
<i>Personalizing News Headlines with Retrieval-Augmented Generation</i> Jiajing Wan, Samia Touileb, Lubos Steskal and Lilja Øvrelid	55
<i>Building Multi-turn Intent Classification with LLM-based Labeling</i> Biancen Xie, Kaiqi Bian, Jai Ranjan Singh Gusain, Manikandarajan Ramanathan and Raj Maragoud	68
<i>Cross-Tokenizer LLM Distillation through a Byte-Level Interface</i> Avyav Kumar Singh, Yen-Chen Wu, Alexandru Cioba, Alberto Bernacchia and Davide Buffelli	84
<i>Fine-grained Readability Controlled Summarization of Scientific Documents via Control Vectors</i> Isabel Cachola, Kuleen Sasse and Mark Dredze	97
<i>Building a Custom Taxonomy of AI Skills and Tasks from the Ground Up with Job Postings</i> Stephen Meisenbacher and Peter Norlander	117
<i>Using Topological Data Analysis to Characterize the Layers of Language Models Before and After Word Substitution Attacks</i> Adam Tang, Catherine Liu, Kimberly Lopez, Shreya Subramanian, Leif Zinn-Brooks, Alexia E. Schulz and Adaku Uchendu	131
<i>Customizing ASR for Language Documentation and Resource Prioritization</i> Alexandra Fort and Shobhana Lakshmi Chelliah	149
<i>Improving Medical Hallucination Detection with System Combination and Rule-based Customization</i> Jonathan Lasko, Damianos Karakos and Francis Keith	160
<i>Asking the Right Questions: Can expert-prompted LLMs reformulate legal queries from non-experts?</i> Katherine Atwell, Morgan A. Gray, Jaromir Savelka, Len Rial, Sera Linardi and Malihe Alikhani	167
<i>When Valid Signals Fail: Regime Boundaries Between LLM Features and RL Trading Policies</i> Zhengzhe Yang	182
<i>Unintended Effects of Geographic Conditioning in Large Language Models</i> Naz Col and David M. Chan	191
<i>Efficiency vs. Verifiability in Evidence-Aware RAG: Does Prompt Compression Preserve Citation Grounding?</i> Aiyu Li, Qian Peng and Bin Chen	202
<i>When Gradients Collide: Failure Modes of Multi-Objective Prompt Optimization for LLM Judges</i> Parth Darshan and Abhishek Divekar	216

Program

Friday, July 3, 2026

09:00 - 09:15 *Opening Remarks*

09:15 - 10:45 *Poster Session*

BAID: A Benchmark for Bias Assessment of AI Detectors

Priyam Basu, Yunfeng Zhang and Vipul Raheja

Small Language Models for the Democratization of Financial Literacy: Challenges and Opportunities

Tagore Rao Kosireddy, Jeffrey David Wall and Evan Lucas

From Understanding to Engagement: Personalized pharmacy Video Clips via Vision Language Models (VLMs)

Suyash Mishra, Qiang Li, Anubhav Girdhar and Srikanth Patil

A Versatile Multi-Modal Agent for Rare Disease Diagnosis and Risk Gene Prioritization

Tianyu Liu, Wangjie Zheng, Weihao Xuan, Rui Yang, Botao Yu, Kexin Huang, Nan Liu and Hongyu Zhao

Evaluating Customized vs. Generalist Transformer-based Models for Legal Contract Classification

Amrita Singh, H. Suhan Karaca, Aditya Joshi, Hye-young Paik and Jiaojiao Jiang

Personalizing News Headlines with Retrieval-Augmented Generation

Jiajing Wan, Samia Touileb, Lubos Steskal and Lilja Øvreliid

FD-RAG: Federated Dual-System Retrieval-Augmented Generation

Tianhao Gao, Kai Yang and Yiyang Li

Building Multi-turn Intent Classification with LLM-based Labeling

Biancen Xie, Kaiqi Bian, Jai Ranjan Singh Gusain, Manikandarajan Ramanathan and Raj Maragoud

Cross-Tokenizer LLM Distillation through a Byte-Level Interface

Avyav Kumar Singh, Yen-Chen Wu, Alexandru Cioba, Alberto Bernacchia and Davide Buffelli

Fine-grained Readability Controlled Summarization of Scientific Documents via Control Vectors

Isabel Cachola, Kuleen Sasse and Mark Dredze

Friday, July 3, 2026 (continued)

Building a Custom Taxonomy of AI Skills and Tasks from the Ground Up with Job Postings

Stephen Meisenbacher and Peter Norlander

10:45 - 11:00 *Coffee Break*

11:00 - 12:15 *Poster Session (cont.)*

Using Topological Data Analysis to Characterize the Layers of Language Models Before and After Word Substitution Attacks

Adam Tang, Catherine Liu, Kimberly Lopez, Shreya Subramanian, Leif Zinn-Brooks, Alexia E. Schulz and Adaku Uchendu

Customizing ASR for Language Documentation and Resource Prioritization

Alexandra Fort and Shobhana Lakshmi Chelliah

Improving Medical Hallucination Detection with System Combination and Rule-based Customization

Jonathan Lasko, Damianos Karakos and Francis Keith

Asking the Right Questions: Can expert-prompted LLMs reformulate legal queries from non-experts?

Katherine Atwell, Morgan A. Gray, Jaromir Savelka, Len Rial, Sera Linardi and Malihe Alikhani

Hint-Assisted Reasoning: Improving Mathematical Problem Solving in Small Language Models

Jawad Hossain, Xiangyu Guo, Jiawei Zhou and Chong Liu

When Valid Signals Fail: Regime Boundaries Between LLM Features and RL Trading Policies

Zhengzhe Yang

Unintended Effects of Geographic Conditioning in Large Language Models

Naz Col and David M. Chan

Efficiency vs. Verifiability in Evidence-Aware RAG: Does Prompt Compression Preserve Citation Grounding?

Aiyu Li, Qian Peng and Bin Chen

Who's Asking? Simulating Role-Based Questions for Conversational AI Evaluation

Navreet Kaur, Hoda Ayad, Hayoung Jung, Shravika Mittal, Munmun De Choudhury and Tanu Mitra

Friday, July 3, 2026 (continued)

When Gradients Collide: Failure Modes of Multi-Objective Prompt Optimization for LLM Judges

Parth Darshan and Abhishek Divekar

Low-resource Authorship Style Transfer via Dynamic Style Inference and Parameter Modulation

Jongkyung Shin, Minguk Jeon, Chan Woo Park and Chiehyeon Lim

12:15 - 12:30 *Closing Remarks*