ICLR 2024 Workshop on Large Language Models for Agents

https://a2llm.github.io/

Workshop summary

This workshop seeks to delve into the significance of language agents, marking a transformative step in AI's evolution. Building on the current huge progress on large language models, we'll focus on **interactive agents** that **perform intricate tasks** in both real and simulated environments guided by **natural language instructions**. Uniquely, these agents employ language prompts to a sophisticated degree, utilizing them not merely for communication but also for reasoning. Using language for interactions among agents (including humans) paves the way for advanced multi-agent systems. The latest LLM-driven agents have also adopted language for their reasoning processes, consequently simplifying the incorporation of varied external stimuli and enabling multi-step planning and reasoning in a non-programmed and explicit manner. This progressive development amplifies the variety and complexity of challenges such agents can tackle autonomously.

We would like to discuss the following topics in this workshop:

(1) Memory Mechanisms and Linguistic Representation: This session will analyze the similarities between LLMs and human memory and will discuss the mechanisms of storage and formation of the linguistic representation in LLMs.

- How can we bridge the gap between language agent memory and human memory mechanisms?
- What promising methods are on the horizon for updating parametric memory and storing new information?
- How could a combined approach of a parametric memory (holding implicit memory and some semantic memory) and a vector database (storing more up-to-date semantic memory and episodic memory) better mimic human long-term memory systems?

(2) Tool Augmentation and Grounding (interaction with the environment): Addressing the enhancement of LLMs through tool augmentation provides up-to-date and/or domain-specific information. It aslo enables a language agent to take actions in real-world environments.

- How does tool augmentation advance the capabilities of LLMs and language agents, particularly in providing up-to-date/domain-specific information, specialized capabilities, and enabling actions in real-world environments?
- What are the current challenges and potential solutions in enabling language agents to figure out when and how to use a tool automatically, and integrating the execution results into their reasoning process?
- What are the implications of the advancements in tools, and the emerging research on tool set expansion, learning new tools, and creating new tools using LLMs? And what is the role of human involvement in ensuring the safe use of tools?

(3) Reasoning, Planning, and Risks: This session will discuss the intertwined processes of reasoning and planning in language agents and highlight the potential hazards associated with language agents' ability to autonomously operate in the real world.

- What types of reasoning are existing LLMs particularly proficient or deficient at, and how can we understand the origins of these reasoning capabilities or the lack thereof?
- In what ways have LLM-powered embodied language agents started to reason explicitly using language, and how does this impact their generalizability, explainability, and learning efficiency?
- How have recent developments like chain-of-thought and tree-of-thought reasoning algorithms impacted the capability of language agents? Given the field's evolution and our growing understanding, what potential advancements might we expect to see in reasoning algorithms to utilize this newfound language-driven reasoning capability in language agents?

(4) Multi-modality and Embodiment: This will explore how language agents can integrate multiple modalities such as vision, sound, and touch to enhance their understanding and interaction with the environment. The discussion will also touch upon the challenges and benefits of multi-modal learning and how it can lead to a more holistic and robust AI system.

- How can language agents adopt a multi-modal approach that integrates various modalities such as vision, sound, and touch to enhance environmental interactions and understanding?
- As we begin to extend language agents into the realm of embodiment, what future developments could we expect, particularly in regards to their application in various modalities such as robotics?
- As language agents become more autonomous and capable of interacting with their environment, what potential risks (e.g., bias, fairness, hallucination, privacy, irreversible actions) must we consider and mitigate? Given the potential risks of language agents, how should we approach their deployment in the real world, and what collaborative measures from researchers, practitioners, regulators, and the public are required to ensure effective risk management?

Workshop format

This **1 day** workshop will involve 7 invited talks (30 min presentation), 3 oral presentations selected from the peer-reviewed submissions (10 min presentation), one panel discussion (50 min presentation), and two poster sessions for all accepted peer-reviewed submissions (all talks are in a large-attendance talk format).

Organizers

- Xinyun Chen (Senior Research Scientist, Google DeepMind)
- Robert Tang (PhD student, Yale University)
- Di Jin (Senior Applied Scientist, Amazon)
- Devamanyu Hazarika (Applied Scientist, Amazon)
- Daniel Fried (Assistant Professor, CMU LTI)
- Dawn Song (Professor, UC Berkeley)
- Shafiq Joty Rayhan (Research Director, Salesforce AI & Associate Professor, NTU)
- Meredith Ringel Morris (Director & Principal Scientist for Human-AI Interaction Research, Google DeepMind)

Invited speakers

All our invited speakers have confirmed their attendance.

