ALTA 2024

Proceedings of the 22nd Annual Workshop of the Australasian Language Technology Association

December 2-4, 2024
Australian National University
Canberra, Australia

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

Platinum

Australian Government
Defence

Gold

Google

Silver

ARDC
Australian Research Data Commons

Bronze

Commonwealth Bank

POSTERA CRESCAM LAUDE
THE UNIVERSITY OF
MELBOURNE

unsloth

Host sponsors

Australian
National
University

ANU
SCHOOL OF
CYBERNETICS

ANU
CULTURE
HISTORY
LANGUAGE

ANU
LITERATURE
LANGUAGES
LINGUISTICS

Introduction

Welcome to the **22nd Annual Workshop of the Australasian Language Technology Association (ALTA 2024)**. Hosted on the Acton campus of the Australian National University in Canberra, ALTA 2024 will provide a platform for the exchange of ideas, exploration of innovations, and discussion of the latest advancements in language technology. The conference acknowledges the significance of its location on the *traditional lands of the Ngunnawal and Ngambri peoples*, underscoring a commitment to inclusivity and respect.

ALTA 2024 convenes leading researchers, industry experts, and practitioners in the fields of natural language processing (NLP) and computational linguistics. This year, ALTA will focus on the critical role of large language models (LLMs) in shaping contemporary research and industrial applications.

ALTA has seen a remarkable growth in 2024. We received 43 submissions, a 1.79 times increase from the 24 submissions in 2023. This trajectory aligns with trends observed in global NLP research communities such as ACL and EMNLP. Following a rigorous and competitive review process, 21 submissions were accepted, comprising 10 long papers, 6 short papers, and 5 abstracts (not included in proceedings). The acceptance rate for papers included in the proceedings is 37.21% (16/43), reflecting a more selective process compared to 2023's 66.67% acceptance rate (16 out of 24 papers). We are also delighted to observe an increase in international participation. Of the accepted submissions, 85.71% (18 submissions) originate from Australia, 9.52% (2) from the USA, and 4.76% (1) from Malaysia.

This year's submissions showcase advancements across a wide array of topics. From educational applications such as personalised tutoring systems to healthcare-focused advancements like dementia self-disclosure detection and synthetic clinical text generation, the accepted papers demonstrate the versatility of NLP technologies. There is an evident focus on low-resource language processing, multilingual NLP, and domain-specific applications, with papers exploring practical solutions for real-world problems such as hate speech detection and legal document processing. A clear emphasis is given to bridging the gap between research and application. The focus on small-scale LLMs resonates with the community's efforts to develop resource-efficient and accessible AI systems.

We want to sincerely thank everyone who helped make ALTA 2024 a reality. A special thank you to our keynote speakers for fantastic presentations: Prof. Eduard Hovy (University of Melbourne), Prof. Jing Jiang (Australian National University), Prof. Steven Bird (Charles Darwin University), and Kyla Quinn (Australian Department of Defence). Thank you to the members of the discussion panel for an insightful conversation: Kyla Quinn, Prof. Hanna Suominen (Australian National University), and Luiz Pizzato (Commonwealth Bank). Thank you to the members of organising committee and volunteers for their hard work in preparing and running ALTA. We extend our heartfelt appreciation to the reviewers: your diligence and insightful feedback played an integral role in upholding the quality and rigor of the review process. Lastly, ALTA 2024 gratefully acknowledges the support of our sponsors: Defence Science and Technology Group (Platinum), Google (Gold), ARDC (Silver), and Commonwealth Bank, University of Melbourne, and Unsloth AI (Bronze). We are also proud to have The Australian National University as our host. The success of this workshop would not be possible without your invaluable contributions.

Welcome to ANU and Canberra! We hope that you enjoy ALTA 2024, and look forward to a rewarding and inspiring time together.

