HuCLLM 2024

The First Human-Centered Large Language Modeling Workshop

Proceedings of the Workshop

August 15, 2024

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 979-8-89176-152-0

Introduction

A word's meaning resides in the heart and soul of its "generator" - people. How do we include human (personal, social, cultural, situational) context, ethically, into LLMs – the base models of our NLP systems?

Language modeling in the context of its source [author] and target [audience] can enable NLP systems to better understand human language. Advances in *human-centered NLP* have established the importance of modeling the human context holistically, including personal, social, cultural, and situational factors in NLP systems. Yet, our NLP systems have become heavily reliant on large language models that do not capture the human context.

Human language is highly dependent on the rich and complex human context such as (a) *who* is speaking, (b) to *whom*, (c) *where* (situation/environment) and (d) *when* (time and place). It is additionally moderated by the changing human states of being such as their mental and emotional states.

Current large language models can possibly simulate some form of human context given their large scale of parameters and pre-training data. However, they do not explicitly process language in the higher order structure of language – connecting documents to people, the "source" of the language.

Prior work has demonstrated the benefits of including the author's information using LLMs for downstream NLP tasks. Recent research has also shown that LLMs can benefit from including additional author context in the LM pre-training task itself. Progress in the direction of merging the two successful parallels, i.e., human-centered NLP and LLMs, drives us toward creating a vision of human-centered LLMs for the future of NLP in the era of LLMs.

Human-centered large language modeling has the potential to bring promising improvements in human-centric applications through multiple domains such as healthcare, education, consumerism, etc. Simultaneously, this new research focus also brings multitudes of unexplored architectural, data, technical, fairness, and ethical challenges. With our first edition of the Human-Centered Large Language Modeling (HuCLLM) workshop, we aim to create a platform where researchers can present rising challenges and solutions in building human-centered NLP models that bring together the ideas of human and social factors adaptation into the base LLMs of our NLP systems.

We received 35 submissions, of which 18 were accepted for presentation at the workshop. These papers will be presented at oral and poster sessions on the day of the workshop. The workshop day will also include keynote talks and a panel session on human-centered large language modeling. We thank all our participants and reviewers for their work. We hope you enjoy the first edition of HuCLLM and the research published in these proceedings.

Nikita Soni, Lucie Flek, Ashish Sharma, Diyi Yang, Sara Hooker, H Andrew Schwartz

HuCLLM 2024 Chairs

Acknowledgements This workshop is supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via the HIATUS Program contract 2022-22072200005, and LAMARR Institute for Machine Learning and Artificial Intelligence. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

Organizing Committee

Workshop Organizers

Nikita Soni, Stony Brook University
Lucie Flek, University of Bonn
Ashish Sharma, University of Washington
Diyi Yang, Stanford University
Sara Hooker, Cohere For AI
H. Andrew Schwartz, Stony Brook University

Program Committee

Emergency Reviewers

Dana Moukheiber, Massachusetts Institute of Technology

Program Committee

Gavin Abercrombie, Heriot-Watt University
Maria Antoniak, Allen Institute for AI
Shaina Ashraf, Math and CS, Marburg University and Rheinische Friedrich-Wilhelms Universität Bonn
Giuseppe Attanasio, Instituto de Telecomunicações
Shaily Bhatt, Carnegie Mellon University
Siva Uday Sampreeth Chebolu, University of Houston
Wei-Fan Chen, Rheinische Friedrich-Wilhelms Universität Bonn
Jaemin Cho, University of North Carolina, Chapel Hill
Elizabeth Clark, Google
Amanda Cercas Curry, Bocconi University
Athiya Deviyani, School of Computer Science, Carnegie Mellon University
Sayan Ghosh, University of Southern California
Salvatore Giorgi, University of Pennsylvania
Karina H Halevy, Carnegie Mellon University
Shreya Havaldar, University of Pennsylvania
Chia-Chien Hung, NEC Laboratories Europe and Universität Mannheim
Akbar Karimi, Rheinische Friedrich-Wilhelms Universität Bonn
Harmanpreet Kaur, University of Minnesota - Twin Cities
Vivek Kulkarni, University of California Santa Barbara
Jonathan K. Kummerfeld, University of Sydney
Hwanhee Lee, Chung-Ang University
Inna Wanyin Lin, University of Washington
Shijia Liu, Northeastern University
Li Lucy, Allen Institute for Artificial Intelligence and University of California Berkeley
Meryem M'hamdi, University of Southern California
Nicole Meister, Stanford University
Jimin Mun, Carnegie Mellon University
Monica Munnangi, Northeastern University
Matthias Orlikowski, Universität Bielefeld
Barbara Plank, Ludwig-Maximilians-Universität München and IT University of Copenhagen
Joan Plepi, Rheinische Friedrich-Wilhelms Universität Bonn
Jielin Qiu, Carnegie Mellon University
Hannah Rashkin, Google
Naba Rizvi, University of California, San Diego
Paul Röttger, Bocconi University
Vahid Sadiri Javadi, Technische Universität Chemnitz
Ian Stewart, Pacific Northwest National Laboratory
Zeerak Talat, Mohamed bin Zayed University of Artificial Intelligence
Anvesh Rao Vijjini, University of North Carolina, Chapel Hill
Akhila Yerukola, Carnegie Mellon University

