

ACL 2026

**The 64th Annual Meeting of the Association for
Computational Linguistics (ACL 2026)**

Proceedings of the Conference - Volume 2: Short Papers

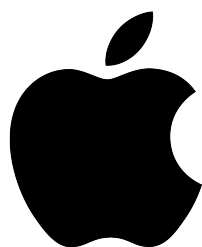
July 2-7, 2026

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

Diamond



Platinum



Gold



JPMorganChase



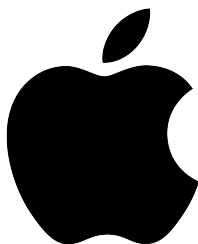
Silver



Bronze



Academic Ally



©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 979-8-89176-391-3

Message from the General Chair

It is my great pleasure and honor to welcome you to the 64th Annual Meeting of the Association for Computational Linguistics (ACL 2026), held in San Diego, California, USA, July 2–7, 2026. ACL 2026 continues our field’s tradition of excellence in scholarship, innovation, and inclusivity, and I am deeply grateful to the many volunteers who have worked tirelessly to bring this event to life.

It is exciting to see the many developments in the field reflected at our conference, such as the advancement of large language models (born out of our community) into many application directions. This is not without challenges, partly due to unsolved problems, and partly due to misuse of the technology. All this is reflected well at ACL 2026. The conference also reflects that the field is not only about advancing technology, but also the science of fundamentally understanding language and how it is used by many different speaker communities.

I want to express my deepest thanks to our Program Chairs — Maria Liakata, Viviane P. Moreira, Jijun Zhan, and David Jurgen — who have overseen the reviewing and selection process, and shaped a compelling and diverse scientific program. It is impossible to overstate how much work and dedication they put in to make this conference successful. Overall, we received more than 12,148 submissions and accepted 2,296 at the main conference and 2,163 as findings. This was made possible through close coordination with the ACL Rolling Review (ARR) team and I sincerely thank them. We are also deeply grateful to the many senior area chairs, area chairs, reviewers, and the Best Paper Committee (led by Antonios Anastasopoulos and Eunsol Choi), whose dedication ensured the high quality of our program. Isabelle Augenstein served as Conflict Chair to shepherd papers that were submitted by conference chairs.

I am also grateful to our Technical Open Review Chairs, Shuoyang Ding and Yi R. (May) Fun, for their behind-the-scenes work. For those joining virtually, our Virtual Infrastructure Chairs — Gael Guibon and Parisa Kordjamshidi — have worked to make the hybrid conference experience inclusive and engaging. I would also like to recognize the efforts of our Ethics Chairs — Margaret Mitchell and Paco Guzman — who have led the work aimed at ensuring that submissions meet the ethical standards of our field.

ACL 2026 hosts a rich set of 37 workshops and 6 tutorials, thanks to the dedicated efforts of our Workshop Chairs — Loic Barrault and Yang Zhao — and our Tutorial Chairs — Kenton Murray and Jacob Andreas. Our Demonstration Track Chairs — Greg Durrett and Ping Jian — have put together an impressive set of system demonstrations of cutting-edge innovations in NLP. I am proud to say that we also have a rich Industry Track, thanks to Yunyao Li, Georg Rehm, and Mei Tu who created a “conference in the conference”. Their experience from organizing the industry track already last year helped advance the visibility and impact of the event. We also hosted a special linguistic symposium entitled “Linguistics and NLP in the LLM Era” which featured Allyson Ettinger, Richard Futrell, Zoey Liu and Tom McCoy. As is customary, TACL and CL papers are presented at the conference. Thanks go to the TACL Editors-in-Chief — Asli Celikyilmaz, Roi Reichart, Dilek Hakkani-Tur, and Yun-Nung (Vivian) Chen — and the CL Editor-in-Chief, Wei Lu, for their coordination efforts.

This conference could only have been made possible thanks to the invaluable work of our ACL Business Manager and Director of Events, Jenn Rachford, and her team. Our Visa Chairs — David Chiang and Swabha Swayamdipta — have helped numerous participants with their visa needs and any logistical difficulties in attending the conference. I also thank our Student Volunteer Chairs — Jingbo Shang, Lianhui Qin, Xuezhe Ma, and Muhao Chen — for their behind-the-scenes work, which proved indispensable during the conference.

Additional thanks go to Santosh T.Y.S.S., Juan Diego Rodriguez, and Ona de Gibert, our Student Research Workshop Chairs, who, with the support of faculty advisors Yves Scherrer, Elias Stengel-Eskin, and Snigdha Chaturvedi, devoted their work to supporting and mentoring the next generation of researchers. They also secured additional funding from the Vienna Meeting Fund — congratulations!

Our Publication Chairs — David Stap, Haoran Xu, and Jan Niehues — have ensured timely and high-quality conference proceedings. I also thank the Handbook Chairs, Jia Xu and Ryan Cotterell, for assembling the handbook that will guide attendees through the conference program. Of course our conference

could not exist without the support of our sponsors! We are deeply grateful to our Sponsorship Chairs — Baotian Hu and Steve Richardson — and to Joel Tetrault, ACL’s Sponsorship Director, for securing the generous support that enables us to keep the conference accessible to as many participants as possible. Fostering an inclusive environment remains a key goal of ACL. Our Diversity and Inclusion Chairs — Ximena Gutierrez-Vasques, Anietie Andy, Gabrielle Gaudeau, and Felermino Ali — have worked to support diverse participation and organize activities that reflect the global nature of our vibrant community. An inclusive ACL allows researchers from all areas and backgrounds to contribute, collaborate, and thrive — enriching both the scientific conversation and the broader societal impact of our work. We remain committed to building a community where everyone feels welcomed and valued.