Tim Baldwin
Sergio José Rodríguez Méndez
Nicholas I-Hsien Kuo
ALTA 2024 Program Chairs

Organizing Committee

General Chair

Gabriela Ferraro, Australian National University

Program Chairs

Tim Baldwin, Mohamed bin Zayed University of Artificial Intelligence; University of Melbourne
Sergio José Rodríguez Méndez, Australian National University
Nicholas Kuo, University of New South Wales

Publication Chair

Anton Malko, Australian National University

Technology Chair

Dawei Chen, Australian National University

Finance Chair

Shunichi Ishihara, Australian National University

Sponsorship Chair

Charbel El-Khaissi, Australian National University

Local Chairs

Ned Cooper, Australian National University
Anton Malko, Australian National University

Publicity Chair

Kathy Reid, Australian National University

Program Committee

Area Chairs

Karin Verspoor, Royal Melbourne Institute of Technology
Mark Dras, Macquarie University
Sarvnaz Karimi, CSIRO

Reviewers

Massimo Piccardi, University of Technology Sydney
Sergio José Rodríguez Méndez, Australian National University
Nicholas I-Hsien Kuo, University of New South Wales
Gabriela Ferraro, Australian National University
Dawei Chen, Australian National University
Sarvnaz Karimi, CSIRO
Anudeex Shetty, University of Melbourne
Antonio Jimeno Yepes, Royal Melbourne Institute of Technology
Mark Dras, Macquarie University
Inigo Jauregi Unanue, University of Technology Sydney
Karin Verspoor, Royal Melbourne Institute of Technology
Jonathan K. Kummerfeld, University of Sydney
Jing Jiang, Australian National University
Mike Conway, University of Utah
Kemal Kurniawan, University of Melbourne
Hanna Suominen, Australian National University
Daniel Beck, Royal Melbourne Institute of Technology
Xiang Dai, CSIRO
Fajri Koto, Mohamed bin Zayed University of Artificial Intelligence
Diego Mollá, Macquarie University
Gisela Vallejo, University of Melbourne
Anushka Vidanage, Australian National University
Meladel Mistica, University of Melbourne
Shunichi Ishihara, Australian National University
Ming-Bin Chen, University of Melbourne
Lin Tian, University of Technology Sydney
Rongxin Zhu, University of Melbourne
Ekaterina Vylomova, University of Melbourne
Rena Wei Gao, University of Melbourne
Jey Han Lau, University of Melbourne

Public lecture: Generative LLMs: what they are and where they are heading

Eduard Hovy
University of Melbourne

2024-12-02 17:30:00 – Room: **Innovation space, Birch building**

Abstract: Generative AI has unleashed hype and concern. But it is surprising how few people understand how simple it is at heart, and how some of its shortcomings spring from its essential nature and will remain hard to overcome. In this talk I briefly describe the essential process and explore the three principal directions of GenLLM research: making them usable, useful, and understandable.

Bio: Professor Eduard Hovy is Executive Director, Melbourne Connect - a dynamic collaboration between leading organisations and interdisciplinary institutions aimed at leveraging research and emerging technologies to address global challenge - and a Professor in the School of Computing & Information Sciences, University of Melbourne.

Keynote Talk: LLM Evaluation: Writing Styles, Role-playing, and Visual Comprehension

Jing Jiang
Australian National University
2024-12-03 09:00:00 – Room: **Innovation space, Birch building**

Abstract: Large language models (LLMs) have demonstrated exceptional abilities that extend beyond language understanding and generation. This underscores the need for a more comprehensive evaluation of LLMs that covers a broader spectrum of capabilities beyond traditional NLP tasks. In this talk, I will share some of our recent work on LLM evaluation, with a focus on LLMs' writing styles and role-playing capabilities, and the abilities of large vision-language models to combine and interpret visual and linguistic signals in complex scenarios.

Bio: Jing Jiang is a Professor in the School of Computing at the Australian National University. Previously she was a Professor and Director of the AI & Data Science Cluster in the School of Computing and Information Systems at the Singapore Management University. Her research interests include natural language processing, text mining, and machine learning. She has received two test-of-time awards for her work on social media analysis, and she was named Singapore's 100 Women in Tech in 2021. She holds a PhD degree in Computer Science from the University of Illinois Urbana-Champaign.

Keynote Talk: Language Technology and the Metacrisis

Steven Bird
Charles Darwin University

2024-12-03 12:00:00 – Room: **Innovation space, Birch building (via Zoom)**

Abstract: Despite their manifold benefits, language technologies are contributing to several unfolding crises. Small screens deliver mainstream content across the world and entice children of minoritised communities away from their ancestral languages. The data centres that power large language models depend on the mining of ever more rare earth metals from indigenous lands and emit ever more carbon. Malicious actors flood social media with fake news, provoking extremism, division, and war. Common to these crises is content, i.e. language content, increasingly generated and accessed using language technologies. These developments – the language crisis, the environmental crisis, and the meaning crisis – compound each other in what is being referred to as the metacrisis. How are we to respond, then, as a community of practice who is actively developing still more language technologies? I believe that a good first step is to bring our awareness to the matter and to rethink what we are doing. We must be suspicious of purely technological solutions which may only exacerbate problems that were created by our use of technology. Instead, I argue that we should approach the problem as social and cultural. I will share stories from a small and highly multilingual indigenous society who understands language not as sequence data but as social practice, and who understands language resources not as annotated text and speech but as stories and knowledge practices of language owners. I will explore ramifications for our work in the space of language technologies, and propose a relational approach to language technology that avoids extractive processes and centres speech communities.

