ALTESOL Language & Technology Research Group

2nd Annual

Artificial Intelligence Research in Applied Linguistics (AIRiAL) Conference

Teachers College
COLUMBIA UNIVERSITY

AI in Education: Empowering Learners & Preparing Educators

September 27-28, 2024

Plenary Speakers:
Yanis Ben Amor, Columbia University
Monica Arés, Imperial College Business School

Table of Contents

Welcome to AIRiAL 2024

Welcome to the second annual Artificial Intelligence Research in Applied Linguistics (AIRiAL) 2024 Conference at Teachers College, Columbia University in New York City. I am excited to see so many innovative and dedicated AI researchers gathered in-person to explore the intersections of artificial intelligence research and language education. This conference provides an opportunity to share cutting-edge research, exchange ideas, and foster collaboration to advance the field of applied linguistics.

This year's theme "AI in Education: Empowering Learners & Preparing Educators" highlights efforts to reflect on the skills that educators and learners need in order to benefit from AI technologies during education and in the workplace. We are honored to include two invited speakers to discuss this theme, Yanis Ben Amor, Executive Director of Center for Sustainable Development at the Climate School at Columbia University, and Monica Arés, Executive Director of Innovation, Digital Education, and Analytics (IDEA) Lab, Imperial College Business School.

AIRiAL 2024 was organized by members of the Applied Linguistics & TESOL Language & Technology Research Group. Thank you to our 2024 planning committee Shamini Shetye (chair), Dan Eskin, Xin Li, Kedi Mo, and Yilin Zhang, and for their dedication throughout the summer, for their many hours they have invested and for their passion and interest in this field of research.

In addition to the paper presentations, the program this year includes two sessions featuring posters and technology demonstrations and a colloquium. Presenters and attendees from 10 countries and 17 US states are gathering at Teachers College.

We extend our deepest gratitude to our sponsors, Language Testing International (LTI), Duolingo English Test (DET), Science Connected, British Council, and ESL Plus Educational Technologies and Consulting for their generous support for this event and for supporting artificial intelligence research in applied linguistics. I would also like to express appreciation to the Department of Arts & Humanities, the Applied Linguistics & TESOL program and the Digital Futures Institute at Teachers College, Columbia University for supporting AIRiAL 2024 and to President Bailey for opening the conference.

Thank you to our abstract reviewers: Emad Alghamdi, Jill Burstein, Ikkyu Choi, Sarah Chepkirui Creider, Larissa Goulart, Peter Kim, Stephanie Link, Sha Liu, Hoa Nguyen, Olena Rossi, Sowmya Vajjala, and Xiaoming Xi. Thank you to our student volunteers: Yuchen Fan**,** Xin Guan**,** Anthony Guzman, Ruotong Li, Zehao Li, Huiqi Meng, Joel Tillman, Biyin Xu, Weiting Yuan, and Yifan Zhang.

We hope you enjoy meeting colleagues, discussing research in artificial intelligence, and finding inspiration to build, design, and evaluate next generation models, applications, assessments, and policies. We hope to see you at AIRiAL 2025.

Erik Voss, Assistant Professor
Teachers College, Columbia University, New York City (2024)

Conference Organization

Organizing Committee

Erik Voss, Faculty Advisor
Shamini Shetye, Chair
Dan Eskin
Yilin Zhang
Kedi Mo
Xin Li

Abstract Reviewers

Emad Alghamdi
Jill Burstein
Ikkyu Choi
Sarah Chepkirui Creider
Larissa Goulart
Peter Kim
Stephanie Link
Sha Liu
Hoa Nguyen
Olena Rossi
Sowmya Vajjala
Xiaoming Xi

Best Student Paper Award Judges

Matthew Poehner
Ikkyu Choi
Dan Eskin

Student Volunteers

Yuchen Fan
Xin Guan
Anthony Guzman
Ruotong Li
Zehao Li (Photographer)
Huiqi Meng
Joel Tillman
Biyin Xu
Weiting Yuan
Yifan Zhang (Photographer)

Special Acknowledgements

Kiana Howerton, AL & TESOL Program Secretary

College Support

Applied Linguistics & TESOL Program
Department of Arts & Humanities Digital Future Institute (DFI)

PLATINUM SPONSOR

Language Testing International®
EXCLUSIVE LICENSEE OF ACTFL

GOLD SPONSOR

duolingo english test

SILVER SPONSORS

Science Connected

BRITISH COUNCIL

ESL PLUS
ESL Plus Educational Technologies and Consulting

Additional support provided by

Teachers College Columbia University AL & TESOL

Digital Futures Institute Teachers College Columbia University

Arts & Humanities

Plenary Speaker: Yanis Ben Amor

Closing the AI Gap between High-Income Countries and Emerging Markets and Developing Economies

Yanis Ben Amor

Columbia University

Abstract:
While AI technologies, particularly Large Languages Models such as ChatGPT, offer numerous opportunities for enhancing education, research, and innovation, their widespread adoption and effective utilization require substantial resources, infrastructure, and expertise. This discrepancy in resources between developed and developing countries could widen the gap between their respective universities, and their ability to use or develop these AI technologies.

This session will explore how the availability of computational resources, data, expertise and talent readiness, as well as financial resources all play a crucial role and will determine or hinder the capacity to train and deploy advanced AI models. Training AI models like ChatGPT requires significant computational power and storage capabilities. Universities in developed countries often possess powerful computing infrastructure and can afford high-performance hardware or cloud-based services for AI training. Developed countries, with their advanced technological infrastructure and access to abundant data sources, are also more likely to possess the necessary data for training sophisticated AI models. In addition, developed countries tend to have a more established ecosystem for AI research and a larger pool of skilled professionals. They often attract top talent, have well-funded research programs, and offer extensive opportunities for collaboration. Finally, developed countries, with greater financial resources, can allocate substantial funds to AI initiatives in universities, enabling them to pursue cutting-edge research, establish specialized AI centers, and offer competitive salaries to AI professionals. This is in stark contrast with the resources available in developing countries.

The session will then suggest solutions, particularly for developers when designing, developing or deploying AI models, to prevent a widening digital divide brought by AI between the Global South and the Global North.

Bio:
Professor Ben Amor has over 20 years of research experience developing digital tools. Professor Ben Amor led programs in Uganda and India to develop and allow access to Social and Emotional Learning curricula to children and adolescents. He has created mental health programs for school children in low-resource settings, and Digital Education tools to monitor children's school absenteeism in India using fingerprint technology.

Prof. Ben Amor is currently working on several projects involving Artificial Intelligence, as part of a digital application providing health access to vulnerable populations in Low-and-Middle-Income Countries (LMICs). Prof. Ben Amor has also launched a Columbia-wide initiative to study the impact of Artificial Intelligence on the Future of Work and Education, with a particular focus on LMICs.

Prof. Ben Amor has a PhD in Molecular Biology. He has published widely in infectious diseases, global health, and health education.

Plenary Speaker: Monica Arés

Learning Reimagined: Harnessing AI to Empower Educators and Learners

Monica Arés

Imperial College

Abstract:
As artificial intelligence continues to transform the educational landscape, we stand on the brink of revolutionary change. Imagine a world where AI provides personalized, adaptive learning experiences that cater to individual needs, enhancing engagement and deepening understanding. This session will explore the vast potential of AI in education, from intelligent tutoring systems and simulations to advanced language processing tools and multimodal learning platforms.

These technologies are not just new gateways to knowledge but keys to unlocking curiosity, creativity, connection, and innovation at scale. By merging human intellect with advanced AI technologies, we can create learning experiences that integrate our voices, languages, gestures, and gazes. At Imperial College London, we are pioneering these innovations, redefining traditional educational models through advanced learning ecosystems and simulations that enhance fluency and confidence in critical areas like problem-solving, communication, and leadership.

However, we must also confront the question: what if we don't embrace these advancements? Failure to integrate AI in education could mean perpetuating outdated learning models that fail to engage students, stifle creativity, and do not prepare learners for the challenges of the future.

Yet, there is immense hope and possibility. AI paves new pathways for lifelong learning, making education more accessible, dynamic, and personalized. By empowering learners and equipping educators with cutting-edge tools, we can inspire the next generation of global leaders and innovators. This session will also address how to bring pilot projects to scale and transform visionary ideas into actionable strategies, ensuring the practical implementation of AI-driven educational models.

Join me to envision a future where AI redefines education.

Bio:
Monica is a leading expert in immersive technology, using it to empower people to learn, connect and create in ways they never could before. She is deeply committed to using XR technology to increase access to education and fundamentally transform the way we learn.

As the former Head of Immersive Learning at Meta, and a pioneering leader on Interactive Learning Labs at Amazon, Monica is a visionary and leading figure in the field of immersive learning. She is dedicated to developing innovative products, programs, and content to make learning more powerful and accessible.

Conference Schedule: Friday

Time	AIRiAL 2024 Paper Presentations Day 1: Friday, September 27, 2024
8:00 - 9:00	Registration
9:00 - 9:15	Welcome and Opening Remarks ***Thomas R. Bailey**, President of Teachers College, Columbia University* ***Erik Voss**, Teachers College, Columbia University*
9:20 - 10:40 *Paper 1:* 9:20- 9:40 *Paper 2:* 9:40 - 10:00 *Paper 3:* 10:00 - 10:20 *Paper 4:* 10:20 - 10:40	**Session 1** **Exploring Learner Engagement with ChatGPT Feedback in Language Learning: A Case Study** *Jini Jung, Georgia State University* **Is ChatGPT Corrective Feedback Comprehensible to L2 Learners?** *Susanne Rott, University of Illinois, Chicago* **Investigating the Impact of AI on L2 Vocabulary Knowledge and Writing Performance of Turkish EFL Learners** *Merve Bako, Yildiz Technical University, Türkiye* *Ezgi Aydemir Altaş, Yildiz Technical University, Türkiye* **The Impact of AI-Enabled Personalized Recommendations on L2 Learners' Engagement, Motivation, and Learning Outcomes** *Babak Daneshvar Ghorbani, Iran University of Science and Technology, Iran* *Golnoush Haddadian, Georgia State University*

10:45 - 11:45 *Paper 1:* 10:45 - 11:05 *Paper 2:* 11:05 - 11:25 *Paper 3:* 11:25 - 11:45	**Session 2** **TBD** **Limits of Text-Based Prompting in Multimodal AI Models: Theoretical Implications for Art(s) Education** *Lauri Vakeva, University of Helsinki, Finland* **Facilitating Interactions in Less Commonly Taught Languages: A Case Study about Catalan and ChatGPT** *Mireia Toda Cosi, University of Maryland*
11:45 - 1:15	Lunch Break *(On your own)*
1:15 - 2:15	**Keynote Speaker** ***Yanis Ben Amor,*** *Columbia University* *Executive Director of the Center for Sustainable Development at the Climate School* **<u>Title:</u>** **Closing the AI Gap between High-Income Countries and Emerging Markets and Developing Economies** (Followed by Friday Group Picture)
2:15 - 3:15	**Poster & Technology Demonstration Session #1** (See presenters below)