Thank you to the Publicity and Social Media Chairs — Raj Dabre, Wei Xu, Flor Miriam Plaza-del-Arco, Debora Nozza, and Yaping Zhang — for their communication work through social media, extending the conference’s reach to a wider audience. And this brings me to our Website and App Chairs — Abulhair Saparov and Najoung Kim — who kept our community informed with timely updates about all aspects of the conference. The huge number of emails and communications that we received was smoothly managed throughout the year by our Internal Communication Chairs, Ming Jiang and Margot Mieskes. We are very grateful!

Last but not least, I wish to thank the ACL Executive Committee for their guidance, and my fellow past General Chairs for generously sharing their experience and insights.

ACL 2026 is a community-wide effort. Whether you are presenting a paper, leading a workshop or tutorial, volunteering, mentoring, supporting or attending to learn and connect, your presence and participation make a difference at this conference. Thank you for being part of it!

Welcome to San Diego. Welcome to ACL 2026!

Philipp Koehn

Johns Hopkins University
General Chair, ACL 2026

Message from the Program Chairs

Welcome to the 64th Annual Meeting of the Association for Computational Linguistics! ACL 2026 will be held in a hybrid format, offering attendees the option to join us in person in San Diego, California, or to participate remotely from anywhere in the world.

Review Process. All submissions to ACL 2026 went through a two-stage review process. First, papers were submitted to the ACL Rolling Review (ARR), where they were reviewed by reviewers and received meta-reviews from Area Chairs. Then, authors had the option to commit their reviewed papers to ACL via a separate ACL 2026 commitment site. At this stage, Senior Area Chairs provided recommendations, and the Program Chairs made final acceptance decisions based on SAC recommendations, but also with a view to obtaining a balanced program and prioritizing work at a mature state of development. This process is consistent with previous conferences, including ACL 2025, EMNLP 2025, and EACL 2026. We worked closely with the ARR team, particularly the ARR January 2026 Editors-in-Chief, and served as guest Editors-in-Chief for this round. We helped recruit new reviewers and Area Chairs to ARR, resulting in 8,594 reviewers and 1,434 Area Chairs in the January ARR cycle, which received 10,518 submissions. ACL also recruited 255 Senior Area Chairs to oversee the review and meta-review process. Additionally, we collaborated with the ARR October 2025 cycle, which was shared with EACL 2026, and received 1,630 submissions listing ACL as their preferred venue. We recruited 23 Senior Area Chairs and also served as “super” Senior Area Chairs, overseeing the reviewing process of a large number of submissions.

Coordinating a review process of this scale required the efforts of a large number of volunteers. Delays at multiple stages of the process, including reviewing, meta-reviewing, and recommendation, placed additional burdens on the Program Chairs and ARR Editors. We are particularly grateful to those who absorbed this additional workload, serving as emergency Reviewers, Area Chairs, and Senior Area Chairs. For the ACL commitment phase, Senior Area Chairs made recommendations for 6,564 committed papers based on the reviews, meta-reviews, author responses, and the papers themselves, with final acceptance decisions made by the Program Chairs. The vast majority of papers committed to ACL 2026 came from the ARR January 2026 cycle (88%), followed by ARR October 2025 (10.4%). A small number of papers were committed directly from earlier cycles, including July 2025 (1.2%) and May 2025 (0.4%).

During the ARR review process, reviewers identified many non-existent references, which led to those submissions being desk rejected. During the final consistency checks of the camera-ready versions of papers accepted to ACL 2026, we identified over 100 papers that contained citations to non-existent literature. Since these papers were going to be published and were no longer subject to the anonymity requirements, WisPaper was used to identify potential citation issues, and then all flagged citations were reviewed by human experts (PCs and SACs) to confirm that references did not exist. Whether human-authored or LLM-generated, the inclusion of these non-existent references was a clear violation of the ACL Policy on Publication Ethics. For this reason, we had to reject such papers at this late stage. We recommend that future cycles should run such checks both at the submission stage and at the camera-ready stage to prevent the contamination of the scientific literature with false references and to identify such cases as early as possible.

Acceptance Rate. In total, there are 2,296 papers accepted to the Main Conference and 2,163 papers accepted to Findings. The acceptance rate calculation follows the precedent set by previous conferences that go through ACL Rolling Review (ARR), e.g., ACL 2025. The calculation takes into account the multi-stage process of ARR, in which a paper may be revised in ARR and later committed to the conference. In total, we had 12,148 unique submissions across the October 2025 (1,630) and January 2026 ARR (10,518). The acceptance rate is 18.9% for Main Conference papers and 17.8% for Findings papers.

