IJCNLP-AACL 2026

The 15th International Joint Conference on Natural Language Processing and the 5th Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics

Proceedings of the Conference - Volume 2: Short Papers

November 6-10, 2026

The IJCNLP-AACL organizers gratefully acknowledge the support from the following sponsors.

Co-organizer



Platinum



Gold





Venue Supporter



©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 979-8-89176-440-8

Message from the General Chair

Welcome to AACL-IJCNLP 2026 in Hengqin, China!

This joint conference brings together the Asia-Pacific and international natural language processing communities to share research, exchange ideas, and build connections. We look forward to conversations that connect languages, disciplines, institutions, and different perspectives on the future of our field.

We hope that the conference will provide a welcoming setting for presenting your work, learning from colleagues, and meeting new collaborators. Whether you are joining us in Hengqin or participating online, thank you for contributing to this community.

We thank the organizing committee, reviewers, tutorial instructors, workshop organizers, invited speakers, volunteers, sponsors, and all participants for the work that makes this meeting possible. We wish you a rewarding conference and an enjoyable stay in Hengqin.

Yue Zhang, General Chair
Ruifeng Xu, Ayu Purwarianti, Tong Rong, and Jun Suzuki, Program Chairs

Message from the Local Organizing Chairs

We are delighted to welcome you to the 5th Asia-Pacific Chapter of the Association for Computational Linguistics and the 15th International Joint Conference on Natural Language Processing. It is our great pleasure to host the AACL-IJCNLP community in Hengqin, a dynamic area of the Greater Bay Area just across from Macau.

We sincerely thank the General Chair, Program Chairs, organizing committee members, reviewers, volunteers, sponsors, and all contributors for their dedication and support in making this conference possible. We are also grateful to all authors, speakers, and participants for bringing their work, ideas, and energy to AACL-IJCNLP 2026.

We hope your time at AACL-IJCNLP 2026 will be both productive and enjoyable. Beyond the technical program, we encourage you to experience the local culture and hospitality of Hengqin and the surrounding region, including nearby Macau. We wish you inspiring discussions, fruitful collaborations, and a memorable conference experience.

AACL-IJCNLP 2026 Local Organization Chairs

Derek F. Wong (University of Macau)
Haofen Wang (Tongji University)

Organizing Committee

General Chair

Yue Zhang, Westlake University, China

Program Chairs

Ruifeng Xu, Harbin Institute of Technology, Shenzhen, China
Ayu Purwarianti, Institut Teknologi Bandung, Indonesia
Tong Rong, Singapore Institute of Technology, Singapore
Jun Suzuki, Tohoku University, Japan

Local Organizing Chair

Derek F. Wong, University of Macau
Haofen Wang, Tongji University, China

Local Organizing Committees

Runzhe Zhang, University of Macau
Zhiwen Xie, Central China Normal University, China

Workshop Chair

Xuefeng Bai, Harbin Institute of Technology, Shenzhen, China
Ratish Puduppully, IT University of Copenhagen, Denmark

Tutorial Chair

Zhongqing Wang, Soochow University, China
Naoaki Okazaki, Institute of Science Tokyo, Japan

Demonstration Chair

Ningyu Zhang, Zhejiang University, China
Derry Wijaya, Monash University, Indonesia

Student Research Workshop Chair

Yuxin Cao, Fudan University, China
Shio Hong, Fudan University, China
Yuhang Zhou, Shanghai Innovation Institute, China
Wenhe SUn, Fudan University, China

Best Paper Award Chair

Rui Yan, Wuhan University, China

Publication Chair

Junxian He, Hong Kong University of Science and Technology
Thuy-Trang Vu, Monash University, Australia

Handbook Chair

Leiyang Cui, Tencent AI Lab, China

Publicity Chair

Aditya Joshi, The University of New South Wales, Australia
KyungTae Lim, Korea Advanced Institute of Science Technology, South Korea

Web and App Chair

Xinchi Chen, Fudan University, China
Piji Li, Nanjing University of Aeronautics and Astronautics, China
Shudong Liu, University of Macau

Finance Chair

Hou Pong Chan, Alibaba DAMO Academy

Registration Chair

Wenxuan Zhang, Singapore University of Technology and Design, Singapore

Ethics Chair

Yangqiu Song, Hong Kong University of Science and Technology
Ekaterina Vylomova, The University of Melbourne, Australia

Diversity Inclusion Chair

Mamoru Komachi, Hitotsubashi University, Japan
Emily Ohman, Waseda University, Japan

Student Volunteer Chair

Juntao Li, Soochow University, Japan

Sponsorship Chair

Yan Song, University of Science and Technology of China, China
Yuanhe Tian, Zhongguancun Institute of Artificial Intelligence, China