3:15 - 4:35	**Session 3**
Paper 1: 3:15 - 3:35	**Evaluating AI-Generated Prompts in CILS DUE-B2 exams: A Comparative Validation Study** *Sabrina Machetti, University for Foreigners of Siena, Italy* *Giulia Peri, University for Foreigners of Siena, Italy* *Paola Masillo, University for Foreigners of Siena, Italy*
Paper 2: 3:35 - 3:55	**Multi-Stage Adaptive Design of an Oral Proficiency Assessment: BEST Plus 3.0** *Yage Leah Guo, Center for Applied Linguistics* *Francesca DiSilvio, Center for Applied Linguistics* *Wen Sun, Center for Applied Linguistics* *Rachel Myers, Center for Applied Linguistics* *Anna Zilberberg, Center for Applied Linguistics*
Paper 3: 3:55 - 4:15	**Automated Scoring of Student Generated Questions in Reading Comprehension Assessment** *Hajung Kim, University of Toronto, Canada* *Angelie Ignacio, University of Toronto, Canada* *Eunice Eunhee Jang, University of Toronto, Canada*
Paper 4: 4:15 - 4:35	**Using GPT Responsibly to Create Complex Reading and Listening Tasks** *Geoffrey LaFlair, Duolingo* *Andrew Runge, Duolingo* *Yigal Attali, Duolingo* *Sarah Goodwin, Duolingo* *Yena Park, Duolingo* *Phoebe Mulcaire, Duolingo* *Alina A. von Davier, Duolingo*
4:35 - 5:00	Coffee Break

5:00 - 6:20 *Paper 1:* 5:00 - 5:20 *Paper 2:* 5:20 - 5:40 *Paper 3:* 5:40 - 6:00	**Session 4** **A Systematic Review of Artificial Intelligence in Language Education from 2013 to 2023: Current Status and Future Implications** *Chaoran Wang, Colby College* *Meina Zhu, Wayne State University* **Empowering Korean English Teachers with ChatGPT: Insights from Professional Development Workshops on AI Integration** *Hyunjoo Moon, Ohio University* *Yesung Chung, Ohio University* *Greg Kessler, Ohio University* **GenAI in Language Teacher Education: A Trainer-Trainee Dialogue** *Amanda Brown, Syracuse University* *Jiaxing Lin, Syracuse University*
6:00 - 6:15	Announcements
6:30 - 8:00	Opening Reception (Wine & Cheese in Everett Lounge) Music provided by Chen Wang, Teachers College, Columbia University

AIRiAL 2024 Poster and Technology Demonstration Session #1

Friday, September 27, 2024
Time: 2:15 - 3:15 pm

POSTERS

Discursive Multidimensionality in Educational Management: Generative AI as an Insight Tool for Decision-Making in Colleges
Fernanda Peixoto Coelho, Pontifícia Universidade Católica de São Paulo, Brazil
Tatiana Schmitz de Almeida Lopes, Pontifícia Universidade Católica de São Paulo, Brazil
Adilson Gomes, Pontifícia Universidade Católica de São Paulo, Brazil

Integrating the Public Perspective into AI-Driven Language Curriculum and Pedagogy Development
Kayvan Shakoury, University of Western Ontario, Canada

Stance Classification in Applied Linguistics Using ChatGPT
Yaser Shamsi, Oklahoma State University

Leveraging Artificial Intelligence and Educational Data Mining to Address Challenges in Education for Students, Educators, Institutions, and Governments
Ecem Kopuz, City University of New York Graduate Center
Birkan Kopuz, Beykent University, Türkiye
Mustafa Kürşat Halat, Boğaziçi University, Türkiye
Ege Yıldırım, University of San Francisco

TECHNOLOGY DEMONSTRATIONS

Optimize My Resume: An AI-Driven Online Tool for Job Search Optimization and Task-Based Language Teaching
Cicely Anne Rude, Teachers College, Columbia University
Michael Campbell Stone, Science Connected

VR meets AI: Unleashing Potentials in Language Acquisition and Communication Skills Development
Minh Le, Teachers College, Columbia University

Adaptive Immersive Language Learning with Socrat.ai
Nina Bamberg, Pedagogy.Cloud

Exploring Generative AI for the ESL/EFL Classroom: Twee.com
Sharon Chan, Teachers College, Columbia University

Conference Schedule: Saturday

Time	AIRiAL 2024 Paper Presentations Day 2: Saturday, September 28, 2024
8:00 - 9:00	Registration
9:00 - 9:05	Opening Remarks
9:05 - 10:25	**Session 5**
Paper 1: 9:05 - 9:25	**Examining Language Model Inferences for Applied Linguistics Research: A Case Study with Two-Turn Dialogues** *Ikkyu Choi, Educational Testing Service* *Jiyun Zu, Educational Testing Service* *Saerhim Oh, Educational Testing Service*
Paper 2: 9:25 - 9:45	**Applied Chatbot-Enhanced Education Design (ACED) Cycle: Principles for Chatbot-Enhanced Language Learning in the Generative AI Era** *Burak Senel, Iowa State University* *Ananda Senel, Iowa State University*
Paper 3: 9:45 - 10:05	**Spoken Dialogue Technology Versus ChatGPT: Benefits and Challenges for Practicing and Assessing Oral Interaction** *Veronika Timpe-Laughlin, Educational Testing Service* *Tetyana Sydorenko, Portland State University* *Judit Dombi, University of Pécs, Hungary* *Saerhim Oh, Educational Testing Service* *Rahul Divekar, Bentley University*
Paper 4: 10:05 - 10:25	**The Impact of Generative AI-Powered Chatbots on L2 Comprehensibility** *Sinem Sonsaat-Hegelheimer, Iowa State University* *Şebnem Kurt, Iowa State University*

10:30-11:30	**Keynote Speaker** ***Monica Arés,*** *Imperial College Business School* *Executive Director of Innovation, Digital Education, and Analytics (IDEA) Lab* **Title:** **Learning Reimagined:** **Harnessing AI to Empower Educators and Learners** (Followed by Saturday Group Picture)
11:30 - 1:00	Lunch Break *(On your own)*
1:00 - 2:00	**Poster & Technology Demonstration Session #2** (See presenters below)
2:00 - 3:20 *Paper 1:* 2:00 - 2:20 *Paper 2:* 2:20 - 2:40 *Paper 3:* 2:40 - 3:00 *Paper 4:* 3:00 - 3: 20	**Session 6** **Unveiling ChatGPT's Writing Style: A Stylometric Analysis and AI Detection Study** *Peter Kim, Cambridge Boxhill Language Assessment* **First Steps in Using Generative AI to Enhance Genre-Based Automated Rhetorical Feedback: Evaluating GPT-3.5 and GPT-4 for Move Analysis** *Burak Senel, Iowa State University* *Elena Cotos, Iowa State University* **Interacting with ChatGPT: Exploring Engagement and Feedback in ESL Writing** *Inyoung Na, Iowa State University* *Mahdi Duris, Iowa State University* *Volker Hegelheimer, Iowa State University* **AI as Instructional Text Writer: A Multi-Dimensional Comparison with Human Texts** *Tony Berber Sardinha, Pontifical Catholic University of Sao Paolo, Brazil* *Marilisa Shimazumi, Cultura Inglesa College, Brazil*
3:20 - 3:50	Coffee Break

3:50 - 4:50	**Colloquium Session** **AI Supporting Learning, Teaching, and Assessing across Languages and Modalities: Recent Developments and Insights from Finland** *Ari Huhta, University of Jyväskylä, Finland* *Matthew Poehner, The Pennsylvania State University* *Dmitri Leontjev, University of Jyväskylä, Finland* *Raili Hilden, University of Helsinki, Finland* *Anna von Zanzen, University of Helsinki, Finland* *Roman Yangarber, University of Helsinki, Finland*
4:50 - 6:10 *Paper 1:* 4:50 - 5:10 *Paper 2:* 5:10 - 5:30 *Paper 3:* 5:30 - 5:50 *Paper 4:* 5:50 - 6:10	**Session 7** **Leveraging ChatGPT for Developing Adaptive Personalized Reading Assessment among College ESL Students** *Hamidreza Moeiniasl, University of Toronto, Canada* **Identifying Fairness Issues in Automatically Generated Testing Content** *Kevin Stowe, Educational Testing Service* *Alyssa Francis, Educational Testing Service* **Whose ChatGPT?** **Unveiling Unintended Educational Inequalities Introduced by Large Language Models** *Renzhe Yu, Teachers College, Columbia University* *Zhen Xu, Teachers College, Columbia University* *Sky Wang, Teachers College, Columbia University* **Rethinking Language Education in the Age of Generative AI** *Zhongfeng Tian, Rutgers University, Newark* *Chaoran Wang, Colby College*
6:10 - 6:30	Closing Remarks
7:00 - 10:00	Banquet & Best Student Paper Award Announcement *Pisticci* *125 LaSalle St, New York, NY 10027*

AIRiAL 2024 Poster and Technology Demonstration Session #2

Saturday, September 28, 2024
Time: 1:00 - 2:00 pm

POSTERS

Improving the Digital Sovereignty of Teachers:
Development of a Training Concept for the Use of AI in Foreign Language Teaching
Andreas Grünewald, University of Bremen, Germany
Leona Droste, University of Bremen, Germany

Establishing Transparency Boundaries in AI Adoption:
Student Awareness and Ethical Integration
Krisztina Domjan, American University

AI in Education: Exploring Pre-Service Teachers' Attitudes, Readiness, and Concerns
Dini Arini, Washington State University
Jihee Im, Washington State University
Hadir Alderaan, Washington State University

Artificial Intelligence and Technologies for Education: Learning and Teaching Languages
Ana Emilia Fajardo Turbin, University of Brasilia, Brazil

Traditional vs. AI-Driven Translation Sources in Language Learning
Dale Courtney Jeffery, Fukui University of Technology, Japan
Bradford J. Lee, Fukui University of Technology, Japan

A New Perspective on International Chinese Education:
Integrating Traditional Standards with ChatGPT
Yan Li, The Ohio State University
Zhenyu Park, Yanji, China

TECHNOLOGY DEMONSTRATIONS

Pangea Chat
Will Crowell Jordan-Cooley, Pangea Chat

Enhancing Language Education through AI:
Predictive Analytics for Assessing Reading Text Levels
Kamaleddine Tabine, US Department of State
Michel Tinguiri, US Department of State

Enhancing Flipped Learning with Artificial Intelligence
Meryem Ezgi Yalcin, Yale University

AceIt MR: Intelligent Job Interview Simulator for Non-Native Speakers
Amany Alkhayat, Teachers College, Columbia University

Conference Abstracts

Paper Abstracts: Friday

Session 1: 9:20 - 10:40 am

Exploring Learner Engagement with ChatGPT Feedback in Language Learning: A Case Study

Jini Jung, Georgia State University

ChatGPT distinguishes itself from other AI tools with its unique capability to comprehend and generate human-like texts. While its outcome, a human-like text, may raise concerns, such as the risk of plagiarism, unethical utilization leading to academic dishonesty, and the potential for misinformation, the utilization of ChatGPT for synchronous written feedback holds significant potential for enhancing engagement and facilitating language learning. However, it is essential to note that ChatGPT lacks the autonomy to initiate the feedback process independently, as it is solely programmed to respond to given prompts. Therefore, cultivating digital literacy skills is essential to crafting meaningful prompts tailored to learners' proficiency levels and feedback requirements, thus stimulating productive feedback.