- **Denny Zhou** (speaker & panelist): Denny is the founder and lead of LLM Reasoning Team in Google DeepMind.
  Research Topics: reasoning and chain-of-thought prompting.
- **Dilek Hakkani-Tur** (speaker & panelist): Dilek is a Professor of Computer Science at University of Illinois, Urbana-Champaign. More recently, she worked as a senior principal scientist at Amazon Alexa AI focusing on enabling natural dialogues with machines.
  Research Topics: building conversational agents.
- **Asli Celikyilmaz** (speaker): Asli is a Research Manager at Fundamentals AI Research (FAIR) Labs at Meta AI.
  Research Topics: exploring machine thinking.
- **Chelsea Finn** (speaker): Chelsea is an Assistant Professor in Computer Science and Electrical Engineering at Stanford University. She also spent time at Google as a part of the Google Brain team.
  Research Topics: vision-language models and robotic control.
- **Karthik Narasimhan** (speaker & panelist): Karthik is an Assistant Professor in Computer Science at Princeton University.
  Research Topics: language agents.
- **Graham Neubig** (speaker): Graham is an Associate Professor at the Language Technologies Institute of Carnegie Mellon University.
  Research Topics: *large language models for code generation.*
- **Joyce Chai** (speaker & panelist): Joyce is a Professor in the Department of Electrical Engineering and Computer Science at the University of Michigan.
  Research Topics: *embodied language agents.*

Panel Discussion / Q&A: *Exploring the methods, tasks, theories, and risks of language agents*

Covering the broadness of agents through the lens of memory modeling, tool enhancement, and grounded language understanding, our panel invites some of our speakers to explore the challenges and opportunities in building versatile language agents. The panelist will discuss the methods, tasks, theories, and risks associated with LLM-driven agents that are capable of using language as a tool for thought and communication. This insightful discussion will shed light on both theoretical underpinnings and practical implications in the field.

Logistics

This workshop will be a **hybrid** event, combining both in-person and virtual participation. We plan to utilize physical conference rooms for presentations, poster sessions, and panel discussions, while also synchronizing sessions through **ZOOM** to accommodate remote attendees (we have Zoom Licensed). Furthermore, the workshop content will be recorded and uploaded to YouTube for broader accessibility. Some of our organizers will be present on-site to facilitate and coordinate the workshop, while we will also have online organizers assisting with tasks such as collecting questions and feedback from remote participants.

We expect more than **100** submissions and we will accept around 40 papers. More than **150** people are expected to attend in-person, and more than **100** virtually. The workshop will provide an opportunity for discussion through two main venues. First, the standard poster sessions will allow for discussion among all workshop attendees in the usual manner. Second, half of Q&A questions in the panel discussions will be solicited from attendees. These questions can be collected and selected based on popularity before the workshop or live during the session (e.g. using the sli.do question and answering platform).

Moreover, we will be operating a Slack community to facilitate communications and online discussions. And we plan to leverage Twitter and other media outlets to publicize our workshop. These platforms will be instrumental in disseminating information about the conference, as well as fostering an online academic dialogue (e.g. spontaneous paper reading groups initiated by the community).

We will use OpenReview to manage the submission process. We aim to raise $1,500 in sponsorship funds to support our workshop. We have initiated contact with potential sponsors, including Google, Meta, and some startups, like XihuXinchen and AI Waves. To ensure effective communication and collaboration, we have reached out not only to their researchers but also to the relevant departments, such as Marketing and HR, through their researcher lead.

The funds raised will primarily be utilized for purchasing food and beverages for the workshop attendees. Additionally, we are setting up **DEI (Diversity, Equity, and Inclusion) grants** and the **Best Paper Award**. These are integral to our commitment to fostering inclusive and encouraging high-quality submissions.

Workshop submissions

All paper submissions will be managed through the OpenReview platform. The reviewing process will be double-blind. We invite submissions of both short (up to 4 pages) and long (up to 8 pages) research papers. These submissions should contain original research and authors have the choice of making these submissions **archival** or **non-archival**. non-archival submissions allow for dual submission, where it is permitted by third parties. All papers can be published elsewhere at a later date.