Keynote Talk

NLP for Mental Health: Risk, Promise, Responsibility

Iryna Gurevych
Technical University of Darmstadt
2026-11-06 09:30:00 – Room: **Muhe B+C**

Abstract: More than a billion people live with a mental health condition, yet care remains scarce and often a poor fit for the person receiving it — so people increasingly turn to unsupervised chatbots to fill the gap. Because language is the richest signal we have into someone's psychological state, natural language processing is the key to doing this better. In this talk, I argue that NLP can genuinely help — but solving it takes clinical grounding, real data, and responsible evaluation, not just larger models. Drawing on work from my group, I show how large language models can turn therapy transcripts into treatment-relevant structure, how clinically grounded synthetic data eases the field's privacy bottleneck, and why evaluation must reflect what clinicians actually care about — not just what's easy to benchmark. I close with an invitation to build mental health AI that is personalized, privacy-aware, and responsibly evaluated: tools that support clinicians, not replace them.

Bio: Iryna Gurevych is a full professor of computer science at the Technical University of Darmstadt, Germany, where she founded and heads the Ubiquitous Knowledge Processing (UKP) Lab. She is a pioneer of computational argumentation and cross-document text modeling, with landmark applications to fact verification and misinformation detection, and her research extends to large language models and interdisciplinary NLP for the social sciences and humanities. Prof. Gurevych served as President of the Association for Computational Linguistics (ACL) in 2023 and is an ACL Fellow, a member of the German National Academy of Sciences Leopoldina and Academia Europaea, and a recipient of the 2025 Royal Society Milner Award.

Keynote Talk

Beyond Self-Play: Towards Effective Self-Evolution in LLMs

Yulan He
King's College London
2026-11-06 14:00:00 – Room: **Muhe B+C**

Abstract: Can large language models keep improving through experience they generate themselves? Self-play and self-generated experience offer promising routes. But producing more synthetic data does not necessarily lead to sustained model capability growth. This talk argues that effective self-evolution depends on whether the learning pipeline continues to provide learnable information. Drawing on our recent work, I will examine several ways such information can arise, including generating informative self-play experience, learning directly from environment-generated experience, and actively seeking novel evidence rather than waiting for tasks to be provided. These studies suggest that effective self-improvement requires more than generating its own experience; it needs experience that contains useful information that the model can actually learn from.

Bio: Yulan He is a Professor in Natural Language Processing at King's College London and a Turing AI Fellow. Her research focuses on making Large Language Models (LLMs) more reliable, interpretable, and safe, particularly in reasoning and agentic AI, with applications spanning across education, health, and science. She has received several awards for her research. Her five-year Turing AI Fellowship project was awarded the Best Research Project (Research Excellence) at the RAi UK AI Robotics Awards 2026. She also received a Best Research Paper Award at the same awards ceremony. In addition, she is the recipient of a SWSA Ten-Year Award and a CIKM Test-of-Time Award.

Keynote Talk

Uncertainty, Belief, and Trust in Language Models

Tim Baldwin
MBZUAI
2026-11-07 09:00:00 – Room: **Muhe B+C**

Abstract: Language models and other AI technologies are being used in an ever-increasing range of applications, including being trialled for decision support or automated decision-making in sensitive applications. A critical component of any such usage is the capability for the model to output a confidence/uncertainty assessment associated with any output, or (statistical) belief underlying a particular action. In this talk, I will present recent work on uncertainty estimation and belief updates, with a particular focus on practical utility in improving model performance and efficiency, and enabling humans to perform tasks more effectively.

Bio: Tim Baldwin is Professor of Natural Language Processing at Mohamed bin Zayed University of Artificial Intelligence (MBZUAI), in addition to being a Melbourne Laureate Professor in the School of Computing and Information Systems at The University of Melbourne, and Co-founder and Chief Scientist of startups including LibrAI (focused on AI safety) and Examn AI (focused on dialogue-based candidate profiling). At MBZUAI he has also held positions including Provost, Associate Provost, and Foundation Chair of the Department of Natural Language Processing.
Tim completed a BSc(CS/Maths) and BA(Linguistics/Japanese) at The University of Melbourne in 1995, and an MEng(CS) and PhD(CS) at the Tokyo Institute of Technology in 1998 and 2001, respectively. He joined MBZUAI at the start of 2022, prior to which he was based at The University of Melbourne for 17 years. His research has been funded by organisations including the Australian Research Council, Google, Microsoft, Amazon, Xerox, ByteDance, SEEK, NTT, and Fujitsu. He is the author of over 500 peer-reviewed publications across diverse topics in natural language processing and machine learning, and the recipient of a number of awards at top conferences.