Bio: Over the past three decades, Steven Bird has been working with minoritised people groups in Africa, Melanesia, Amazonia, and Australia, and exploring how people keep their oral languages and cultures strong. He has held academic appointments at Edinburgh, UPenn, Berkeley, and Melbourne. Steven established the ACL Anthology, the Open Language Archives Community and the Natural Language Toolkit, and is past president of the Association for Computational Linguistics. Since 2017 he has been research professor at Charles Darwin University, where he collaborates with Indigenous leaders and directs the Top End Language Lab, http://language-lab.cdu.edu.au. Steven pursues other language-related projects at http://aikuma.org.

Keynote Talk: LLMs are great but ...

Kyla Quinn
Australian Department of Defence
2024-12-04 09:00:00 – Room: **Innovation space, Birch building**

Abstract: Knowledge workers are crying out for ways to industrialised the boring parts of their jobs, company executives are looking for ways to get a computer to replace all the humans and everyone thinks an LLM will solve all of their problems. But how do we ensure that we aren't creating a catastrophic failure when we deploy LLMs in situations where we can't afford to fail?
In this keynote, I will explore some of the issues we need to contend with when we put LLMs and other language technologies into an enterprise. I will touch on data preprocessing, governance, user trust and interpretation.

Bio: Kyla Quinn is the Technical Director of Data and Analytic Services Branch at the Australian Signals Directorate. In this role she provides strategic direction for staff involved in developing analytic tooling, from the AI and ML used in the back end through to user interfaces. Kyla has a background in engineering and linguistics and has recently submitted her PhD which is an evolutionary exploration of paradigm syncretism in the world's languages through Bayesian analysis and LLM embeddings.

Table of Contents

Shared Task (Not Peer Reviewed)

Tutorial (Not Peer Reviewed)

Program

Monday, December 2, 2024

13:00 - 14:00	*Tutorial Part 1*
14:00 - 14:30	*Afternoon Tea*
14:30 - 15:30	*Tutorial Part 2*
15:30 - 17:30	*Break*
17:30 - 18:30	*Public Lecture: “Generative LLMs: How they work and where they are headed” (Professor Eduard Hovy).*

Tuesday, December 3, 2024

08:45 - 09:00	*Opening*
09:00 - 10:00	*ALTA Keynote 1: “LLM Evaluation: Writing Styles, Role-playing, and Visual Comprehension” (Professor Jing Jiang).*
10:00 - 10:30	*Morning Tea*
10:30 - 11:00	*Minute Madness*
11:00 - 12:00	*Oral presentations, session 1: Education and Data Visualisation*
12:00 - 13:00	*ALTA Keynote 2: “Language Technology and the Metacrisis” (Professor Steven Bird).*
13:00 - 14:00	*Lunch*
14:00 - 15:00	*Oral presentations, session 2: Healthcare, Biomedical, and Legal Applications*
15:00 - 15:15	*Afternoon Tea*
15:15 - 16:15	*Panel Discussion [Panellists: Kyla Quinn, Professor Hanna Suominen, Luiz Pizzato]*
16:15 - 17:15	*Oral presentations, session 3: Multilingual NLP and Low-Resource Language Processing*
18:00 - 21:00	*Dinner at Badger & Co*

Wednesday, December 4, 2024

09:00 - 10:00	*ALTA Keynote 3: “LLMs are great but ...” (Kyla Quinn).*
10:00 - 10:30	*Morning Tea*
10:30 - 12:00	*Oral presentations, session 4: Advances in NLP Models and Techniques*
12:00 - 13:00	*Lunch*
13:00 - 14:30	*Oral presentations, session 5: Ethical Considerations and Social Media Analysis*
14:30 - 14:45	*Afternoon Tea*
14:45 - 15:30	*Oral presentations, session 6: Shared Task*
15:30 - 16:00	*ALTA AGM*
16:00 - 17:00	*Best Paper Award / Shared Task Award / Closing*