We are also introducing a **special track** dedicated to system reports or demonstrations. These can be comprehensive descriptions of open-source frameworks or detailed explanations of specific system operations. This platform aims to present the latest technological advancements in a more accessible and interactive manner for workshop attendees. Thus, the paper submission formats comprise research papers, technical papers, position papers, and system demonstrations. The submission should use the LaTeX style files provided for ICLR 2024: https://github.com/ICLR/Master-Template/raw/master/iclr2024.zip.

Proposed schedule

Time	**Event**	**Duration**
Morning Session		

9:00 - 9:10	Opening Remarks	10 min
9:10 - 9:40	Invited Talk #1	30 min
9:40 - 10:10	Invited Talk #2	30 min
10:10 - 10:20	Oral Presentation #1	10 min
10:20 - 10:30	Oral Presentation #2	10 min
10:30 - 10:40	Oral Presentation #3	10 min
10:40 - 11:10	Coffee Break	30 min
11:10 - 11:40	Invited Talk #3	30 min
11:40 - 12:30	Poster Session #1	50 min
12:30 - 13:30	Lunch Break	60 min
Afternoon Session		
13:30 - 14:00	Invited Talk #4	30 min
14:00 - 14:30	Invited Talk #5	30 min
14:30 - 15:00	Invited Talk #6	30 min
15:00 - 15:30	Invited Talk #7	30 min
15:30 - 16:20	Panel Discussion	50 min
16:20 - 16:30	Closing Remarks	10 min
16:30 - 17:30	Poster Session #2 and Happy Hour	60 min

Relation to other workshops

Our Language Agents workshop has associations and overlaps with several other workshops in the AI community. Below, we outline the key relationships and interactions with each one:

- Foundations and Applications in Large-scale AI Models - Pre-training, Fine-tuning, and Prompt-based Learning Workshop @ KDD 2023: This workshop explores the paradigm shifts in AI models, specifically focusing on large-scale models' pre-training, fine-tuning, and prompting. Our workshop builds on this premise, emphasizing the significance of sophisticated language use within these AI models, and the potential implications for autonomous task handling.
- The Symposium on Advances and Open Problems in Large Language Models @ IJCAI 2023: LLM Symposium primarily explores LLMs applications and their associated challenges,

specifically with regards to trustworthiness and ethical concerns. Our workshop extends this discussion, focusing on the role of LLM-driven agents in complete AI system functioning.

- Instruction Workshop @ NeurIPS 2023: The Instruction workshop probes the advancements in the instruction-following capabilities within LLMs. Our workshop relates to this by exploring the implications of such instructions-based improvements on language agent capacity in both real and simulated environments.
- Workshop on Enormous Language Models @ ICLR 2021: This workshop investigates the scalability and boundaries of Large Language Models (LLMs). Our workshop resonates with such discussions by considering these factors from a unique vantage point, focusing on the development and integration of tool using within LLMs.
- The Workshop on Knowledge Retrieval and Language Models @ ICML 2022: The ICML workshop delves into the discussion of knowledge retrieval limits in LLMs. Our workshop intersects with this theme, examining the interplay between knowledge retrieval and multi-modal integrations.
- Challenges & Perspectives in Creating Large Language Models @ ACL 2022: This workshop addresses the broader concerns related to the creation of LLMs such as infrastructure, ethical and legal frameworks, and evaluation measures. Our workshop fits into the discourse around these themes by emphasizing the need for a conceptual framework to guide the development and performance of language agents.

It's important to note that while our workshop interfaces with all these other workshops, what sets us apart is our detailed focus on the abilities of AI language agents. We're addressing not just their construction or use, but delving deeper into their function as entities capable of meaningful interaction with the world.

Diversity and inclusion

In organizing this workshop, we have made extensive considerations for diversity to promote an inclusive workshop without sacrificing quality. With respect to minority representation, **four** of our invited speakers (Dilek, Asli, Chelsea, Joyce) and **three** of our organizers (Xinyun, Meredith, Dawn) identify as women in AI. Our invited speakers include researchers from both industry (Google, Meta, Amazon, Salesforce) and academia (Stanford, CMU, Princeton University, UIUC, Michigan). Our invited professors span the full range of seniority from the assistant to the full professor level.

Our keynote sessions will showcase a diverse range of topics, presented by experts from various research areas including Dilek, specializing in speech and language processing, Chelsea, an authority in robotics, and Denny Zhou, focused on machine learning.