Keynote Talk

The Genie of the LLM: What Shall We Wish for the Future of NLP?

Nizar Habash
New York University Abu Dhabi
2026-11-08 09:00:00 – Room: **Muhe B+C**

Abstract: Large language models have given NLP extraordinary new capabilities. The genie is out of the lamp. The question is no longer only what it can do, but what we should ask it to do, how we know we got what we really wanted, and what it will do to us, our research, and our education.
I reflect on these questions through a multilingual journey from early work on English, Spanish, and Chinese to decades of Arabic NLP. Arabic provides a revealing case study: its morphology, dialects, scripts, and linguistic variation expose assumptions that broad multilingual experiments can overlook. Recent work illustrates both the remarkable capabilities of LLMs and the continuing value of other models, linguistic knowledge, targeted data, careful evaluation, and human expertise.
The future will not simply be a more powerful version of the present. As answers become easier to generate, good questions, deep understanding, and judgment become more valuable. Our challenge is not merely to build more powerful genies, but to choose our lamps wisely, and become better wish-makers.

Bio: Nizar Habash is a Professor of Computer Science at New York University Abu Dhabi (NYUAD). He is also the director of the Computational Approaches to Modeling Language (CAMeL) Lab. Professor Habash specializes in natural language processing and computational linguistics. He received his PhD in Computer Science from the University of Maryland College Park in 2003. His research includes extensive work on machine translation, morphological analysis, and computational modeling of Arabic and its dialects. Professor Habash has been a principal investigator or co-investigator on over 30 research grants. And he has over 300 publications including a book entitled "Introduction to Arabic Natural Language Processing". Professor Habash is one of the inaugural recipients of the King Salman Academy for Arabic Language Award (2022); he is the recipient of the Antonio Zampolli Prize (2024); and he was selected as a Fellow of the Association for Computational Linguistics (2025). His website is www.nizarhabash.com.

Table of Contents

An Empirical Study of On-Device Translation for Real-Time Live-Stream Chat on Mobile Devices
Jeiyoon Park, Leedaehwan, Changmin Yeo, Yongshin Han and Minseop Kim 1

ImmunoFOMO: Are Language Models missing what oncologists see?
Aman Sinha, Bogdan-Valentin Popescu, Xavier Coubez, Marianne Clausel and Mathieu Constant 11

Constructive Comment Extraction from Japanese YouTube Shorts: A Dataset and Empirical Study
Siyuan Feng, Luecheng Wang, Koyo Ogura, Yuma Iwamoto, Seinosuke Sato, Yuki Kometani, Masaki Itabashi, Shintaro Ozaki, Yusuke Sakai, Shogo Matsuno, Hidetaka Kamigaito, Katsuhiko Hayashi, Ryoma Obara, Sho Takase and Taro Watanabe 22

When Does Knowledge Distillation Hurt? Reliability-Aware Distillation for Low-Resource Language Summarization
Dipto Sumit, Ankan Kumar Roy Srizon, Sadia Khair Rodela, Atia Haque Asha, Mourchona Afrin, Niloy Farhan and Farig Sadeque 35

Generate, Verify, and Refine: Event Timelines for Robust Video-Grounded Dialogue Systems
Wiradee Imrattanatrai, Masaki Asada, Kimihiro Hasegawa, Susan Holm, Ken Fukuda and Teruko Mitamura 49

WikiOriginQA: Automated Construction of a QA Dataset with Knowledge Origins for Analyzing Cultural Bias
Kento Haneda, Yosuke Kishinami, Ryo Fujii, Keisuke Sakaguchi and Makoto Morishita 69

Evaluating Contextually Mediated Factual Recall in Multilingual Large Language Models
Yihong Liu, Bingyu Xiong and Hinrich Schuetze 78

MortalMATH: Evaluating the Conflict Between Reasoning Objectives and Emergency Contexts
Etienne Lanzeray, Stephane Meilliez, Malo Ruelle and Damien Sileo 88

Improving Japanese-English Patent Claim Translation via Reasoning and GRPO with Syntactic Reward
Shogo Tsujimoto, Katsuki Chousa, Masaaki Nagata and Ryohei Sasano 95

Fairness in Prompting: How Demographic Variations Affect Language Model Outputs
Ryandito Diandaru, Gleb Kuzmin, Muhammad Dehan Al Kautsar, Zeerak Talat, Xudong Han, Haonan Li, Tatsuki Kuribayashi, Artem Shelmanov, Timothy Baldwin and Fajri Koto 106