Special Theme: Explainability of NLP Models. Following the precedent of theme tracks from previous events, ACL 2026 features a special theme designed to reflect on and stimulate discourse regarding

the current trajectory of the NLP field. Explainability encompasses the methodologies and techniques intended to render the internal mechanisms of complex NLP models, such as Large Language Models (LLMs), transparent and human-interpretable. It seeks to move beyond the “black box” paradigm, where predictions are accepted without scrutiny, and instead aims to elucidate the underlying reasoning of specific model outputs. Explainability is fundamental to establishing trust, ensuring fairness, and enabling responsible deployment. By identifying a model’s potential dependence on spurious correlations or inherent societal biases, explainability empowers developers to diagnose system errors, enhance robustness, and ensure accountability. This is particularly vital in high-stakes domains—including healthcare, finance, and law—where interpreting the rationale behind a decision is as essential as the decision itself. The theme track invited empirical and theoretical work, as well as surveys and position papers reflecting on the Explainability of NLP Models. Suggested topics of discussion included: (i) How do explainability methods need to be adapted for different model architectures? Can we develop a unified framework to evaluate explanations across these architectures? (ii) How can we rigorously and quantitatively evaluate the quality of an explanation? What metrics can reliably measure the faithfulness (accuracy of the model’s reasoning) and plausibility (human-perceived reasonableness) of an explanation? (iii) Can explanations be used to reliably detect when a model is making a biased prediction based on sensitive attributes? How can input-based explanations help mitigate social biases during model training? (iv) Can we use explanations to systematically find and fix problems in the training data itself, such as spurious correlations or annotation errors? How can explainability facilitate a human-in-the-loop process for iterative data refinement? (v) Can we identify specific directions, mechanisms, patterns, or “knobs” within a model’s internal activations that control high-level behaviors like abstaining from unanswerable questions? Can we design models that are inherently more interpretable?

We received 124 submissions to the special theme track during the review phase. Among these, 29 papers were accepted to the main conference and a further 27 to the Findings of ACL 2026. The conference will also feature a panel discussion on this theme, with leading experts from both academia and industry, with one of the keynote talks, by Tania Lombrozo, also focused on explainability.

Best Paper Selection. ACL 2026 implemented the updated ACL award policy that seeks to expand the pool of work recognized as outstanding. In total, 158 papers were nominated by the reviewers, area chairs, and senior area chairs for awards. The best paper committee assessed these papers to select the best papers, outstanding papers, and special awards for social impact and best resource. Separately, the senior area chairs also nominated their favorite papers as SAC Highlights. The selection process was double-blind, using anonymized camera-ready versions of the papers. The final selection was made by the best paper committee, and the winners will be announced during the opening ceremony to provide early recognition of the selected contributions and to maximize their visibility within the conference program. In addition, we have the 2026 ACL Dissertation Award, the CL high impact paper award, and the test of time awards.

Program Composition & Presentation Modes. Based on feedback from the conference support staff and the Underline team after ACL 2025, we decided to hold the virtual presentation sessions during the main conference. This enables us to align the virtual sessions with time slots that suit participants in Europe and Asia. This approach allows virtual attendees to participate concurrently with the physical event, avoiding the need for organizers and attendees to engage with the conference twice and separately. We recognize that many virtual attendees may not be able to attend keynotes during their scheduled times at ACL, so we added a third virtual-only keynote in a timezone aligned with Europe and Asia to allow greater live participation and visibility on this year’s theme.

The in-person conference will hold 7 Oral/Poster Sessions running in parallel across 10 rooms. This year, 504 main conference papers were selected for oral presentation by the program chairs. The goal was to create a balanced program that featured a diverse set of topics, rather than selecting papers solely on the basis of their review scores. To accommodate a larger number of oral presentations, presentation time was reduced to 8 minutes, followed by 2 minutes of questions. The in-person conference will have

2,078 poster presentations from Main and Findings papers. Rather than have all posters from a track in the same session, sufficiently-sized tracks were divided into two or more sessions to allow authors presenting in one session to attend the other poster sessions for their track. This year, Findings papers will be presented alongside the Main Conference posters, demos, and the Student Research Workshop. This follows the positive experience and feedback of ACL 2025 of holding Main and Findings posters concurrently. In addition to the main conference papers, the ACL program also includes 4 papers accepted by Computational Linguistics (CL) and 27 papers accepted by Transactions of the ACL (TACL). Among these, 18 journal papers will be presented in person as oral presentations, thematically distributed across appropriate sessions.

The virtual conference will feature over 1K presentations, including 752 Main Conference and 176 Findings Presentations, as well as Demos, Student Research Workshop, TACL, CL, Industry papers, and a virtual-only keynote talk.

Keynotes. We are happy to have three keynote talks at the conference, two in-person and one virtual.

- Philip Resnik (University of Maryland): A New Balancing Act: Reflections on the Relationship between Computational Linguistics and AI
- Yue Zhang (Westlake University): How Can AI Help Us Manage AI Conferences (where Troubles Are Caused by AI) – virtual
- Tania Lombrozo (Princeton University): Explanation and Understanding: A Perspective from Cognitive Science

Acknowledgments. Organizing ACL 2026 has been a collaborative effort, made possible by the dedication and hard work of thousands of people. We gratefully acknowledge the support and contributions of the following people:

- The General Chair, Philipp Koehn;
- The ARR Editors-in-Chief of the January 2026 cycle (Sarvnaz Karimi, Vincent Ng, Xiaodan Zhu, and Margot Mieskes) and the entire ARR team, especially Holy Lovenia, Freda Shi, and Yuntian Deng from the tech team;
- The technical OpenReview chairs, Shuoyang Ding, May Fung, and the OpenReview support team, in particular, Rachel, Emilia, and Harold, for technical help in setting up ACL 2026 on the OpenReview platform;
- The 255 Senior Area Chairs;
- The 1,434 Area Chairs and the 8,594 reviewers, and 171 ethics reviewers;
- The best paper committee chairs, Antonios Anastasopoulos and Eunsol Choi, and the best paper committee members;
- The ethics chairs, Margaret Mitchell and Paco Guzman;
- The conflicts chairs, Isabelle Augenstein and Joel Tetreault;
- The workshop chairs, Loic Barrault and Yang Zhao;
- The tutorial chairs, Kenton Murray and Jacob Andreas;
- The industry track chairs, Georg Rehm, Yunyao Li, and Mei Tu;
- The demonstration chairs, Greg Durrett and Ping Jian;