This study delves into the engagement levels of two language learners with differing proficiency levels and feedback requirements in ChatGPT feedback. The advanced learner exhibited active cognitive engagement, meticulously analyzing suggestions, and seeking clarification. In contrast, the lower proficiency learner engaged more behaviorally, investing time in navigating and internalizing feedback. Despite these differences, both learners demonstrated a sense of authorship in their outcomes.

These findings underscore the significance of critical digital literacy skills for learners to take an active role in the feedback process and for educators to design personalized prompts and training programs tailored to individual learner needs. This research advocates for a learner-centered approach in utilizing ChatGPT as a valuable tool for language learning. By doing so, educators can harness its potential while mitigating risks associated with its use, ensuring a productive learning experience for language learners.

Is ChatGPT Corrective Feedback Comprehensible to L2 Learners?

Susanne Rott, University of Illinois, Chicago

AI tools, such as ChatGPT, can provide opportunities for language learning but also lead to misuse, particularly to complete writing assignments. To minimize misuse, Poole and Polio (2023) promote the teaching of digital literacies, which means the effective use of AI tools as opposed to plagiarism. While automated writing evaluation (AWE) tools have significantly advanced most of them have been textbook-specific or fee-based. Likewise, these tools generally provide feedback on all grammar errors and do not allow focusing on a limited set of structures, as instructors do. Acknowledging that corrective feedback in the context of a production, reviewing, and revising model (e.g., Williams, 2012) guides learners to interact with the feedback (Ranalli et al., 2017), the current study investigated whether and how corrective feedback provided by ChatGPT can be effectively used by 63 learners of beginning and intermediate German to self-correct their essays. The study compared ChatGPT output (Pfau & Polio, 2023) with instructor feedback. Findings showed that ChatGPT can focus on one specific

structure and allow the creation of a scaffolded feedback process. In addition, ChatGPT identified a majority of grammatical errors but not all, but neither did instructors. Yet, ChatGPT did not perform equally well on all structures. Learners were able to use a majority of ChatGPT's feedback, yet some students verbalized frustration because they were not familiar with all of the grammatical terminology used by ChatGPT.

Investigating the Impact of AI on L2 Vocabulary Knowledge and Writing Performance of Turkish EFL Learners

Merve Bako, Yildiz Technical University, Türkiye
Ezgi Aydemir Altaş, Yildiz Technical University, Türkiye

One of the prevalent concerns among EFL students is retaining newly acquired words and recalling them during writing tasks. Responding to this challenge, EFL educators started to experiment with AI which emerges as a promising resource that engages learners actively in the learning process. Therefore, this study aims to explore the impact of an AI tool (Mizou) on the L2 vocabulary knowledge and writing performances of Turkish EFL learners. The participants of the study were forty students, aged 18-19, who were enrolled in the English preparatory program at a state university in Türkiye and they were randomly assigned to control and experimental groups. Quantitative data were collected via Vocabulary Knowledge Scale (VKS), adapted from Wesche and Paribakht, (1996), and the target vocabulary to be taught and tested included thirty words taken from the coursebook. The VKS was administered before and after the treatment, serving as the pre-test and post-test. During the four-week treatment, experimental group students learned the target words with an AI tool called "Mizou". A chatbot was prepared using the target vocabulary of each listening and students were asked to practice the vocabulary using it in the classroom. However, the instructor in the control group provided a worksheet to practice target vocabulary. Furthermore, at the end of the treatment, learners' writings were examined using a rubric to assess their use of target vocabulary. Finally, students' perceptions on the use of AI were collected through semi structured interviews. This study revealed significant implications regarding vocabulary teaching in language classrooms

The Impact of AI-Enabled Personalized Recommendations on L2 Learners' Engagement, Motivation, and Learning Outcomes

Babak Daneshvar Ghorbani, Iran University of Science and Technology, Iran
Golnoush Haddadian, Georgia State University

The incorporation of generative AI tools, including expansive language models and chatbots gives rise to significant ethical and pedagogical concerns. In two longitudinal studies, AI-enabled personalized recommendations were investigated for their effects on engagement, motivation, and learning outcomes. Study 1 used a quasi-experimental design with repeated measures over a six-week period with 50 intermediate students acquiring English as a second language (17 males, 33 females). Control group received feedback from human tutors, while experimental group received feedback from ChatGPT (GPT-4). Using generative AI techniques, the study aimed to identify potential individual differences variables between the two groups. Study 2 centered on examining the perspectives of a distinct cohort comprising 50 participants (27 males and 23 females) who received feedback from both ChatGPT and their tutors. This study compared learners' perceptions of AI-generated feedback with that of human-generated feedback. Data collected from questionnaires were used to determine the strengths and weaknesses of each feedback form. According to Study 1, the experimental group and control group did not differ statistically significantly in learning outcomes. The results of Study 2 demonstrated that learners preferred both AI-generated and human-generated feedback nearly

equally. Based on the findings, AI-generated feedback can be incorporated into L2 learning evaluation without adversely impacting learning outcomes. AI-generated feedback can be included in L2 learning assessment through a well-rounded strategy that incorporates AI and human input. We recommend a blended approach that combines AI and human feedback. This paper discusses the findings in detail, underscoring their relevance and implications.

Session 2: 10:45 - 11:45 am

Limits of Text-Based Prompting in Multimodal AI Models: Theoretical Implications for Art(s) Education

Lauri Vakeva, University of Helsinki, Finland

This paper explores the limits of text-based prompting in current multimodal Generative Artificial Intelligence (Gen AI) models, focusing on the pragmatics of performative language use in the context(s) of art(s) education. As Gen AI increasingly becomes involved in artistic communication, language use in these contexts evolves, offering new ways to transform content from one modality to another. Through examples, the paper discusses how multimodal Gen AI can serve as a tool for performing diverse meanings and identities in artistic communication across linguistic and non-linguistic domains. It also emphasizes the need to develop ethical AI literacy among art(s) teachers by proposing a general framework for AI pedagogy that helps student teachers understand the value of diversity in technologically mediated learning environments. This paper contributes to applied linguistics in the context of teacher education by underscoring the critical role of language in shaping and interpreting AI-generated content and highlighting ethical considerations related to questions of representation and inclusivity within and beyond the art(s) subjects.

Facilitating Interactions in Less Commonly Taught Languages: A Case Study about Catalan and ChatGPT

Mireia Toda Cosi, University of Maryland

Interaction plays a pivotal role in language learning (Long, 1996; Ellis, 2003). However, when learning a less commonly taught language (LCTL) in a non-immersive setting, opportunities for interaction are limited. While prior research has highlighted the potential of Artificial Intelligence (AI; Bibauw et al., 2023) for learning, we need to evaluate its usability and validity.

We tested ChatGPT as a conversational partner on: (1) ecological validity in its interactions and (2) its ability to effectively adjust the difficulty of interactions. Fifteen intermediate learners of Catalan with diverse backgrounds completed three 10-15 minute interactions with ChatGPT. These were structured as tasks: discussing a YouTube video, engaging in a polarizing debate, and making dinner arrangements with a difficult friend. All chats were reviewed by 3 native speakers (NSs) who reported on the naturalness and quality of the interactions. In addition, each task was randomly assigned to different difficulty manipulations that ChatGPT was not able to execute. Lastly, participants shared their demographics, motivation (Nagel, 2023), technology use and attitudes (Berghal et al., 2023), and completed a qualitative interview.

ChatGPT received positive assessments from learners and NSs, although it was flagged as lacking internal consistency within the conversation. In general, limited access to NSs made learners feel more positively about the interaction, while they also reported that the tendency of ChatGPT to use recasts allowed them to learn and incorporate new more complex structures. Lastly, the perceived quality of the interaction had an impact on the complexity of the language used by participants.

Session 3: 3:15 - 4:35 pm

Evaluating AI-Generated Prompts in CILS DUE-B2 exams: A Comparative Validation Study

Sabrina Machetti, University for Foreigners of Siena, Italy
Giulia Peri, University for Foreigners of Siena, Italy
Paola Masillo, University for Foreigners of Siena, Italy

This study presents a comparative validation analysis of generative AI (ChatGPT-4; OpenAI, 2023), in crafting prompts for the CILS (Certification of Italian as a Foreign Language) DUE-B2 written exams. The goal is to assess test-takers performance differences when responding to prompts from human writers compared to AI-generated ones, by analyzing variations in response quality and characteristics. The aim is to investigate how AI integration might shift the construct of writing in L2 Italian. The study will also lay the ground for the examination of the implications of AI-generated prompts on the validity and reliability of CILS exams, including an analysis of potential biases and ethical considerations.

Following Chapelle and Voss (2021)'s indications on argument-based validation for technology-mediated language assessment, we seek to shed light on the affordances and challenges of AI for the Italian S/FL assessment, a field where technology's role is still emerging. This will highlight the competencies required for language testers, learners and teachers in ethically using AI tools.

Considering the emergent nature of this research area, especially in the context of the Italian language, this study marks an advancement in comprehending and utilizing the capabilities of AI in the field. The presentation will discuss the research context, the methodology and the findings from the pilot administration conducted, promoting a discourse that will pave the way for more extensive future research.

Multi-Stage Adaptive Design of an Oral Proficiency Assessment: BEST Plus 3.0

Yage Leah Guo, Center for Applied Linguistics
Francesca DiSilvio, Center for Applied Linguistics
Wen Sun, Center for Applied Linguistics
Rachel Myers, Center for Applied Linguistics
Anna Zilberberg, Center for Applied Linguistics

Recent calls for increased rigor in adult education led to the release of revised National Reporting System (NRS) Educational Functioning Level (EFL) Descriptors for English as a Second Language (ESL). Consequently, adult ESL educational programs in the U.S. receiving federal funding are mandated to measure and report on the students' educational gains using a standardized assessment. One of the assessments approved for this purpose, BEST Plus 3.0, is aligned to measure all six NRS EFLs and features an innovative multi-stage computerized adaptive design.

BEST Plus 3.0 is an individually administered, adaptive, face-to-face oral interview measuring integrated listening and speaking skills of adult ESL learners in the U.S. It follows a conversational format and is scored live by a trained administrator using a scoring rubric. Multi-stage computer adaptive testing (CAT) is an AI-based method that uses IRT ability estimates in an automated iterative process to select a set of tasks tailored to each test-taker's

ability level. The multi-stage adaptive design has several advantages over a fixed-form assessment, including test time and task difficulty custom-tailored to each test taker. The multi-stage design is especially beneficial in the context of an oral interview because several tasks are fixed within each stage at each difficulty level and are presented in logical order thus preserving a conversational flow. This session describes the development and validation of both conceptual and psychometric models underpinning BEST Plus 3.0 and highlights the ways in which this adaptive technology enhances the human experience of both test takers and test administrators.