Opportunity for discussion

The workshop will provide an opportunity for discussion through two main venues. First, the standard poster sessions will allow for discussion amongst all workshop attendees in the usual manner. Second, the panel discussions and invited talks will include Q&A sessions at the end for which questions can be solicited from attendees. These questions can be collected and selected based on popularity before the workshop or live during the session (e.g. using the sli.do question and answering platform).

Timeline

We plan on using the following timeline, which gives three weeks after the ICLR decisions for authors to prepare their submissions, and two weeks to collect reviews, in order to meet the March 3rd deadline for notifying contributions of their acceptance status.

January 15th, 2024: ICLR decision notification
January 21th, 2024: Workshop submission deadline
February 26th, 2024: Acceptance notification

Organizers

Our organizing committee has rich research experience in LLM and autonomous agents across several domains. Their diverse research topics can synergistically contribute to the workshop and ensure coverage of the various topics.

They have extensive experience in organizing workshops at venues including ICLR, NeurIPS, AAAI, ACL, EMNLP, CVPR, and ECCV. Specifically, Xinyun has co-chaired **13** workshops in top AI conferences, Di and Xinyun both led the organizing team of previous workshops at ICLR before. Di, Dev, and Robert have organized a previous version of this workshop series (1st Workshop on Taming Large Language Models at SIGDIAL 2023), and this proposed workshop could be seen as the second-year version of their collaboration.

Xinyun Chen (xinyunchen@google.com) is a senior research scientist at Google DeepMind. She obtained her Ph.D. in Computer Science from University of California, Berkeley. Her recent research focuses on large language models, code generation, and reasoning. She received the Facebook Fellowship in 2020, and Rising Stars in Machine Learning in 2021. She has co-organized 13 workshops and tutorials at ICLR, ICML, CVPR, ECCV, ICCV, AAAI, etc., and she is a co-chair of AISec 2022 and 2023.

Robert Xiangru Tang (xiangru.tang@yale.edu) is a Ph.D. student at Yale University advised by Mark Gerstein. Previously, he completed his Master's at Yale as well advised by Dragomir R. Radev. His research interests are large language models and code generation. He interned at Tencent and Microsoft Research. He also serves as a program committee for multiple conferences and workshops, including ACL, EMNLP, NAACL, NeurIPS, ICML, etc.

Di Jin (djinamzn@amazon.com) is a Senior Applied Scientist at Amazon. His research interests focus on large language models, conversational modeling, transfer learning, and model robustness of Natural Language Processing. He received his Ph.D. degree from MIT, supervised by Dr. Peter Szolovits. He has co-organized the RobustML Workshop at ICLR 2021, Meta Learning and Its Applications to Natural Language Processing Workshop at ACL 2021, Taming Large Language Models workshop at SigDIAL 2023. He has been serving on the Program Committee (as a reviewer, meta-reviewer, or area chair) of multiple conferences, including ACL, EMNLP, NAACL, NeurIPS, ICLR, ICML, AAAI, amongst others.

Devamanyu Hazarika (devamanyu@u.nus.edu) is an Applied Scientist at Amazon Alexa AI. His research goal is to build controllable dialog agents in an efficient and scalable manner. Before Amazon, he received his Ph.D. degree from the National University of Singapore. Dev has helped organize SSNLP 2020 as a Volunteer chair and is currently serving as a publicity chair at ACL 2023, SPC at AAAI '23, '24,

AC in ACL 2023 and EMNLP 2023. He has also served on the program committee of multiple conferences, including ACL, EMNLP, NAACL, AAAI, NeurIPS, amongst others.

Daniel Fried (dfried@cs.cmu.edu) is an assistant professor at the Language Technologies Institute in the School of Computer Science at Carnegie Mellon University, working on natural language processing: language interfaces, code generation, and language grounding. Before CMU, he was a postdoc at FAIR Seattle and the University of Washington, and completed a PhD at UC Berkeley. His research has been supported by an Okawa Foundation Research Award, a Google PhD Fellowship, an NDSEG Fellowship, and a Churchill Scholarship. He has been a workshop co-organizer for four previous workshops at NLP and ML venues.