- The internal communications chairs, Ming Jiang and Margot Mieskes;
- The website and conference app chairs, Abulhair Saparov and Najoung Kim;
- The publication chairs, David Stap, Haoran Xu, and Jan Niehues;
- The handbook chairs, Jia Xu and Ryan Cotterell;
- The visa chairs, David Chiang and Swabha Swayamdipta;
- The publicity chairs, Raj Dabre, Wei Xu, Flor Miriam Plaza-del-Arco, Debora Nozza, and Yaping Zhang;
- The student research workshop chairs, Santosh T.Y.S.S., Juan Diego Rodriguez, and Ona de Gibert;
- The student research workshop chairs faculty advisors, Yves Scherrer, Elias Stengel-Eskin, and Snigdha Chaturvedi;
- The student volunteer chairs, Jingbo Shang, Lianhui Qin, Xuezhe Ma, and Muhao Chen;
- The diversity and inclusion chairs, Ximena Gutierrez-Vasques, Aniete Andy, Gabrielle Gaudeau, and Felermino Ali;
- The sponsorship chairs, Baotian Hu and Steve Richardson;
- The virtual infrastructure chairs, Gael Guibon, and Parisa Kordjamshidi;
- The ACL Anthology Director Matt Post and his team;
- The TACL editors-in-chief (Asli Celikyilmaz, Roi Reichart, Dilek Hakkani-Tur, and Vivian Chen) and CL Editor-in-Chief Wei Lu for coordinating TACL and CL presentations with us;
- Past program chairs for *CL conferences, in particular Christos Christodoulopoulos (EMNLP 2025), Vera Demberg (EACL 2026), and Ekaterina Shutova (ACL 2025), for information and support;
- Qi Zhang, developer of WisPaper, for checking camera-ready papers to ensure they did not contain hallucinated references;
- Damira Mrsic and the Underline team for help with the program;
- Jennifer Rachford and the entire conference support staff for their crucial help and patience as we learned our way through running such a complex process;
- All the authors of papers who submitted their papers for review in the ARR 2025 October and ARR 2026 January cycles, and those who committed to the ACL 2026 conference.

We hope you enjoy this year's huge and diverse program!

Maria Liakata (Queen Mary University of London, UK)
 Viviane Moreira (UFRGS, Brazil)
 Jiajun Zhang (Chinese Academy of Sciences, China)
 David Jurgens (University of Michigan, USA)
 ACL 2026 Program Co-Chairs

Organizing Committee

General Chair

Philipp Koehn, Johns Hopkins University

Program Chairs

Maria Liakata, Queen Mary University of London
Viviane P. Moreira, Institute of Informatics at UFRGS, Brasil
Jiajun Zhang, Chinese Academy of Sciences
David Jurgens, University of Michigan

Workshop Chairs

Loic Barrault, Meta
Yang Zhao, Chinese Academy of Sciences

Tutorial Chairs

Kenton Murray, Johns Hopkins University
Jacob Andreas, Massachusetts Institute of Technology

Demonstration Chairs

Greg Durrett, New York University
Ping Jian, Beijing Institute of Technology

Student Research Workshop Chairs

Santosh T.Y.S.S., Technical University of Munich & Amazon
Juan Diego Rodriguez, University of Texas at Austin
Ona de Gibert, University of Helsinki

Student Research Workshop Chairs: Faculty Advisors

Yves Scherrer, University of Oslo
Elias Stengel-Eskin, University of Texas at Austin
Snigdha Chaturvedi, University of North Carolina

Publication Chairs

David Stap, NXAI
Haoran Xu, Microsoft
Jan Niehues, Karlsruhe Institute of Technology

Handbook Chairs

Jia Xu, Stevens Institute of Technology

Ryan Cotterell, ETH Zürich

Sponsorship Chairs

Baotian Hu, Harbin Institute of Technology
Steve Richardson, Brigham Young University

Diversity and Inclusion Chairs

Ximena Gutierrez-Vasques, Universidad Nacional Autonoma de Mexico
Anietie Andy, Howard University
Gabrielle Gaudeau, University of Cambridge
Felermimo Ali, Microsoft Research Africa

Publicity Chairs

Raj Dabre, Google
Wei Xu, Georgia Tech
Flor Miriam Plaza-del-Arco, University of Leiden
Debora Nozza, Bocconi University
Yaping Zhang, Chinese Academy of Sciences

Website and Conference App Chairs

Abulhair Saparov, Purdue University
Najoung Kim, Boston University

Ethics Chairs

Margaret Mitchell, Huggingface
Paco Guzman, Meta

Student Volunteer Chairs

Jingbo Shang, University of California at San Diego
Lianhui Qin, University of California at San Diego
Xuezhe Ma, University of Southern California
Muhao Chen, University of California at Davis