Automated Scoring of Student Generated Questions in Reading Comprehension Assessment

Hajung Kim, University of Toronto, Canada
Angelie Ignacio, University of Toronto, Canada
Eunice Eunhee Jang, University of Toronto, Canada

Students' active questioning, through student-generated questioning (SGQ), fosters critical thinking and engagement in learning (Chin & Brown, 2002). Automated scoring of SGQs can help overcome challenges in scoring it and further offer diagnostic feedback beyond text comprehension. This study examined automated machine learning algorithms to score young children's SGQs. The study engaged 265 students in Grades 3 to 8. Data was collected using BalanceAI, a web-based literacy assessment platform (Jang, 2023). Students read a passage and were prompted to write three questions. Four human raters evaluated 2,920 SGQs using a 6-point rubric (Maplethorpe et al., 2020). Cohen's Kappa ranged from 0.46 to 0.83, and were used to train ML models. Two different ML approaches were tested in the present study. The first feature-based approach combined several features, including TF-IDF, bigram, type-token ratio, topic words, word embedding for SGQs and the corresponding passage. The second approach, a word embedding model, included only word embedding for SGQs as features using OpenAI. We evaluated ML models in terms of accuracy, precision, recall, and F1 score. OpenAI's embedding model demonstrated comparable performance to human scoring, exhibiting a higher Cohen's Kappa but a lower level of agreement. The results show the potential to leverage future developments in pre-trained embeddings into language assessment practices. However, considering the model was built solely using word embeddings, establishing validity evidence using rubrics can be challenging. We call for further research on valid automated assessment of students' ability to generate high-quality questions beyond responses to predetermined questions.

Using GPT Responsibly to Create Complex Reading and Listening Tasks

Geoffrey LaFlair, Duolingo
Andrew Runge, Duolingo
Yigal Attali, Duolingo
Sarah Goodwin, Duolingo
Yena Park, Duolingo
Phoebe Mulcaire, Duolingo
Alina A. von Davier, Duolingo

ChatGPT has brought the capabilities of large-language models (LLMs) to the forefront of applied linguistics research. In the field of test development it has started conversations about how it can be used responsibly (Burstein, 2023) while also maintaining standards that uphold the validity of the interpretations and uses of test scores.

Burstein (2023) outlines four standards (Validity and Reliability, Fairness, Privacy and Security, and Transparency and Accountability) which can be used to guide the development of assessment developed with AI. Through this lens of responsible AI, we present work that leverages LLMs to develop complex reading and listening tasks for a high stakes English language test.

We describe how we ensure construct representation as we replenish our item bank by applying the concept of test specifications to our GPT-driven prompting process. We discuss our iterative approach to item development, which leans on a feedback loop between pilot results, item content, and format enhancements. Additionally, we describe our human-in-the-loop review process to ensure that GPT-generated items meet quality and content standards to provide a fair and high-quality experience for our diverse test taker population.

The availability of LLMs is a catalyst for a paradigm shift in how content and questions are developed for assessments of all types. However, maintaining standards of validity, reliability, and fairness remain paramount while folding this technology into existing test development processes is a necessary component of the RAI standards. This presentation will share how this can be accomplished in an operational setting.

Session 4: 5:00 - 6:20 pm

A Systematic Review of Artificial Intelligence in Language Education from 2013 to 2023: Current Status and Future Implications

Chaoran Wang, Colby College
Meina Zhu, Wayne State University

This presentation provides a systematic review of selected empirical studies on artificial intelligence (AI) in language education, spanning from 2013 to 2023. Data for this review were gathered from the Web of Science, Eric ProQuest, Scopus, and five top specialized language education journals. Following the principles delineated in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021), a total of 125 studies met the selection criteria and were analyzed using multiple methodologies, including bibliometric analysis, inductive content analysis, and topic modeling. This article furnishes an overview of the current landscape of AI in language education research, emphasizing specific AI technologies, their applications, and their educational impact. The most prevalent AI technologies encompass automated writing evaluation, bots, machine translation, automatic

speech recognition, and intelligent systems. The results also reveal frequent utilization of AI to assist students in learning writing and speaking. Extensive discussions about practical implications and outlines for future research directions are provided from multiple perspectives. The evolution of AI necessitates initiatives addressing diversity, equity, and inclusion (DEI) concerns in language education. Future research demands large-scale collaborative efforts with a focus on long-term research and development endeavors. This presentation offers insights into the trajectory of the research trends and practices within the field of AI-assisted language education, as well as highlights significant directions for future research to address existing gaps in the field.

Empowering Korean English Teachers with ChatGPT: Insights from Professional Development Workshops on AI Integration

Hyunjoo Moon, Ohio University
Yesung Chung, Ohio University
Greg Kessler, Ohio University

This study investigates the impact of professional development workshops and online learning communities on ChatGPT integration in Korean English classrooms. The research team held two workshops, including ChatGPT Integration in English Classrooms, targeting Korean English teachers. Quantitative surveys were distributed before and after the workshops to collect data on teacher perceptions, experiences, and proficiency with ChatGPT. The data was analyzed via descriptive statistics, correlation analysis to explore teachers' perceptions and experiences with ChatGPT, and paired-sample t-tests to assess the impact of the workshop. Before the workshop, teachers had a generally positive point of view on ChatGPT integration in education, however, their integration was limited to creating test questions and conducting research. After the workshop, teachers' perceptions of ChatGPT became more positive and their self-reported proficiency levels increased significantly. This finding shows that professional development workshops have a positive impact on teachers' perceptions and proficiency with ChatGPT. For those who wanted to have continued communication with the other teachers, they joined the online learning community, AIGURU. In AIGURU, teachers share prompts for creating teaching materials or learning activities. They also share their teaching experiences with ChatGPT in synchronous online meetings. The qualitative feedback about the workshop highlighted a demand for tangible examples of how ChatGPT can be integrated into classroom settings. Therefore, these online learning community activities provide an excellent opportunity for teachers to exchange insights and learn from the experiences of their peers. This collaborative environment allows for a richer understanding of the practical applications of AI tools in education.

GenAI in Language Teacher Education: A Trainer-Trainee Dialogue

Amanda Brown, Syracuse University
Jiaxing Lin, Syracuse University

The use of Generative AI (GenAI) in language teacher education is critical but underexplored, and teacher-educators report a lack of knowledge and confidence in preparing teachers-in-training for GenAI use (Moorhouse & Kohnke, 2024). This qualitative case study reports on efforts by a language teacher-educator and a teacher-in-training to include GenAI in their respective professional practices.

In informal polling, the teacher-educator found that 75% or more of her teachers-in-training in the 2203-24 academic year had limited to no experience of GenAI use in general and especially for professional purposes, with reasons expressed including fear of committing academic

integrity violations after consistent negative messaging across the curriculum and a lack of understanding of how to leverage GenAI effectively and ethically. The teacher-educator initially explored a laissez-faire approach, inviting teachers-in-training to experiment with GenAI and share results, with minimal uptake by teachers-in-training. Subsequent efforts involved positive messaging around effective and ethical use of GenAI, active encouragement of individual experimentation, an invited talk on the topic, and finally a requirement to use GenAI for collaborative in-class activities. A teacher-in-training documented his evolutionary GenAI journey, describing initial concerns about effectiveness and ethics. He reported being influenced by positive instructor messaging, and a transformation beginning with peer in-class sharing on professional GenAI use, followed by explicit demonstration during the invited talk, and then required personal use. In the latter third of the semester, he transformed into an active user and began educating peers. Implications of this qualitative case study for teacher education will be discussed.

Paper Abstracts: Saturday

Session 5: 9:05 - 10:25 am

Examining Language Model Inferences for Applied Linguistics Research: A Case Study with Two-Turn Dialogues

Ikkyu Choi, Educational Testing Services
Jiyun Zu, Educational Testing Services
Saerhim Oh, Educational Testing Services

Empirical investigations in applied linguistics often involve expert evaluation of language use, which takes time and is difficult to scale up. Recent language models have achieved human-like performances in multiple benchmark tasks (e.g., Min et al., 2023) and present a promising opportunity to address these challenges. However, it is not yet clearly understood how language models achieve such performance or how well performance on benchmark tasks generalize. It is thus essential to examine how and why language models make inferences for a specific task. In this talk, we present a case study illustrating the process of fine-tuning a language model for a custom evaluation task and probing its performance. The custom task involves the classification of two-turn dialogues, which constitute part of a listening comprehension assessment, into two classes: whether the second turn is a direct response of the first turn or not. We fine-tuned an open source, transformer-based encoder (Liu et al., 2019) for the classification task on a data set consisting of approximately 5000 dialogues and identified most impactful tokens in the classification utilizing multiple methods including saliency mapping (Wallace et al., 2019). The resulting classifier achieved 97% accuracy on a separate data set consisting of approximately 1400 dialogues, and the impactful tokens included a mixture of statistical signals potentially specific to our data set and meaningful linguistic signals likely to generalize. These findings highlight the importance of task-specific model training and examination to properly gauge the capacity and generalizability of a resulting model.

Applied Chatbot-Enhanced Education Design (ACED) Cycle: Principles for Chatbot-Enhanced Language Learning in the Generative AI Era

Burak Senel, Iowa State University
Ananda Senel, Iowa State University

The increasing presence of chatbots in language learning contexts, particularly with the advent of generative AI chatbots like ChatGPT, necessitates a principled, research-informed approach to their design and integration. However, the field of chatbot-enhanced language learning (CELL) currently lacks such a framework. This paper introduces the Applied Chatbot-enhanced Education Design (ACED) Cycle, a comprehensive, iterative framework for designing, selecting, and integrating chatbots in language learning contexts. Developed from a systematic review of 21 CELL studies, the ACED Cycle consists of six key phases: Pedagogical Considerations, Task Design/Selection, Chatbot Design/Selection, Chatbot Piloting, Chatbot Use, and Reflection and Evaluation. Each phase of the ACED Cycle is guided by a set of principles derived from the findings of the systematic review. For instance, in the Pedagogical Considerations phase, the principle "Justified Use" emphasizes the importance of outlining clear learning goals prior to chatbot design or selection. In the Task Design/Selection phase, the principle "Task Variation" advocates for diverse tasks and topics to maintain engagement and contextual relevance. The Chatbot Design/Selection phase includes principles such as "Traced Interactions", which focuses on selecting or designing chatbots that can record interactions for assessment. This presentation will cover the ACED Cycle framework, the principles associated with each phase,

and a hypothetical pedagogic scenario demonstrating how the ACED Cycle can be applied to integrate chatbots into language teaching effectively. The scenario will illustrate how teachers can use the ACED Cycle to design and integrate chatbots that align with specific learning goals, learner proficiency levels, and age groups.