Dawn Song (dawnsong@cs.berkeley.edu) is a Professor in the Department of Electrical Engineering and Computer Science at UC Berkeley. Her research interest lies in deep learning and security. She has studied diverse security and privacy issues in computer systems and networks, including areas ranging from software security, networking security, database security, distributed systems security, applied cryptography, to the intersection of machine learning and security. She is the recipient of various awards including the MacArthur Fellowship, the Guggenheim Fellowship, the NSF CAREER Award, the Alfred P. Sloan Research Fellowship, the MIT Technology Review TR-35 Award, the George Tallman Ladd Research Award, the Okawa Foundation Research Award, the Li Ka Shing Foundation Women in Science Distinguished Lecture Series Award, the Faculty Research Award from IBM, Google and other major tech companies, and Best Paper Awards from top conferences. She obtained her Ph.D. degree from UC Berkeley. Prior to joining UC Berkeley as a faculty, she was an Assistant Professor at Carnegie Mellon University from 2002 to 2007.

Shafiq Joty (sjoty@salesforce.com) is a Research Director at Salesforce AI Research in Palo Alto, USA. He served as an Assistant, and then Associate Professor at Nanyang Technological University in Singapore (on leave). He started as a research scientist at Qatar Computing Research Institute after receiving a PhD in computer science from the University of British Columbia. He is the PC co-chair for SIGDIAL-2023 and a SAC for EMNLP-23, AC for ICLR-23, NeurIPS-23, and ACL-23. He is honored to be part of the ICLR-23 and NAACL-22 best paper selection committees.

Meredith Ringel Morris (meredith.morris@gmail.com) is a computer scientist conducting research in the areas of human-computer interaction (HCI), human-centered AI, human-AI interaction, computer-supported cooperative work (CSCW), social computing, and accessibility. She is honored to have been recognized as an ACM Fellow and ACM SIGCHI Academy member for my contributions to HCI. She is currently Director for Human-AI Interaction Research for Google DeepMind. Prior to joining Google DeepMind (previously Google Brain), she was the Director of the People + AI Research team in Google Research's Responsible AI organization. She is also an Affiliate Professor in The Paul G. Allen School of Computer Science & Engineering and in The Information School at the University of Washington, where she participates in the dub research consortium. Previously, she was a Sr. Principal Researcher and Research Area Manager for Interaction, Accessibility, and Mixed Reality at Microsoft Research, where she founded the Ability research group.

Program Committee

We are expecting approximately 100 paper submissions, and have thus reached out to around 120 individuals to assist in the review process.
Thus, each paper can have three reviews, and each person only needs to review up to three papers.

Lisa Adams, Mihali Felipe, Declan Clarke, Joel Rozowsky, Prashant Emani, Shaoke Lou, Timur Galeev, Beatrice Borsari, Cagatay Dursun, Christopher J.F. Cameron, Gaoyuan Wang, Matthew Jensen, Mor Frank, Muhammed Erguder, Pengyu Ni, Ran Meng, Xiao Zhou, Yan Xia, Anna Su, Dingyao Zhang, Donglu Bai, Eric Ni, Jason Liu, Jiahao Gao, Jiaqi Li, Suchen Zheng, Tianxiao Li, Weihao Zhao, Xin Xin, Xuezhu (Michelle) Yu, Yuhang Chen, Yunyang Li, Yunzhe Jiang, Zhiyuan (Andy) Chu, Eve Wattenberg, Israel Yolou, Maya Geradi, Michael Gancz, Susanna Liu, Ananya Rajagopalan, Andrei Onut, Andrew Tran, Andy Gu, Andy Yang, Bill Qian, Danielle Lee, Daraeno(Dara) Ekong, Howard Dai, Jacqueline Wang, Jeffrey Tan, Jiakang Chen, Leo deJong, Michael Ofodile, Nawal Naz Tareque, Raymond Lee, Rick Gao, Sahil Chhabra, Serena Ulammandakh, Tom Sutter, Yifang (Lucy) Zha, Aatmik Mallya, Cengizhan Buyukdag, Elizabeth (Elianna) Knight, Eric Nguyen, Eva Skarabot, Heng-Le Chen, Heyu (Leah) Wang, Isabella Wu, Ke Xu, Luoshu Zhang, Yihan Liu, Yuan Gao, Zhongzheng Mao.

We also plan to send additional reviewer invitations to authors of the submissions. We maintain a pool of emergency reviewers whom we have not yet contacted but will do so if necessary.