Technical Open Review Chairs

Shuoyang Ding, NVIDIA
Yi R. (May) Fung, HKUST

Virtual Infrastructure Chairs

Gael Guibon, Université Sorbonne Paris Nord
Parisa Kordjamshidi, Michigan State University

Internal Communication Chairs

Ming Jiang, Indiana University Indianapolis
Margot Mieskes, University of Applied Sciences, Darmstadt

Industry Track Chairs

Georg Rehm, German Research Institute for Artificial Intelligence
Yunyao Li, Adobe
Mei Tu, Samsung

Best Paper Chairs

Antonios Anastasopoulos, George Mason University
Eunsol Choi, New York University

Visa Chairs

David Chiang, University of Notre Dame
Swabha Swayamdipta, University of Southern California

Conflict Chair

Isabelle Augenstein, University of Copenhagen

Keynote Talk

A New Balancing Act: Reflections on the Relationship between Computational Linguistics and AI

Philip Resnik
University of Maryland



Sunday, July 5, 2026 – Time: 09:30 – Room: Harbor Ballroom Level 2

Abstract: Beginning in the mid-1980s, a “statistical revolution” in our field redefined what it meant to work on language from a computational perspective. Steve Abney, in his essay opening *The Balancing Act: Combining Symbolic and Statistical Approaches to Language* (Klavans and Resnik, eds., MIT Press, 1996), wrote: “In the space of the last ten years, statistical methods have gone from being virtually unknown in computational linguistics to being a fundamental given”. We are again caught up in revolutionary times. Except this time around, the revolution is hurtling forward even faster, and it is redefining not only the research we do, but what it means to do research in the first place, the role of research in society, and society itself. What does this mean for ACL – what defines us as a community, what is our role, and where do we go from here?

Bio: Philip Resnik is a Professor at the University of Maryland with joint appointments in the Department of Linguistics and the Institute for Advanced Computer Studies. In 2020 he was named an ACL Fellow for significant contributions to symbolic-statistical methods for natural language processing, multilinguality, and the interdisciplinary study of language. Philip’s most recent research has focused in three main areas. The first is computational social science, with an emphasis on qualitative analysis and connecting the signal available in people’s language use with underlying mental state. The second is the computational cognitive neuroscience of language, using computational modeling in connection with brain imaging to look at the role of context and predictive processing during online language comprehension. The third involves fundamental questions about how current AI models relate to human cognition and to human society. Outside academia, Philip’s industry experience includes research at Bolt Beranek and Newman and Sun Microsystems Laboratories, as well as an internship at IBM T.J. Watson Research Center, and in entrepreneurial life he has been a technical co-founder of CodeRyte (clinical NLP, acquired by 3M in 2012), an advisor to FiscalNote (machine learning and analytics for government relations, went public in 2022), and he currently serves as an advisor to Trustible (technology provider for responsible AI governance). Philip was an undergrad in CS at Harvard and earned his PhD in Computer and Information Science at the University of Pennsylvania.

Keynote Talk

How Can AI Help Us Manage AI Conferences (where Troubles Are Caused by AI)

Yue Zhang
Westlake University



Monday, July 6, 2026 – Time: 13:00 – Room: Underline (Virtual)

Abstract: ACL has become 30+ times larger than two decades ago, and we face issues such as overwhelming participants, outdated papers, and low quality review. The reason can largely be attributed to increasing research in LLM, and AI accelerating research. Can AI help us address challenges in academic publication and evaluation? How can we accommodate AI-driven research? Could the research and publication paradigm be improved? In this talk, I will try to give some thoughts, sharing some of our recent work on AI review automation, paper recommendation, and AI arXiv, before briefly discussing a bit of AI research automation.

Bio: Yue Zhang is a Professor at Westlake University, and a fellow of ACL (<https://frcchang.github.io>). His research interests include fundamental NLP and its machine learning algorithms, and his recent research focuses on LLM reasoning and AI scientist. His major contributions to the field include machine learning algorithms for structured prediction (e.g., parsing and IE), neural NLP models (i.e., lattice and graph LSTM), and generalization for NLP/LM (e.g., OOD and logical reasoning). He co-authored the Cambridge University Press book *Natural Language Processing – a Machine Learning Perspective*. He served as a PC co-chair for CCL 2020, EMNLP 2022, and LMG 2025, and a general chair for AACL 2026 and NLPC 2026. He was a co-chair for the test-of-time award committee of ACL 2024 and 2025. He serves as editor in chief for the LLM journal, action editor for TACL, and associate editor for ACM/IEEE TASLP, ACM TALLIP, and IEEE TBD.

Keynote Talk
**Explanation and Understanding: A Perspective from
Cognitive Science**

Tania Lombrozo
Princeton University



Monday, July 6, 2026 – Time: 16:45 – Room: Harbor Ballroom Level 2

Abstract: What makes some explanations more satisfying than others? Why are humans so driven to explain? In this talk I'll take a cognitive scientific approach to explanation and understanding, drawing on decades of work in psychology, education, and philosophy to argue for a function-first approach to explanations—one on which we begin by characterizing what it is that explaining achieves. I'll suggest that the process of explaining plays a crucial role in discovery and generalization, and that we can understand the structure and phenomenology of explanations by first appreciating how explaining drives learning and helps construct our understanding of the world. I'll also consider some potential implications for explanation and understanding in AI systems.