Spoken Dialogue Technology Versus ChatGPT: Benefits and Challenges for Practicing and Assessing Oral Interaction

Veronika Timpe-Laughlin, Educational Testing Services
Tetyana Sydorenko, Portland State University
Judit Dombi, University of Pécs
Saerhim Oh, Educational Testing Services
Rahul Divekar, Bentley University

Interactional competence (IC) is crucial for L2 English learners to adeptly communicate across contexts. However, providing opportunities to practice interactive oral language can be time and resource-intensive. Spoken Dialogue Systems (SDS) can elicit evidence of linguistic skills similarly to human interactions, but often produce relatively transactional communication (Timpe-Laughlin et al., 2024) lacking the richness of human conversation. Large Language Models (LLMs) like ChatGPT, capable of generating human-like responses, might bridge this gap, bringing more flexibility to conversations. This study compared the effectiveness of SDS and ChatGPT in eliciting evidence of language ability and IC in role-play tasks. A task simulating a coffee shop purchase, including small talk, was conducted with 50 tertiary-level L2 English learners using both SDS and a voice-adapted ChatGPT in a counterbalanced design. Participants' interactions were audio-recorded and analyzed for linguistic and interactional features, and participants' perceptions were surveyed post-task. To compare indices and observations across modalities, we deployed descriptive statistics as well as paired samples t-tests with Bonferroni corrections. Additionally, we analyzed user perceptions by means of both quantitative and qualitative approaches. Findings showed that both systems were relatively equal in eliciting language that shows participants' linguistic abilities. Additionally, conversations with ChatGPT and the SDS appeared similarly transactional in that both showed little sequential unfolding of language functions across turns. However, ChatGPT and SDS dialogues varied considerably in terms of conversation length, repair, and topic management. We will discuss the affordances and limitations of both systems for practicing and assessing oral English skills.

Generative AI-Powered Pronunciation Practice for L2 English

Sinem Sonsaat-Hegelheimer, Iowa State University
Şebnem Kurt, Iowa State University

Computer assisted pronunciation teaching (CAPT) tools specialize in providing corrective feedback for L2 pronunciation by relying on automatic speech recognition (ASR) (Cucchiarini & Strik, 2018). However, these tools have constraints on meaning to optimize feedback accuracy (Hincks, 2015). Specifically, they have limitations when it comes to facilitating open-ended conversations. Advances in AI-based chatbots address these meaning constraints, yet chatbots typically do not provide feedback on language use, more specifically L2 pronunciation.

In this study, we investigate fifteen L2 English learners' use of two voice-enabled, AI-based chatbots: Google's Gemini (n= 6) and Pronounce (n = 9). Both chatbots allow users to interact with them by recording their voices. Gemini does not provide any explicit feedback for pronunciation of speakers other than the implicit feedback coming from the real-time transcriptions of speech. Pronounce (Pronounce Inc., 2024) provides feedback for pronunciation, grammar, and vocabulary.

By employing a mixed-methods design including quantitative and qualitative data, we address the following research questions: (i) What is the nature of human-chatbot interactions for the two groups of L2 learners? (ii) How do the learners work with the feedback they receive from the two chatbots? and (iii) What are the learners' perceptions about the use of the chatbots and their feedback? The data include screen recordings of interactions between the chatbots and L2 learners, chat logs of Gemini users, feedback reports of Pronounce users, a user experience survey, and a semi-structured interview. Findings of the study will provide insights on how to integrate AI-based chatbots in L2 pronunciation.

Session 6: 2:00 - 3:20 pm

Unveiling ChatGPT's Writing Style: A Stylometric Analysis and AI Detection Study

Peter Kim, Cambridge Boxhill Language Assessment

The AI revolution has led to an increasing reliance by language learners for writing assistance and, in some cases, generate text ostensibly as their own. Given this backdrop, the purpose of this study is to investigate ChatGPT's writing style using two approaches, 1) employing stop-words to conduct a stylometry analysis with Burrow's delta method. 2) In the second method, content words devoid of stop-words were used to detect AI generated text by using random forest classification (RFC). Prompt engineering was done to further revise GPT generated essay to avoid AI detection. The Uppsala Student English corpus (USE) served as the dataset, comprised of 385 writers and 1407 essays, including 538 essays generated by ChatGPT (GPT 3.5). After fitting the model, feature importance was extracted for further analysis for both methods. The first method primary uses stop-words for classification while the second method uses content words only. These two complementary methods were designed to shed light on the distinct features and styles of writing employed by ChatGPT. Comparison between Burrow's delta and RFC revealed that while stop-word-based classification recognized most instances of GPT-generated text, it also had a high false positive rate. Taken together with the results of RFC, this study concludes that ChatGPT tends to favor a distinct style of writing essays. This has implications for the "teaching" of expository, persuasive, and argumentative essays by LLMs: it is possible that as the use of LLMs become more pervasive, diversity of writing styles may be negatively impacted.

First Steps in Using Generative AI to Enhance Genre-Based Automated Rhetorical Feedback: Evaluating GPT-3.5 and GPT-4 for Move Analysis

Burak Senel, Iowa State University
Elena Cotos, Iowa State University

Advances in generative AI, particularly large language models (LLMs), offer promising opportunities for supporting academic writing instruction, as it has been attempted by genre-based automated writing evaluation (GB-AWE) tools (Cotos, 2023) that provide individualized feedback on 'moves' (communicative goals) and 'steps' (rhetorical strategies) (Swales, 1990). Developing such tools is challenging, if not daunting. It requires manual move/step analysis and annotation of a large corpus of texts, and then training high-performing machine learning models to identify specific moves/steps. LLMs may have the potential to generate rhetorical feedback in natural language similar to, or perhaps more efficiently, than existing GB-AWE tools. This hypothesis, however, calls for proof-of-concept research evaluating LMM performance on move/step analysis tasks. Our study investigated the ability of GPT-3.5 and GPT-4 to identify the Establishing-a-Territory move in research article Introductions under different prompting conditions. An expert-annotated corpus of Introductions from thirty disciplines served as the gold standard. Six prompt conditions provided varying amounts of genre knowledge: Prompt0–NoInformation, Prompt1–MoveDefinition,

Prompt2–MoveStepsDefinitions, Prompt3–MoveStepsExamples, Prompt4–MoveExpressions, and Prompt5–AllCombined. For each prompt, the models classified Introduction sentences across ten iterations at three randomness settings. Accuracy, precision, recall, F-1, and overall reliability metrics reveal that GPT-4 demonstrates more consistent performance across different prompts and randomness values, while GPT-3.5 outperforms GPT-4 on all prompts except Prompt1. Follow-up replication of these experiments on two more Introduction moves promise to inform the future development of LLM-augmented GB-AWE tools to better support novice writers' genre learning and writing development in their disciplines.

Interacting with ChatGPT: Exploring Engagement and Feedback in ESL Writing

Inyoung Na, Iowa State University
Mahdi Duris, Iowa State University
Volker Hegelheimer, Iowa State University

Advancements in generative AI tools, such as ChatGPT, have opened up new possibilities for language learners. These chatbots provide a simple interface to prompt feedback on writing tasks, driving the need to explore how learners interact with them and evaluate their effectiveness as Automated Writing Evaluation (AWE) tools specifically for second language (L2) writers.

In this exploratory study, we investigated the language learning potential (Chapelle, 2001) of ChatGPT, focusing on eight international students from two university-level ESL writing courses at different proficiency levels. By analyzing screen-capture recordings, interviews, and surveys, we explored students' interactions with ChatGPT during three revision sessions. We focused on learner prompts (e.g., asking for feedback, revisions), the nature of feedback received (e.g., summative comments, direct corrections), and the extent of revisions made based on the AI feedback, ranging from minimal to substantial.

Findings indicate that the learning potential of ChatGPT depends on the specificity of learner prompts and ChatGPT's feedback. Manageable and balanced feedback encouraged students to revise on their own, while overwhelming or predominantly positive feedback led them to seek further clarification or additional feedback. Students engaged with the tool in various ways, but most used it to produce revised drafts, often incorporating large portions of the generated feedback. These insights can inform educators on effective strategies for integrating ChatGPT into language classrooms to maximize its benefits for language development.

AI as Instructional Text Writer: A Multi-Dimensional Comparison with Human Texts

Tony Berber Sardinha, Pontifical Catholic University of Sao Paolo, Brazil
Marilsa Shimazumi, Cultura Inglesa College, Brazil

In this paper, we explore the extent to which AI can replicate human-crafted instructional content, more specifically, the sample texts used as models for writing tasks in English as a Foreign Language textbooks, from a corpus perspective. We collected two corpora: (1) the English Language Teaching Textbook (ELTT) corpus, comprising 106,840 words from 500 texts, from 43 textbooks published over 25 years (1996-2021), and (2) a mirror, AI-generated corpus that replicates the human-authored ELTT corpus. Our analytical framework was Multi-Dimensional Analysis (Biber, 1988, 1995; Berber Sardinha & Veirano Pinto, 2014, 2019). The corpus was tagged with the Biber Tagger and post-processed with the Biber TagCount program. Five dimensions were identified, namely (1) Persuasion, speaker engagement, and personal opinion vs Expression of analysis and technical information; (2) Expressive,

interactive, speculative discourse with stance marking; (3) Formal, informative, detailed composition; (4) Narrative and descriptive accounts; (5) Summarized abstracted overviews. We scored both the human and the AI texts on these dimensions, revealing discernible differences between AI-generated and human-authored texts. In general, the study revealed notable differences between AI-generated and human-written EFL coursebook texts. Firstly, AI struggles with crafting texts that necessitate persuasion, engagement, and the conveyance of personal opinions, aspects typically enriched by human authorship. Conversely, texts produced by AI tend to lean more towards delivering analytical and technical information, and less toward expressive, interactive, and speculative language, with a notable reduction in stance-marking features. These distinctions allowed for a successful differentiation between AI and human texts, exceeding an 80% accuracy rate.

Session 7: 4:50 - 6:10 pm

Leveraging ChatGPT for Developing Adaptive Personalized Reading Assessment among College ESL Students

Hamidreza Moeiniasl, University of Toronto, Canada

The emergence of ChatGPT has empowered language teachers to develop adaptive and more effective tests by harnessing its natural language processing (NLP) capabilities. This research followed suit, commencing with the creation of a diverse collection of reading passages to encompass a wide array of topics, difficulty levels, and genres. Then, ChatGPT's NLP underwent training on the compiled dataset, focusing on the intricacies of language inherent in reading passages. This training enabled ChatGPT to generate item bank, analyze and interpret text, and provide feedback on reading responses, as well as explanations for correct or incorrect answers. To investigate the effectiveness of this approach to reading test development, the study included a representative sample of college-level ESL students (187 students) in a Canadian college. Feedback from test-takers' experiences collected through focus groups led to continuous refinement on the tests to enhance the tests' adaptability. Finally, the effectiveness of the reading tests was also compared against the validated DIALANG reading test. The results showed a high correlation of .78, affirming the validity of the newly developed reading tests. By following this approach, educators and assessment professionals can leverage ChatGPT’s NLP techniques to create adaptive assessments that cater to students' diverse language proficiency levels, ultimately enhancing the learning experience and promoting effective assessment practices. Such an approach allows also teachers to develop dynamic tests that not only serve as an assessment tool but also help students to learn from the assessment tool, and uncover the students' matured and maturing skills and knowledge.