Breaking the Silence: A Dataset and Benchmark for Bangla Text-to-Gloss Translation
Sharif Mohammad Abdullah, Abhijit Paul, Shubhashis Roy Dipta, Zarif Masud, Shebuti Rayana and Ahmedul Kabir 127

VIDA: A Dataset for Visually Dependent Ambiguity in Multimodal Machine Translation
Jingheng Pan, Xintong Wang, Longyue Wang, Liang Ding, Weihua Luo and Chris Biemann . 137

Mitigating the Impact of Glitch Tokens via Targeted Retokenization
Muhammad Alif Al Hakim, Dmitrii Kuzmin, Carlos Javier Galán Zavala, Tatsuya Hiraoka, Munachiso Samuel Nwadike, Ayana Niwa, Marco Cognetta, David Pohl and Kentaro Inui 155

A Multilingual Evaluation of LLM-Based High-School Mathematics Assistance Across Indian Languages
Yash Kumar and Subhajit Roy 170

FahMai-TeleBench: A Thai Grounded Tool-Use Benchmark for the Deployable Model Tier
Nutchanon Yongsatianchot, Piyalitt Ittichaiwong, Jitkapat Sawatphol and Tichanat Ruangsiri 185

Composing Within Constraints: A Benchmark for Thai Classical Poetry in Large Language Models
Nutchanon Yongsatianchot and Phuvis Kerdpramote 197

ThaiEditBench: A Mechanism-Typed, Length-Stratified Benchmark for Thai Orthographic Error Correction
Nutchanon Yongsatianchot, Piyalitt Ittichaiwong and Kanyakorn Veerakanjana 211

Signatures of Steerability in Activation Space of Language Models
Prajjwal Bhattarai and Tuka Alhanai 226

DIAM: Discourse Interaction and Addressee Modeling for Malaysia Hansard
Goh Kun Shun, Mohan Raj, Lay-Ki Soon, Wern Han Lim and Ian K. T. Tan 240

Is Brevity the Soul of Wit? An Empirical Test Using 570,000 Reddit Jokes
Robin Young 254

When Does Concept Erasure Change Behavior?
Gagan Aryan 262

Progress Drift: Auditing False Progress in Long-Horizon LLM Agents
Xiangbo Zhang and Xiaoxu Ma 281

Modeling Iterative Text Revisions Through Chain of Revisions
Zhexiong Liu and Diane Litman 295

Penalty-Framed No-Valid-Option MCQA: Analyzing LLM Abstention under Invalid Choices
Jinhyeok Kim and Hye-Young Jung 308

mSFT: Addressing Dataset Mixtures Overfitting Heterogeneously in Multi-task SFT
Woosung Koh, Jeyoung Jeon, Youngjin Song, Yujin Cheon, Soowon Oh, Jaehyeong Choi and Se-Young Yun 327

YouRA: A Persistent-State Architecture for Evidence-Traceable Autonomous Research Agents
Yoonkyu Woo, Woojin Lee and Jin-Xia Huang 355

TRIAGE: Severity-Ranked Multi-Agent Failure Attribution
Xizhi Wang 376

A Reliability Audit of Aggregated MTEB Scores: Coverage Asymmetry and Language-Specific Rank Inversions Across 17 Languages
Archit Rastogi 398

Training-Free Residual-Guided Retrieval for Multi-hop KGQA with Single-Step LLM Inference
Keito Sei, Sunil Kumar Maurya, Yohei Kobashi and Yutaka Matsuo 430

SEAL-PR: Scientific Evidence-Audited and Locally Isolated Peer Review under Untrusted Manuscripts
Zida Yang, Guquan You, Weiqing Zeng and Yu Zheng 443

You Don't Understand: Benchmarking Experiential Comprehension Gaps in LLMs
Yuyang Dai 463

Auditing PEFT Claims: A Stability Probe Unifying Three LoRA-Variant Fragility Patterns
Suan Lee and JaeSeong Kim 474

ILRR: Inference-Time Steering Method for Masked Diffusion Language Models
Eden Avrahami and Eliya Nachmani 495

The Missing I Don't Know": Why Three Reasoning-Reliability Findings Converge on Calibrated Abstention
Srijith Ravikumar 513

Probing for Negation: Extracting Compositional Negation Signals from Pretrained Language Representations for Image Retrieval
Animesh Singh and Brejesh Lall 525

RL-MR: Reinforcement Learning based Budget-Aware Sequential Pretrained Model Ranking
Saddam Mukta, Md. Ahashan Habib, Mir Tafseer Nayeem and Md Adnanul Islam 540