Bio: Tania Lombrozo is the Arthur W. Marks '19 Professor of Psychology at Princeton University, as well as an Associate of the Department of Philosophy and the University Center for Human Values, the director of the Program in Cognitive Science, and co-director of Natural and Artificial Minds, a research initiative within the AI Lab. She received her Ph.D. in Psychology from Harvard University in 2006 after receiving a B.S. in Symbolic Systems and a B.A. in Philosophy from Stanford University. Dr. Lombrozo's research aims to address foundational questions about cognition using the empirical tools of cognitive psychology and the conceptual tools of analytic philosophy. Her work focuses on explanation and understanding, conceptual representation, categorization, social cognition, causal reasoning, and folk epistemology. She is the recipient of numerous early-career awards including the Stanton Prize from the Society for Philosophy and Psychology, the Spence Award from the Association for Psychological Science, a CAREER award from the National Science Foundation, and a James S. McDonnell Foundation Scholar Award in Understanding Human Cognition. Her trade book on explanation, *Why We Ask Why: The Science of Explanation and the Human Drive to Understand*, will be released in October.

Table of Contents

<i>Punctuation-Steered Representation Fine-Tuning</i>	
Zheng Gong, Ying Sun, Ping Li, Yi Zheng and Zhefeng Wang	1
<i>Does Self-Consistency Improve the Recall of Encyclopedic Knowledge?</i>	
Sho Hoshino, Ukyo Honda and Peinan Zhang	9
<i>LaMI: Augmenting Large Language Models via Late Multi-Image Fusion</i>	
Guy Yariv, Idan Schwartz, Yossi Adi and Sagie Benaim	18
<i>Full-Duplex-Bench-v2: A Multi-Turn Evaluation Framework for Duplex Dialogue Systems with an Automated Examiner</i>	
Guan-Ting Lin, Shih-Yun Shan Kuan, Jiatong Shi, Kai-Wei Chang, Siddhant Arora, Shinji Watanabe and Hung-yi Lee	27
<i>Situated Embedding Models for Context-Aware Dense Retrieval</i>	
Junjie Wu, Jiangnan Li, Yuqing Li, Lemao Liu, Liyan Xu, Jiwei Li, Dit-Yan Yeung, Jie Zhou and Mo Yu	37
<i>Big AI is Accelerating the Metacrisis: What Can We Do?</i>	
Steven Bird	50
<i>From Factuality to Meta-Factivity: A Cognitive Blueprint for Trustworthy LLMs</i>	
Liu Daohuan, Xia Lun, Yuer Wang, Jiaoyang Su and Xuri Tang	62
<i>Attention Sinks Are Provably Necessary in Softmax Transformers: Evidence from Trigger-Conditional Tasks</i>	
Yuval Ran-Milo	70
<i>A Mechanistic Account of Attention Sinks in GPT-2: One Circuit, Broader Implications for Mitigation</i>	
Yuval Ran-Milo, Hila Ofek and Shahar Mendel	90
<i>Is a Document Educational or Just Wikipedia-Style? — Pitfalls of Classifier-Based Quality Filtering</i>	
Mateusz Klimaszewski and Piotr Andruszkiewicz	99
<i>On the Hidden Objective Biases of Group-based Reinforcement Learning</i>	
Aleksandar Fontana, Marco Simoni, Giulio Rossolini, Paolo Mori and Andrea Saracino	109
<i>StructMem: Structured Memory for Long-Horizon Behavior in LLMs</i>	
Buqiang Xu, Yijun Chen, Jizhan Fang, Ruobin Zhong, Yunzhi Yao, Yuqi Zhu, Lun Du and Shumin Deng	122
<i>Z3D: Zero-Shot 3D Visual Grounding from Images</i>	
Nikita Drozdov, Andrey Leshchko, Nikita Gavrilov, Anton Konushin, Danila Rukhovich and Maksim Kolodiaznyy	147
<i>Improving Retrieval-Augmented Generation without Taxonomy-based Error Categorization</i>	
Gongbo Zhang, Yifan Peng and Chunhua Weng	155
<i>Deep Kernel Fusion for Transformers</i>	
Zixi Zhang, Zhiwen Mo, Yiren Zhao and Robert D. Mullins	166
<i>Anchoring Depends on Confidence and Post-Training in Language Models</i>	
Hillary N. Owusu and Naomi H. Feldman	174