Identifying Fairness Issues in Automatically Generated Testing Content

Kevin Stowe, Educational Testing Services
Alyssa Francis, Educational Testing Services

Educational Testing Service (ETS) has been exploring automating the generation of content for a large-scale standardized English proficiency test. One problem that has been identified during this process is that the AI-generated content occasionally violates ETS's fairness and sensitivity guidelines, even when the guidelines are included in the initial prompts for content generation. These types of violations are grounds for rejection during human review of the content because a test's fairness guidelines play a pivotal role in ensuring a level playing field for all test takers by minimizing construct-irrelevant sources of score variance. A critical concern, then, is how to maximize sensitivity and minimize bias in AI-generated test content.

The current study aimed to identify the most effective approach to post-production content refinement. ETS tested various methods (fine-tuning, topic-based classification, few-shot prompting, and prompt self-correction) to reduce the amount of fairness-violating content before the output reached the human review stage. By refining post-generation LLM evaluation methods, the team established a system to filter out much of the content that violates ETS's fairness and sensitivity guidelines. Implementation of the system could improve the acceptance rate of the AI-generated test content by up to 20 percent, potentially making the item development process more efficient. For both AI and language assessment experts, this research can shed light on the complex process of refining LLMs.

Whose ChatGPT? Unveiling Unintended Educational Inequalities Introduced by Large Language Models

Renzhe Yu, Teachers College, Columbia University
Zhen Xu, Teachers College, Columbia University
Sky Wang, Teachers College, Columbia University

The universal availability of ChatGPT and other similar tools since late 2022 has prompted tremendous public excitement and experimental effort about the potential of large language models (LLMs) to improve learning experience and outcomes, especially for learners from disadvantaged backgrounds. However, little research has systematically examined the real-world impacts of LLMs on educational equity beyond theoretical projections and controlled studies. In this study, we analyze how learners' academic writing and associated performance changed in response to the initial availability of LLM tools. Based on 2,868,796 writing assignment submissions from 51,021 college students across 3,862 courses between 2019 and 2023 at a public university in the US, we find that students' overall writing proficiency slightly increased following the availability of ChatGPT and that the writing quality gaps between students with higher and lower baseline language proficiency became narrower than before. However, this equalizing effect was concentrated among students with higher socioeconomic status, and the performance gaps between high-SES and low-SES students in these courses went up following the release of ChatGPT. These findings shed light on the digital divides in the era of LLMs and question the equity benefits of LLMs in the early stage. Based on our findings, we create recommendations for researchers, educators, and policymakers on responsible practices to improve educational equity through LLMs.

Rethinking Language Education in the Age of Generative AI

Zhongfeng Tian, Rutgers University, Newark
Chaoran Wang, Colby College

Recent advancements in generative artificial intelligence (AI) technologies have presented new opportunities and challenges to language education (Ilka & Toncelli, 2023). In this presentation, we propose new directions for rethinking language education, drawing upon the conceptual discussions, empirical research, and innovative pedagogical applications featured in our upcoming co-edited volume Rethinking Language Education in the Age of Generative AI (to be published by Routledge in 2025). We will first provide the audience with a first glimpse of the topics and issues included in the book, such as critical AI literacy, generative AI-informed L2 language teaching and assessment, teacher and student perceptions, tool development for language learning, ethical considerations, policies and guidelines. These topics cover various contexts such as ESL, EFL, world language learning, and language teacher education programs in countries like the US, Canada, Hong Kong (China), South Korea, Thailand, Indonesia, and the Philippines. After mapping this research landscape of generative AI in language education, we will discuss how these critical, cross-cultural, and interdisciplinary perspectives can prompt us to contemplate humanizing and reconceptualizing AI and L2/foreign language pedagogies. We will conclude with the implications of the conceptual, empirical, and practical work discussed in our presentation for future research and teaching. We aim for our presentation to spark meaningful dialogue with the audience, inspiring all of us (e.g., researchers, practitioners, and policymakers) to collectively reimagine within this emerging field, leveraging AIs to facilitate the teaching and learning of languages in innovative, critical, ethical, and responsible ways.

Colloquium Abstracts

Saturday, September 28 3:50 - 4:50 pm

AI Supporting Learning, Teaching, and Assessing across Languages and Modalities: Recent Developments and Insights from Finland

Ari Huhta, University of Jyväskylä, Finland
Matthew Poehner, The Pennsylvania State University
Dmitri Leontjev, University of Jyväskylä, Finland
Raili Hilden, University of Helsinki, Finland
Anna von Zanzen, University of Helsinki, Finland
Roman Yangarber, University of Helsinki, Finland

The proposed colloquium presents research in Finland on computerised systems that leverage AI to support L2 learning, teaching, and assessment. The three papers in the colloquium are based on three research projects that, combined, cover both comprehension and production skills and a range of typologically different languages that also vary in size and status. The first project, DigiTala, focused on Finnish and Swedish, the national languages of Finland, and, by using machine-learning approaches, designed an automated speech recognition system for L2 speech, and procedures for automated assessment and delivery of feedback to learners. The second project, Revita, developed an online platform that allows learners or teachers to upload texts, and uses NLP to create exercises based on these texts that focus on specific linguistic constructs (e.g. tense, syntactic constructions, vocabulary) and to monitor learner progress. The system, originally designed to support revitalisation of small, endangered languages (e.g., Mari, Udmurt), has been extended to several majority languages. The third project, DD-LANG, aims to merge dynamic and diagnostic L2 assessment into a new assessment framework that supports the development of L2 reading and writing in English. The study designed innovative computerised reading exercises that provide learners with AI-enhanced mediation that gradually guides them forward whenever they encounter difficulties. The project uses the Revita platform above, including its specific linguistic exercises.

To summarise, we discuss the implementation of AI across different languages and modalities in three related projects in Finland, the challenges encountered, and plans for further contributions by AI.

Presentations of the Colloquium

Digital tool for L2 speaking assessment

The paper is based on a project in which we developed a prototype of a digital tool for assessing second language speaking. We focused on Finnish and Swedish, the two official languages of Finland. The project was a joint initiative of experts in language education, speech and language processing and phonetics from the University of Helsinki, Aalto University and the University of Jyväskylä.

The aim of the project was twofold: Firstly, to create a tool for formative purposes to improve learning by providing immediate feedback to students on multiple aspects of their oral skills, and secondly, to enable and support teacher assessment in large-scale and high-stakes contexts for summative purposes.

The tool is based on speech recognition and consists of several phases, starting with the spoken signal, scoring and evaluating the samples, feeding in training data and finally testing the functionality of the model. Languages with fewer learners face challenges due to the scarcity of training data. Recent advances in machine learning have made it possible to develop systems with a limited amount of data from the target domain. To this end, we propose automatic speech evaluation systems for spontaneous L2 speech in Finnish and Finnish-Swedish, each consisting of six machine learning models, and report their performance in terms of statistical evaluation criteria.

So far, the system is operational for a limited number of speakers and detects auditory signals in monologues, but it is being further trained and refined to map visual cues of interactive communication.

AI-based L2 learning and teaching

Revita is an approach that supports intermediate-to-advanced L2 learners in improving their linguistic skills (Katinskaia &al. 2017, 2018), which currently include: grammar; vocabulary; aural comprehension, where TTS (text-to-speech) models are available; and pronunciation, where STT (speech-to-text) models are available, adapted for speech assessment. Revita's guiding principle is personalization, motivated by current theories from educational science, such as Czikszentmihalyi's Flow Theory and Vygotsky's Zone of Proximal Development (ZPD):

Learning from arbitrary content — as the ability to choose interesting content motivates the learner.

From the chosen content, automatically creating learning tasks/exercises on a wide range of linguistic constructs (Katinskaia &al. 2023).

Continual monitoring of learner performance across all constructs, to identify the ZPD — so the tasks are neither too easy nor too difficult. Providing dynamic feedback whenever the learner makes mistakes, with step-wise guidance to help the learner arrive at the correct answer on her own. Aggregation of large learner data to identify the relations among constructs, to construct optimal learning paths.

Current challenges include identifying the inventory of constructs essential to L2 teaching. Given such an inventory, Revita can support any language; the original focus was on support and

revitalization of smaller endangered languages.

Revita serves as an intelligent tutor or teacher's assistant. From data collected so far, we observe that Revita is able to make accurate assessments of competency based on practice sessions supervised by the tutor, without resorting to exhaustive testing. In the future much more linguistic knowledge needs to be integrated into the underlying models.

Potential of AI in computerized dynamic-diagnostic assessment of L2 reading

We showcase an ongoing assessment project employing an AI system (Revita) to track and support L2 English reading development among upper secondary school learners in Finland preparing for the national Matriculation Exam. Conceptually, the project is distinct in integrating diagnostic language assessment (DiagA) and dynamic assessment (DA) into a coherent framework. The former is an established approach for creating profiles of learners' strengths and weaknesses, with constructs informed by SLA research (Alderson et al., 2015). The latter follows Vygotsky's (2012) position that the inclusion of mediation (i.e., support provided when learner performance breaks down) is central to diagnosing development because the quality of support learners require expands the diagnosis to include abilities that are still emerging (Poehner & Lantolf, 2023). While both DiagA and DA models have previously included computerized procedures, neither has made use of AI in attempting to both diagnose and promote learner language abilities.

We designed mediation for use in a computerized reading assessment. AI is used to vary the phrasing of mediation messages, deliver the mediation in either Finnish or English according to learner preference, and offer it to them in written or spoken form, depending on learner preference. In addition, learners themselves can upload texts into the system that creates automatic exercises with mediation tailored to learners for practice on narrower but still essential reading constructs (e.g., discourse markers, features of grammar). The AI system is demonstrated, and initial research findings, including learner performance and interaction with the system, are reported.