Hye Sun Yun, Northeastern University

Volunteer

Mounika Marreddy, University of Bonn

Table of Contents

Program

Thursday, August 15, 2024

09:00 - 09:10 *Opening Remarks*

09:10 - 09:55 *Keynote 1*

09:55 - 10:30 *Oral Presentation 1*

10:30 - 11:00 *Coffee Break*

11:00 - 11:45 *Keynote 2*

11:45 - 13:45 *Poster Presentation*

Human Speech Perception in Noise: Can Large Language Models Paraphrase to Improve It?
Anupama Chingacham, Miaoran Zhang, Vera Demberg and Dietrich Klakow

Human-Centered Design Recommendations for LLM-as-a-judge
Qian Pan, Zahra Ashktorab, Michael Desmond, Martín Santillán Cooper, James M. Johnson, Rahul Nair, Elizabeth M. Daly and Werner Geyer

Parameter-Efficient Detoxification with Contrastive Decoding
Tong Niu, Caiming Xiong, Yingbo Zhou and Semih Yavuz

To What Extent Are Large Language Models Capable of Generating Substantial Reflections for Motivational Interviewing Counseling Chatbots? A Human Evaluation
Erkan Basar, Iris Hendrickx, Emiel Krahmer, Gert-Jan de Bruijn and Tibor Bosse

Vision-Language Models under Cultural and Inclusive Considerations
Antonia Karamolegkou, Phillip Rust, Ruixiang Cui, Yong Cao, Anders Søgaard and Daniel Hershcovich

Evaluating Large Language Models on Social Signal Sensitivity: An Appraisal Theory Approach
Zhen Wu, Ritam Dutt and Carolyn Rose

Aligning to Adults Is Easy, Aligning to Children Is Hard: A Study of Linguistic Alignment in Dialogue Systems
Dorothea French, Sidney D'Mello and Katharina Von Der Wense

Thursday, August 15, 2024 (continued)

[Non-Archival] Exploring Human-AI Interaction: A Case Study on the Diplomacy Game
Shumin Deng, Jintian Zhang, Ningyu Zhang and Bryan Hooi

[Non-Archival] Learning from Teaching Assistants to Formulate Subgoals for Programming Tasks: Exploring the Potential for AI Teaching Assistants
Changyoon Lee, Junho Myung, Jieun Han, Jiho Jin and Alice Oh

[Non-Archival] Reference-free Medical Multi-document Summary Evaluation Metric via Contrastive Learning
Jimin Lee and Hwanhee Lee

[Non-Archival] Offline RLHF Methods Need More Accurate Supervision Signals
Shiqi Wang, Zhengze Zhang, Wang Xiaoliang, Rui Zhao, Fei Tan and Nguyen Cam-Tu

[Non-Archival (Published Papers)] Direct Preference Optimization with an Offset
Afra Amini, Tim Vieira and Ryan Cotterell

[Non-Archival (Published Papers)] DOSA: A Dataset of Social Artifacts from Different Indian Geographical Subcultures
Agrima Seth, Sanchit Ahuja, Kalika Bali and Sunayana Sitaram

[Non-Archival (Published Papers)] Book2Dial: Generating Teacher Student Interactions from Textbooks for Cost-Effective Development of Educational Chatbots
Junling Wang, Jakub Macina, Nico Daheim, Sankalan Pal Chowdhury and Mrinmaya Sachan

[Non-Archival (Published Papers)] Evaluating Large Language Model Biases in Persona-Steered Generation
Andy Liu, Mona Diab and Daniel Fried

[Non-Archival (Published Papers)] My Answer is C": First-Token Probabilities Do Not Match Text Answers in Instruction-Tuned Language Models
Xinpeng Wang, Bolei Ma, Chengzhi Hu, Leon Weber-Genzel, Paul Röttger, Frauke Kreuter, Dirk Hovy and Barbara Plank

12:45 - 13:45 *Lunch*

13:45 - 14:15 *Keynote 3*

14:15 - 15:00 *Oral Presentation 2*

Thursday, August 15, 2024 (continued)

15:00 - 15:30 *Brainstorming Session*

15:30 - 16:00 *Coffee Break*

16:00 - 16:30 *Keynote 4*

16:30 - 17:25 *Panel Discussion*

17:25 - 17:30 *Closing Remarks*