<i>LLMs Underperform Graph-Based Parsers on Supervised Relation Extraction for Complex Graphs</i> Paolo Gajo, Domenic Rosati, Hassan Sajjad and Alberto Barrón-Cedeño	181
<i>Evolutionary Strategies at Scale lead to Catastrophic Forgetting</i> Immanuel Abdi, Akshat Gupta, Micah Mok, Alex Lu, Nicholas Lee and Gopala Anumanchipalli	194
<i>Too Correct to Learn: Reinforcement Learning on Saturated Reasoning Data</i> Zhenwen Liang, Yujun Zhou, Sidi Lu, Xiangliang Zhang, Haitao Mi and Dong Yu	205
<i>Reliable Use of Lemmas via Eligibility Reasoning and Section-Aware Reinforcement Learning</i> Zhikun Xu, Xiaodong Yu, Ben Zhou, Jiang Liu, Jialian Wu, Ze Wang, Ximeng Sun, Hao Chen and Zicheng Liu	216
<i>Attention Under Attack: Analog Noise Effects and Mechanistic Vulnerabilities in Transformer Models</i> Mafizur Rahman and Lijun Qian	227
<i>ReproEvalCard: A Reporting Standard for Reproducible Evaluation of LLM Pipelines</i> Priyaranjan Pattanyak and Apoorv Bhatia	238
<i>Test-Time Reasoners Are Strategic Multiple-Choice Test-Takers</i> Nishant Balepur, Atrey Desai and Rachel Rudinger	250
<i>MARCH: Multi-Agent Radiology Clinical Hierarchy for CT Report Generation</i> Yi Lin, Yihao Ding, Yonghui Wu and Yifan Peng	273
<i>Prefix Parsing is Just Parsing</i> Clemente Pasti, Andreas Opedal, Timothy J. O’Donnell, Ryan Cotterell and Tim Vieira	286
<i>Privacy-preserving Prosody Representation Learning</i> Kevin Everson and Mari Ostendorf	310
<i>Enhancing Linguistic Competence of Language Models through Pre-training with Language Learning Tasks</i> Atsuki Yamaguchi, Maggie Mi and Nikolaos Aletras	316
<i>Rethinking Data Mixing from the Perspective of Large Language Models</i> Yuanjian Xu, Tianze Sun, Changwei Xu, XinLong Zhao, Jianing Hao, Ran Chen, Yang Liu, Ruijie Xu, Stephen Chen and Guang Zhang	337
<i>Translation or Recitation? Calibrating Evaluation Scores for Machine Translation of Extremely Low- Resource Languages</i> Danlu Chen, Ka Sing He, Jiahe Tian, Chenghao Xiao, Zhaofeng Wu, Taylor Berg-Kirkpatrick and Freda Shi	350
<i>CheckMIABench: Firm Foundations For Membership Inference Attacks on Language Models</i> Jeffrey George Wang, Jason Wang, Marvin Li and Seth Neel	364
<i>Luring as a Proxy: Evaluating Corpus Transferability for Cybergrooming Detection</i> Shiying Fan, Mareike Bassenge and Martin Steinebach	371
<i>From Narrow Unlearning to Emergent Misalignment in LLMs</i> Erum Mushtaq, Anil Ramakrishna, Satyapriya Krishna, Sattvik Sahai, Praseon Goyal, Kai-Wei Chang, Tao Zhang and Rahul Gupta	386
<i>Reliable Evaluation Protocol for Low-Precision Retrieval</i> Kisu Yang, Yoonna Jang, Hwanseok Jang, Kenneth Choi, Isabelle Augenstein and Heuseok Lim	396

<i>When More Words Say Less: Decoupling Length and Specificity in Image Description Evaluation</i> Rhea Kapur, Robert D. Hawkins and Elisa Kreiss	410
<i>Selective Span-Level Unlearning for Large Language Models</i> Chaewon Yoon, Dongjun Kim and Hyun-Je Song	423
<i>Calibrated? Not for Everyone: How Sexual Orientation and Religious Markers Distort LLM Accuracy and Confidence in Medical QA</i> Alberto Testoni and Iacer Calixto	432
<i>DIXITWORLD: Evaluating Multimodal Abductive Reasoning in Vision-Language Models with Multi-Agent Dixit Gameplay</i> Yunxiang MO, Tianshi Zheng, Qing Zong, Jiayu Liu, Baixuan Xu, Yauwai Yim, Chunkit Chan, Jiaxin Bai and Yangqiu Song	448
<i>UERLens: Understanding Event Relations in Large Language Models</i> Yong Guan, Zhiyuan Li and Shaoru Guo	463
<i>GOLEMcoref: A Multilingual Coreference Dataset of Fiction</i> Andreas Van Cranenburgh, Xiaoyan Yang, Alvanita, Cecilia Nicole Di Domenico, Maria Ferragud, Arianna Graciotti, Byungjun Kim, Seonyeong Park, Noa Visser Solissa, Xiaoyu Zhou and Federico Pianzola	472
<i>Language-Aware Token Boosting: LLM Language Confusion Reduction Without Tuning</i> Trapoom Ukarapol, Pakhapoom Sarapat and Nut Chukamphaeng	481
<i>Pref-CTRL: Preference Driven LLM Alignment using Representation Editing</i> Imranul Ashrafi, Inigo Jauregi Unanue and Massimo Piccardi	490
<i>Dark & Stormy: Modeling Humor in Sentences from the Bulwer-Lytton Fiction Contest</i> Venkata S Govindarajan and Laura Biester	501
<i>The Personalization Trap: How User Memory Alters Emotional Reasoning in LLMs</i> Xi Fang, Weijie Xu, Yuchong Zhang, Scott Nickleach, Stephanie Eckman and Chandan K. Reddy	511
<i>Neuro-Symbolic Agentic Reinforcement Learning for Long-Term Original Character Companionship and Interaction</i> Zhenhan Huang	530
<i>Taming Extreme Tokens: Covariance-Aware GRPO with Gaussian-Kernel Advantage Reweighting</i> Cheng Wang, Qin Liu, Wenxuan Zhou and Muhao Chen	540
<i>On the Rejection Criterion for Proxy-based Test-time Alignment</i> Ayoub Hammal, Pierre Zweigenbaum and Caio Corro	547
<i>Defense Against Knowledge Poisoning Attack on GraphRAG</i> Havva Alizadeh Noughabi, Fattane Zarrinkalam and Ali Dehghantanha	555
<i>PExA: Parallel Exploration Agent for Complex Text-to-SQL</i> Tanmay Parekh, Ella Hofmann-Coyle, Shuyi Wang, Sachith Sri Ram Kothur, Srivas Prasad and Yunmo Chen	564
<i>Skill-Aware Data Selection and Fine-Tuning for Data-Efficient Reasoning Distillation</i> Lechen Zhang, Yunxiang Zhang, Wei Hu and Lu Wang	595
<i>Experiments or Outcomes? Probing Scientific Feasibility in Large Language Models</i> Seyedali Mohammadi, Manas Gaur and Francis Ferraro	605