Poster Abstracts

Friday, September 27 2:15 - 3:15 pm

Supporting Vocabulary Acquisition in Virtual Reality: A Perspective from Embodied Learning

Xin Guan, Teachers College, Columbia University

In recent decades, the development of Virtual Reality (VR) technology has led to an increased application in language learning, particularly in vocabulary acquisition. Embodied learning has been widely explored in the field of language learning and many studies have discovered its role in vocabulary memorization. In the VR-assisted language learning setting, a few studies have also explored how the sensorimotor engagement of the human body can be leveraged for vocabulary acquisition through body-based kinesthetic devices like embodied controllers. However, in what ways the vocabulary gains can be benefited or impeded through embodiment or other mechanisms remains under investigation. The current study aims to explore how embodied control influences vocabulary gains in VR-based language learning. Based on Noun Town, a VR language learning application, 45 non-Spanish speaking college students will be recruited for the experiment to study Spanish noun words. They will be divided into three groups: (1) An embodied VR condition: using controllers for sensorimotor-embodied interaction (2) A non-embodied VR condition: watching the movement of objects only without actions (3) A text-only condition. Methodologically, a mixed method is adopted. The learning gains will be quantitatively measured through three vocabulary tests (pre-tests, immediate and delayed post-tests). Semi-structured interviews and observation will also be conducted to collect participants' perceptions of their learning process. The findings implicate the benefits and challenges of incorporating embodied elements in VR-based vocabulary learning. The results will also show how other factors, like motivation, immersion might potentially influence the embodied learning outcomes.

Discursive Multidimensionality in Educational Management: Generative AI as an Insight Tool for Decision-Making in Colleges

Fernanda Peixoto Coelho, Pontifícia Universidade Católica de São Paulo, Brazil
Tatiana Schmitz de Almeida Lopes, Pontifícia Universidade Católica de São Paulo, Brazil
Adilson Gomes, Pontifícia Universidade Católica de São Paulo, Brazil

Artificial Intelligence is advancing globally, reshaping interactions throughout society. Brazil's transitional phase, relative to more developed nations, offers opportunities to study and adapt before or during these changes. Considering the importance of education for social and economic development, as well as the emergence of new technologies in the contemporary era, this work aims to propose a reflection, through the extraction of broad data to build knowledge that will be the insights that involve application of artificial intelligence in educational management and its current trends. The study will apply Lexical Multimodal Analysis (Berber Sardinha, 2019, 2021; (Biber, 1988), as a methodology and theoretical approach in the analysis of a corpus composed of texts from the Annual Management Plan and Institutional Development Plan which belongs to the 76 Faculty of Technology of the State of São Paulo in the years 2023 and 2024, which will be observed in terms of their main interlocutors, seeking to detect in the colleges documents, the discourses and ideologies, which can be unveiled through the methodology used. The motivation for using lexical multidimensionality is the researcher's perception that there is a lot of information not perceived by educational agents in these texts, thus proposing a computational methodology to detect latent factors with greater precision,

since they are not as perceptible to the human eye. After this step, the resulting data will be processed via GPT chat to observe the recurrence of patterns in both resources, checking whether this AI is capable of capturing the most recurring speeches.

Integrating the Public Perspective into AI-Driven Language Curriculum and Pedagogy Development

Kayvan Shakoury, University of Western Ontario, Canada

Most metaphor studies surrounding AI have focused on exploring educational stakeholders' perceptions of various forms of AI, such as chatbots and computers. Only a few studies have directly examined metaphorical framings referring merely directly to the conceptualizations of AI using restrained prompts, such as "AI is like…because…". Hence, there is a caveat as to the metaphoric perceptions of the umbrella concept of AI held by others, including the public, across different domains of life and society, spanning from industry to health, finance, and beyond. Informed by theories such as conceptual metaphor theory (Lakoff & Johnson, 1987 & 2003), this paper aims to explore metaphoric framings of the mere concept of AI to uncover potentially diverse perceptions of AI and the implications embedded within these metaphors. To achieve this, we rely on analyzing a corpus of newspaper articles (including editorials and letters to editors), believing that the corpus offers a broader representation of perspectives, including those of the public, on AI. The findings demonstrate the prevalence of LIVING ORGANISMS (e.g., PERSON and ECOSYSTEM), TOOL, RESOURCE, and THREAT among in the corpus. These metaphors are utilized to illustrate a spectrum of rhetoric surrounding AI, among which the most significant ones emphasize AI's appeal by attributing characteristics such as creativity, power, intelligence, interactivity, problem-solving, decision-making, transformation, and ability enhancement. Regarding cautionary rhetoric, the threat metaphor was predominantly employed to express concerns regarding potential harms such as breaches of privacy, ethical issues, plagiarism, and the necessity for regulated use of AI.

Stance Classification in Applied Linguistics Using ChatGPT

Yaser Shamsi, Oklahoma State University

Deep neural networks (DNNs) have demonstrated significant potential for sentiment analysis and classification of textual content. Open AI's pre-trained language model, Chat-GPT, is distinguished from the earlier models by its ability to provide explanations for its classifications. The current study focuses on the concept of stance in citation practices within the field of Applied linguistics. Traditional analysis of citation functions has relied on a limited context such as the mere statement where the citation occurs (e.g. Harwood, 2009). Here, I argue that Chat-GPT has the potential to expand the context for analytical purposes and even consider the whole article as the unit of analysis. GPT-4 was prompted to classify citations (N=1054) in twenty different articles within the field of Applied Linguistics and decide the article's stance towards them. GPT-4 was also prompted to provide explanations/justifications for its classification, offering a novel insight into the decision-making processes of deep learning models. The data derived from GPT-4's responses were meticulously analyzed and manually coded to identify themes and features within the text that influenced the model's decision. The findings provide insight into how sentiment analysis is conducted in deep neural networks and how these models interpret academic discourse. Additionally, this study juxtaposes the outcomes from GPT-4's automated classification with manual coding of the stance toward each of the citations, exploring the challenges in comparing the two datasets.

Leveraging Artificial Intelligence and Educational Data Mining to Address Challenges in Education for Students, Educators, Institutions, and Governments

Ecem Kopuz, City University of New York Graduate Center
Birkan Kopuz, Beykent University, Türkiye
Mustafa Kürşat Halat, Boğaziçi University, Türkiye
Ege Yıldırım, University of San Francisco

The global application of educational assessment tools remains limited by the challenge of personalizing materials to meet individual student needs, a process that is both labor-intensive and resource-demanding. Artificial Intelligence (AI) offers a transformative solution by fostering a more collaborative and personalized educational experience. Our platform enhances this by leveraging AI to generate multiple-choice questions enriched with detailed metadata, enabling a comprehensive assessment of student knowledge across various disciplines. The scalability of these questions, combined with Educational Data Mining, allows the system to track and adapt to each student's mastery of knowledge components, significantly reducing the burden on educators. An intelligent tutoring system, powered by conversational AI, offers interactive, student-specific guidance, integrating methodologies such as adaptive learning and spaced repetition to optimize knowledge retention. By incorporating gamification, we address socio-economic barriers, encouraging student engagement and offering scholarship opportunities based on both academic performance and effort. Ultimately, our platform is designed to reduce educational costs for students, institutions, and governments by streamlining resource allocation and optimizing time efficiency.

Saturday, September 28 1:00 - 2:00 pm

Improving the Digital Sovereignty of Teachers: Development of a Training Concept for the Use of AI in Foreign Language Teaching

Andreas Grünewald, University of Bremen, Germany
Leona Droste, University of Bremen, Germany

Although the current state of research shows that the integration of AI in the classroom can help to improve the quality of teaching, save time and support learners with regard to their individual needs and abilities, AI applications in foreign language teaching have so far only been used to a limited extent. In addition to the numerous potentials that new AI tools such as ChatGPT, HeyGen and DeepL offer for foreign language teaching, there is also the risk of spreading misinformation and bias. It is therefore very important that both students and teachers use these applications consciously and safely. In order to promote the reflective skills of teachers in dealing with AI, we are developing a training concept as part of a state-funded research project (DiSo-SGW) that is based on the characteristics of effective and evidence-based teacher training by Lipowsky and Rzejak (2021). The first iteration of field research of a design-based research approach is currently being conducted. In the second iteration, the design will be optimized based on the evaluation results and another implementation takes place in August with foreign language teachers. The aim is to offer this training course nationwide. As part of the poster presentation, we would like to provide further insights into the training concept and present the results of the empirical study on improving the digital sovereignty of foreign language teachers.

Establishing Transparency Boundaries in AI Adoption: Student Awareness and Ethical Integration

Krisztina Domjan, American University

As Artificial Intelligence (AI) tools become increasingly common in higher education's Academic English courses, their proper adoption requires transparency and a solid grasp of AI literacy competencies for students and instructors alike. This study, conducted as part of the Artificial Intelligence in Instruction Faculty Fellowship, investigates how AI tools are communicated and utilized in course syllabi, assignments, and task management within a university setting. The survey data revealed high levels of awareness and understanding of AI policies and the use of AI tools as outlined in the course syllabi. The preliminary analysis found that all participants noted the explicit mention of AI tool usage in the syllabus, the clear instructions on the AI Writing Code of Conduct, and a novel effective approach using color-coded banners indicating permitted AI tools on the Canvas LMS. However, while the usage of grading rubrics as a guideline for AI tool integration showed positive trends, there were mixed responses regarding the regularity of referencing these rubrics, highlighting a potential area for enhancing instructional clarity. Additionally, the practice of adding AI disclaimers in assignments was generally well adhered to, though some students noted challenges in citing AI-generated content, suggesting a need for more structured guidance. In this poster session, I will explore the implications of my research and present strategies for transparent AI use. I will demonstrate best practices for explicit communication, and the importance of ongoing development of AI literacy skills to ensure successful integration in order to develop an informed, ethically aware student body.

AI in Education: Exploring Pre-Service Teachers' Attitudes, Readiness, and Concerns

Dini Arini, Washington State University
Jihee Im, Washington State University
Hadir Alderaan, Washington State University

This study investigates the attitudes, readiness, and concerns of pre-service teachers regarding the integration of Artificial Intelligence (AI) into educational settings. As AI continues to permeate various aspects of life, its potential to enhance educational practices through personalized and adaptive learning systems is significant. However, the integration of technologies in classrooms presents challenges and opportunities that need to be critically examined from the perspective of those who will implement them—pre-service teachers.

Conducted among 43 pre-service teachers from four Indonesian universities and a UK university, the research utilized an online survey to gather data on demographic backgrounds, exposure to AI, attitudes towards AI in education, and perceived training needs. Thematic analysis revealed that while there is general optimism about the benefits of AI, such as enhancing educational outcomes and personalizing learning, there are also significant concerns. These include an overreliance on AI, potential loss of teacher roles, and ethical issues.

Findings indicate that pre-service teachers exhibit moderate confidence in their ability to use AI tools effectively, yet they express a strong need for comprehensive training covering both technical and pedagogical aspects. The study underscores the crucial need for educational institutions to provide support and professional development to equip future educators with the skills necessary to leverage AI effectively while maintaining the essential human elements of teaching.

This research contributes to the understanding of how AI can be integrated into educational practices and the preparation required to navigate this transition smoothly, ensuring that future educators are competent and confident in using AI.