CogMem: A Cognitive Memory Architecture for Sustained Multi-Turn Reasoning in Large Language Models
Yiran Zhang, Jincheng Hu, Mark Dras and Usman Naseem 552

Counting Clues: A Lightweight Probabilistic Baseline Can Match an LLM
Furong Jia, Yuan Pu, Finn Guo and Monica Agrawal 563

Dating the Model: Hidden Dates in System Prompts Affect LLM Evaluation
Mario Sanz-Guerrero, Minh Duc Bui, Manuel Mager and Katharina von der Wense 585

Efficient Task Adaptation in Large Language Models: A Survey of Weight-Based, Prompt-Based, and Embedding-Based Adaptations
Jungwon Park, Changin Choi, Jimyeong Kim, Nojun Kwak and Wonjong Rhee 600

K-Prism: A Korean Diagnostic Benchmark for Multi-faceted Hallucination in Vision-Language Models
Min Hyuk Kim, Dong Hyok Shin, Jimi Lee and Jin Yea Jang 632

Mitigating Cross-Image Information Leakage in Multi-Image Understanding with Large Vision-Language Models
Yeji Park, Minyoung Lee, Sanghyuk Chun and Junsuk Choe 645

Closing the Bilingual Gap: Cross-lingual Semantic Scaffolding for Fine-Grained Machine Translation Evaluation in Taiwanese Hokkien
You-Wei Wu, En-Shiun Annie Lee and Richard Tzong-Han Tsai 664

Contextual Cue Sensitivity in LLM Evaluation of Grant Proposals: A Controlled ERC-Style Study
Marco Rospocher 679

Language Triggers Hijack Language Circuits: A Mechanistic Analysis of Backdoor Behaviors in Large Language Models
Théo Lasnier, Wissam Antoun, Francis Kulumba, Benoît Sagot and Djamé Seddah 699

Evaluating Large Reasoning Models on Belief Revision
Nurul Fajrin Ariyani, Zied Bouraoui, Richard Booth and Steven Schockaert 717

Target-Conditioned Gating for Many-to-Many Argument Assignment in Document-Level Event Extraction
Junru Liao, Haoran Wang, Xuanjing Huang and Yatian Shen 731

The Guardrail Trinity: Procedural Safety Schemas, Reasoning Chain Decomposition, and Dual-Mode Inference for LLM Guardrails
Huy H. Nguyen, Pride Kavumba, Jiaxuan Li, Duc Minh Vo and Koki Wataoka 747

Influcoder: Distilling Decoders' Gradient Influence Rankings into an Encoder for Data Attribution
Dimitri Kachler, Damien Sileo and Pascal Denis 780

Lay-Language and Cross-Lingual Disparities in LLM Medical Diagnosis: A Case Study in Rheumatological Skin Conditions
Grayson Hubbell, Yifei Zhang, Saksham Khatwani, Sriram Sankaranarayanan, Indira Sriram and Maria Leonor Pacheco 789

Not All Verifiable Preference Pairs Are Equal: Support-Aware Filtering for Stable Iterative DPO
Viet Anh Ngo, Long Tri Thai Son, Nguyen Viet Duc, Dong Minh Duc, Quang Minh Tran, Nguyen Thi Thuy Linh and Nguyen Viet Ha 807

Learning to Elicit LLM Bias in Multi-Turn Conversations
Da Gao and Ziyu Chen 816

Toward LLMs Beyond English-Centric Development
Sho Takase and Ukyo Honda 830

Transferable Structural Priors in Continuous Diffusion Language Models
Hao Li, Man Fung Zhuo, Xinqi Liu, Jiayu Wu and Yifan Hong 841

Auditing Bias Detection Under Source Shift: A Source-Aware Benchmark and Evaluation Protocol
Ana Pinto and Alvaro Figueira 854

What Improves During Pretraining? Tracking Knowledge Acquisition with Point-in-Time Models
Chang Su, Bin Wu, Emine Yilmaz and Yao Lu 879

Breaking Bureaucracy: Evaluating open-source LLMs for legal document review
Farrukh Baratov, Niki van Stein and Suzan Verberne 892

BG-RAG: Concept-Mediated Bipartite Graphs for Scalable RAG
Hang Gao, Yangmin Ding, Shaobo Han and Zhuocheng Jiang 917

Topo-Narrator: Heterogeneous Topological Memory for Long-form Narrative Generation
Kai Chen, Haoran Zhang, Tian Lan, Zehua Duo, Jiang Li, Guanglai Gao and Xiangdong Su . 923

More Modalities, More Problems: When Multimodal Agentic Debate Hurts Time-Series Reasoning
Patara Trirat 945