<i>CaBSALLM: Efficient Context-Aware Batch Annotation of Conversational Streams with Large Language Models</i>	
Mohammadsadeh Abolhasani, Reza Mousavi and Paul Jen-Hwa Hu	615
<i>Challenging the Explanation Based on Preceding Tokens: Discovering Transferable Non-Literal Biasing</i>	
Yuchen Huang, Junpeng Zhang and Quanshi Zhang	637
<i>One-step Nonautoregressive Natural Language Generation with Shortcut Flow Matching Models</i>	
Jędrzej Warczyński, Ondrej Dusek and Mateusz Lango	646
<i>SED-SFT: Selectively Encouraging Diversity in Supervised Fine-Tuning</i>	
Yijie Chen, Yijin Liu and Fandong Meng	656
<i>Frame-Semantic Knowledge Injection for Event-Level Inference in LLMs</i>	
Shahid Iqbal Rai, Danilo Croce and Roberto Basili	664
<i>Exploring Cross-Client Memorization of Training Data in Large Language Models for Federated Learning</i>	
Tinnakit Udsa, Can Udomcharoenchaikit, Patomporn Payoungkhamdee, Sarana Nutanong and Norrathep Rattanavipanon	679
<i>FL-MSCL: A Unified Figurative Language Detection Model Driven by Multi-Type Signals and Contrastive Learning</i>	
Lu Shijia, Fumiyo Fukumoto, Huang Xiaoxi and Yoshimi Suzuki	694
<i>Revisiting Evaluation of Question Answering Systems in Low-Resource Indic Languages: Bridging Human and Metric Alignment</i>	
Anuj Kumar, Satyadev Ahlawat, Yamuna Prasad and Virendra Singh	703
<i>Data-efficient Targeted Token-level Preference Optimization for LLM-based Text-to-Speech</i>	
Rikuto Kotoge and Yuichi Sasaki	719
<i>How Do Inpainting Artifacts Propagate to Language?</i>	
Pratham Yashwante, Davit Abrahamyan, Shresth Grover and Sukruth Rao	727
<i>LOGICAL-COMMONSENSEQA: A Benchmark for Logical Commonsense Reasoning</i>	
Obed Junias and Maria Leonor Pacheco	746
<i>BioHiCL: Hierarchical Multi-Label Contrastive Learning for Biomedical Retrieval with MeSH Labels</i>	
Mengfei Lan, Lecheng Zheng and Halil Kilicoglu	759
<i>Dialogue is the Plan: From Interface to Joint Action in Agentic AI</i>	
Mert Inan, Malihe Alikhani and Anthony Sicilia	770
<i>Late Code Chunking: A Code Chunking Strategy for Repository-Level Code Completion</i>	
Seungmin Oh and Eunseok Lee	780
<i>Temporal Leakage in Search-Engine Date-Filtered Web Retrieval: A Retrospective Forecasting Case Study</i>	
Ali El Lahib, Ying-Jieh Xia, Zehan Li, Yuxuan Wang and Xinyu Pi	787
<i>A Shared Geometry of Difficulty in Multilingual Language Models</i>	
Stefano Civelli, Pietro Bernardelle, Nicolò Brunello and Gianluca Demartini	796
<i>T*: Progressive Block Scaling for Masked Diffusion Language Models Through Trajectory Aware Reinforcement Learning</i>	
Hanchen Xia, Baoyou Chen, Yutang Ge, Guojiang Zhao and Siyu Zhu	808

<i>Learning from Emptiness: De-biasing Listwise Rerankers with Content-Agnostic Probability Calibration</i>	
Hang Lv, Hongchao Gu, Ruiqing Yang, Liangyue Li, Zulong Chen, Defu Lian, Hao Wang and Enhong Chen	817
<i>Decoupling Generalization and Adaptation in Meta-Learning for Large Language Models</i>	
Nitin Vetcha, Binqian Xu and Dianbo Liu	827
<i>What Makes AI Research Replicable? Executable Knowledge Graphs as Scientific Knowledge Representations</i>	
Yujie Luo, Zhuoyun Yu, Xuehai Wang, Yuqi Zhu, Ningyu Zhang, Lanning Wei, Lun Du, Da Zheng and Huajun Chen	841
<i>Chain-of-Thought Degrades Visual Spatial Reasoning Capabilities of Multimodal LLMs</i>	
Sai Srinivas Kancheti, Aditya Sanjiv Kanade, Vineeth N. Balasubramanian and Tanuja Ganu	862
<i>Reviving Iterative Refinement in Diffusion-based NER with an Initializer-Restorer Approach</i>	
Long Hai Trieu, Phí Minh Hieu and Makoto Miwa	877
<i>Protein-STORY: Semantic Text-Oriented Representation Yields biologically meaningful Protein embeddings</i>	
Nabil Ibtehaz and Daisuke Kihara	883
<i>Diving into the Decoding Space of Non-Autoregressive Models via Lexically Constrained Search</i>	
Chenyang Huang and Osmar Zaiane	898