Artificial Intelligence and Technologies for Education: Learning and Teaching Languages

Ana Emilia Fajardo Turbin, University of Brasilia, Brazil

This article talks about the importance of Artificial intelligence in the context of low vision and blind students of additional languages. The technological advancements are nowadays welcoming the entrance of Artificial intelligence that has been present in our lives, such as Siri, Alexa, ChatGPT. We should mention the ethics regarding this area which happens to be of major importance due to the consequences of bad use of Artificial intelligence by cybercriminals. We also mention beliefs that are spread in the mainstream media which make people mistrust artificial intelligence or even fear it. We support our research on authors such as Cope, Bill and Kalantzis, Mary, 2015 to whom multiliteracies. include the technological discoveries. Other authors are also reference in this article such as Novoa, Antonio, (2023) who highlights the importance of reading and writing besides using Artificial intelligence and other tools. We conducted some field work in a public school of languages called Interschool language Center (CIL) in Brasilia, Brazil. We interviewed Danny, a teacher in charge of low vision students. She showed us the work she does to make low vision students and blind students of all grades more comfortable and included to learn languages. Students are adults and teenagers. She mentioned how she works with Braille and the fact that those students count on their accessibility tools on their cellphones. We conclude that the cellphone already brings a lot of features in their accessibility mode and low vision students know that and use it, including Artificial intelligence tools.

Traditional vs. AI-Driven Translation Sources in Language Learning

Dale Courtney Jeffery, Fukui University of Technology, Japan
Bradford J. Lee, Fukui University of Technology, Japan

With the increasing integration of technology in education, there is a growing reliance on AI-driven platforms and apps to enhance language acquisition and proficiency (Yang & Kyun, 2022). Despite the market being saturated with tools and options, comprehensive comparative analysis of usage and user sentiment remains limited. The current study addressed this gap by objectively examining app usage and subsequent learning outcomes among Japanese students. In addition, subjective factors like learner engagement, perceptions, and preferences between "traditional" translation sources, such as paper/electronic dictionaries and online translators (e.g., Google Translate, DeepL, etc.), compared to emerging AI-driven tools like interactive chatbots, were also elicited. In this presentation, we will elucidate not only what apps are the most popular among students at present, but also how they use them, for what purposes, and there impressions of them. These findings hold significant implications for attendees, who will gain a deeper understanding of students' mindset and behavioral patterns, the role of technology in language education, and obtain empirical evidence and insights to inform evidence-based language education practices and policies (Levy & Stockwell, 2006). Understanding the strengths and limitations of each translation source is crucial for facilitating the integration of AI technologies in a way that enhances, rather than replaces, human capabilities in language learning. As we transition towards an AI-driven future, this research contributes valuable insights into fostering AI literacy among learners and preparing educators for a technologically enriched language teaching environment.

A New Perspective on International Chinese Education: Integrating Traditional Standards with ChatGPT

Yan Li, Ohio State University
Zhenyu Park, Yanji, China

This paper examines current trends and challenges in international Chinese education, focusing on how we can integrate modern technology with traditional teaching standards. Specifically, it explores combining ChatGPT with the ACTFL's "5C" standards and the "International Chinese Language Education Chinese Proficiency Standards" in Chinese language instruction. The importance of international Chinese education is first highlighted, emphasizing its role in language teaching fostering cultural exchange, and spreading values. The discussion then moves to using ChatGPT to improve fundamental Chinese language skills such as listening, speaking, reading, and writing. It showcases ChatGPT's potential in creating realistic communicative scenarios, providing varied teaching materials, and supporting practical activities. The paper also addresses the challenges of using technology in education, such as the risk of students becoming too reliant on digital tools, which could detract from their practical language abilities, and the need for educators to adapt their teaching strategies to technological advances continuously. The conclusion stresses that teachers should use ChatGPT as a supportive tool rather than a replacement for conventional methods, aiming to boost students' cross-cultural communication skills and promote well-rounded language development.

Technology Demonstration Abstracts

Friday, September 27 2:15 - 3:15 pm

Optimize My Resume: An AI-Driven Online Tool for Job Search Optimization and Task-Based Language Teaching

Cicely Anne Rude, Teachers College, Columbia University
Michael Campbell Stone, Science Connected

New language teachers and adult English as a second language (ESL) students complete their courses of study with a wealth of knowledge and skills, but before they can put those skills into practice, our graduates need to find jobs. Today, employers increasingly use AI-powered systems to screen application materials. Applicants must pass that hurdle (Madancian & Taherdoost, 2024).

This tech demonstration will introduce an innovative AI-driven tool designed to a) assist teachers in conducting their own job searches and b) scaffold the learning of an important job search skill for adult English language learners (ELLs) in Task-based Language Teaching (TBLT) (Bygate, Skehan, & Swain, 2001, Robinson, 2003).

With a clean and accessible user-interface backed by a large language model and prompt engineering, the tool evaluates the alignment of resumes and cover letters with specific job descriptions. It then provides actionable feedback in a narrative format, pinpointing areas for improvement and enhancing the chances of passing initial AI screening processes. Beta testing has demonstrated a remarkable success rate, with 73 percent of applications passing AI screening hurdles with this tool compared to 10-30 percent success without.

While similar tools exist, this one fills a gap with its language educator-focused design and ease of use. We will showcase the versatility and potential applications for both teaching English and launching language teaching careers.

VR meets AI: Unleashing Potentials in Language Acquisition and Communication Skills Development

Minh Le, Teachers College, Columbia University

Virtual Reality (VR), with its immersive and interactive nature, when harmoniously integrated with AI can provide learners with highly personalized, contextually authentic, and dynamically experiential learning environment that not only adapts content to individual learners' pace and style but also provides real-time feedback to help them develop their linguistic skills. This demo station will give you a glimpse into the exciting possibilities that the synergy of VR and AI offers in the context of language learning and communication skills development.

Adaptive Immersive Language Learning with Socrat.ai

Nina Bamberg, Pedagogy.Cloud

Socrat.ai introduces an adaptive learning feature that personalizes language learning experiences to enhance student engagement and retention. By tracking a student's learned vocabulary and intelligently selecting new words for introduction, Socrat.ai creates reading passages that seamlessly blend new words, learned words, and the student's native language. This immersive approach gradually exposes students to new vocabulary in context while

reducing cognitive load and aiding with contextual understanding through the strategic use of their native language.

Socrat.ai incorporates spaced repetition techniques to reinforce learned vocabulary at optimal intervals, ensuring long-term retention. The platform also employs mastery learning principles, allowing students to progress at their own pace and achieve proficiency before moving on to new content. Socrat.ai employs large language models for text analysis and adaptation and uses proprietary algorithms to generate personalized learning paths.

This technology demonstration will showcase Socrat.ai's adaptive immersive language learning feature. We invite attendees to experience the platform firsthand and provide valuable feedback. We envision Socrat.ai as a tool that encourages language learning through personalized, engaging technology that empowers students to achieve fluency more effectively and efficiently.

Exploring Generative AI for the ESL/EFL Classroom: Twee.com

Sharon Chan, Teachers College, Columbia University

This technology demonstration will invite participants to explore the AI platform Twee.com, which was created as an AI lesson plan/activity creation tool specifically for English language teachers. The presenter will guide participants in creating activities using the platform's various tools, share with participants their own experience with the platform, and engage participants in discussions on the practical implementations of AI lesson-creation technologies, with a focus on creativity as well as limitations and concerns.

Saturday, September 28 1:00 - 2:00 pm

Pangea Chat

Will Crowell Jordan-Cooley, Pangea Chat

Pangea Chat is an intelligent language learning environment which facilitates "learning by doing" within the context of peer-to-peer instant messaging conversation. Pangea Chat integrates human and artificial intelligence (AI) to immediately augment student communicative abilities while fostering learning and supporting educator practices. By building game, learning and assessment mechanics into AI-powered language tools, Pangea Chat helps students produce and understand their target language while supporting engagement and measuring understanding. Within the open-ended possibility space of natural language, Pangea Chat instantiates core principles of ideal educational games: freedom of choice; integrated content and gameplay; adaptive feedback and instruction; and contextually-embedded assessment. To mitigate growing concerns of security and data control, with both social media and AI, Pangea Chat is built on a decentralized and open-source protocol to allow institutions the option to self-host while building a connected network for language and intercultural exchange.

Enhancing Language Education through AI: Predictive Analytics for Assessing Reading Text Level

Kamaleddine Tabine, US Department of State
Michel Tinguiri, US Department of State

This presentation explores using artificial intelligence (AI) in predictive analytics to advance language education via the Interagency Language Roundtable (ILR) scale. By harnessing state-of-the-art AI technologies, the presenters aim to enhance both instructional and evaluative

processes, thus fostering adaptive learning environments tailored to learners' needs.

Dr. Kamaleddine Tabine will showcase how AI can scrutinize and assess textual complexities, employing advanced tools to analyze linguistic elements (syntax, morphology, lexis) and rhetorical components (argumentative coherence, stylistic features) as stipulated by the ILR descriptors. This rigorous evaluation facilitates the precise assessment and categorization of textual challenges, illustrating the pivotal role of AI in accurately predicting ILR levels and refining assessment strategies within language education.

Michel Tinguiri will illustrate how AI can optimize text readability and instructional content to bolster teaching methodologies. His demonstration will highlight the practical applications of AI in classroom settings, emphasizing how these technologies can be customized to cater to individual learning styles and needs, thereby enhancing the instructional process.

The session will comprehensively analyze how AI employs deep learning and natural language processing to analyze key linguistic attributes for precise ILR-level determinations meticulously. The presenters will discuss the broader implications of their demonstrations, exploring potential pathways for further advancements in educational and technological domains. Ethical concerns, such as data privacy and bias mitigation, will also be addressed, underscoring our commitment to the responsible use of AI in educational contexts.

Enhancing Flipped Learning with Artificial Intelligence

Meryem Ezgi Yalcin, Yale University

Flipped course design presents a valuable opportunity to manage mixed ability class environments by offering self-paced learning activities outside of class. The design involves well-structured and supervised out-of-class learning, with pre-F2F and post-F2F components and optimizes F2F learning time for interactive activities. One of the challenges in this design has been the differences in students' self-regulation abilities in independent learning contexts. As technology advances, learning management systems and emerging AI technologies present promising solutions to this obstacle. AI tools enable the instructors to create engaging asynchronous learning materials, which can help reduce the self-regulation issues. This presentation focuses on how AI-powered voice over tools and text to image technology can be utilized to create multimedia learning content suitable for flipped course design in world language classes.

AceIt MR: Intelligent Job Interview Simulator for Non-Native Speakers

Amany Alkhayat, Teachers College, Columbia University

AceIt XR is currently a research project designed to help non-native speakers improve their communication skills specifically for job interviews. This work-in-progress Mixed Reality application helps learners practice interview skills in an immersive environment that blends the physical with the virtual, mimicking real life scenarios. The MR environment includes an AI trainer that is trained on job interview questions, enabling interactive practice and response to learner queries. AceIt XR also aims to mitigate performance-impacting anxiety (Abbasi, Khalil, & John, 2019), enhancing career prospects (Huffcutt et al. 2011; Macan 2009) for non-native speakers by fostering confidence and competence in a supportive, realistic training setting. The presenter will showcase the application to the attendees in which they will practice with the AI avatar and receive feedback.