TSLL

Technology for Second Language Learning

2024

Table of Contents

TSLL
for Second Language Learning

Welcome

Dear TSLL 2024 Participants,
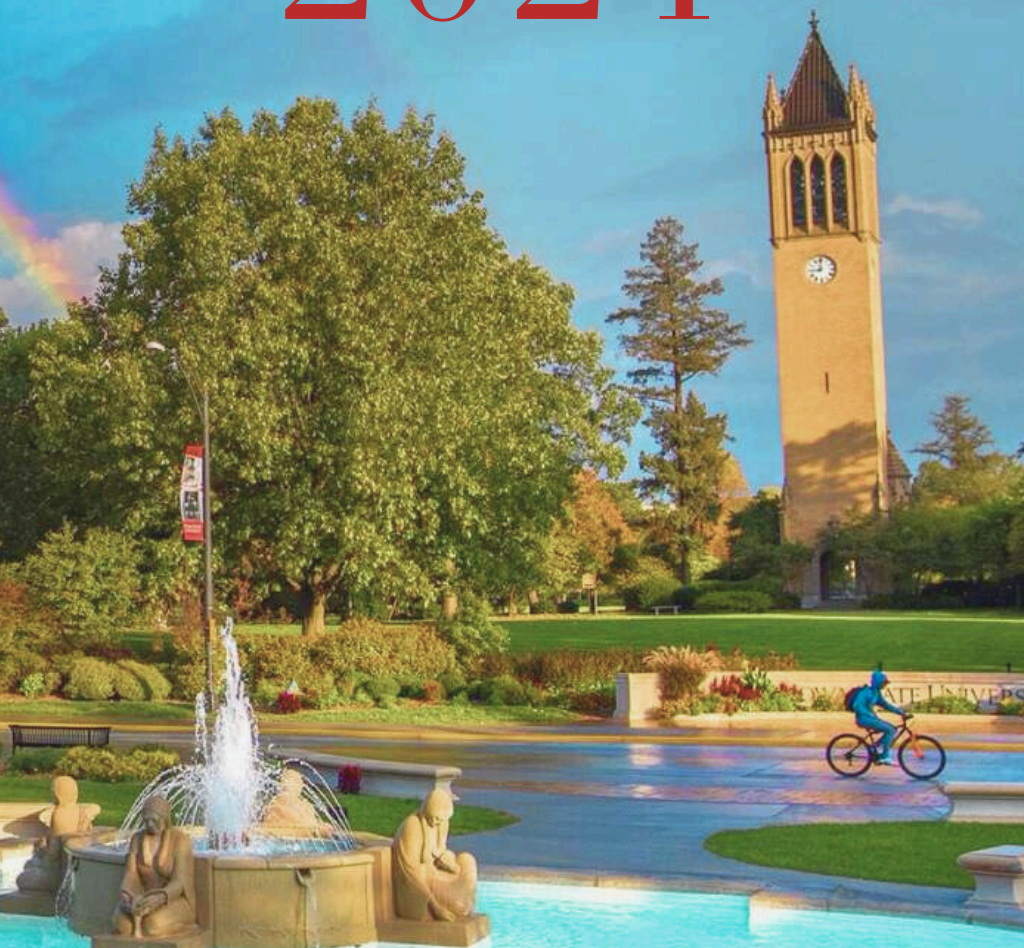
Welcome to the 21st annual Technology for Second Language Learning Conference at Iowa State University! The conference theme "Researching the Language and Use of Generative AI" encompasses issues and questions that GenAI raises for applied linguists. They are grappling with questions such as what are the linguistic, interactional, and pragmatic characteristics of language produced by large language models? How should we expand our understanding of language teaching, learning, and assessment in a world where machines can respond to learners' input with grammatical monologic and dialogic discourse, and where learners have access to a new generation of tools to facilitate their own production of discourse? And how do we build our understanding of the implications of generative AI for language teaching, learning, and assessment through systematic, empirical methods of inquiry?

The three-day conference begins the morning of October 24 and goes through the end of the day on October 26 Central Daylight Time. During this time, participants on the Iowa State University campus and around the world are invited to join the six keynote presentations offered by leaders in the field whose research has provided unique insights into issues surrounding AI in language learning, language assessment, and applied linguistics research. Participants are also invited to attend a range of talks from a rich collection of 52 presentations and one colloquium contributed by researchers from around the world. In addition, participants can access 20 poster presentations asynchronously anytime from Thursday morning until Saturday at noon. The program was developed from over 100 proposals we received.

The presentations delve into issues raised for the field by the widespread accessibility of generative AI built on large language models. They extend the research presented at last year's conference, which are described in the selected papers, Exploring Artificial Intelligence in Applied Linguistics.

We are grateful to the Department of English and the program in Applied Linguistics & Technology at Iowa State University for making TSLL 2024 possible. We thank the conference organizing committee, the applied linguistics faculty, and the English department chair for their enthusiastic support of this event. We thank all participants whose contributions have created an excellent program and rich opportunity for all to learn.

On behalf of the organizing committee, we wish you an enriching conference experience with opportunities to connect with new ideas and colleagues across the globe.

Sincerely,
Carol A. Chapelle, Gulbahar Beckett, Bethany Gray, and Jim Ranalli

Sponsors

Department of English

Applied Linguistics & Technology
Iowa State University

Organizing Committee

Dr. Carol A Chapelle
Iowa State University

Dr. Gulbahar Beckett
Iowa State University

Dr. Jim Ranalli
Iowa State University

Bethany Gray
Iowa State University

Organizing Committee

Mahdi Duris
Iowa State University

Gi Jung Kim
Iowa State University

Danilo Calle Londono
Iowa State University

Hwee Jean (Cindy) Lim
Iowa State University

Jeanne Beck
Iowa State University

Organizing Committee

Sara Afifi
Iowa State University

Andrias Susanto
Iowa State University

Duong Nguyen
Iowa State University

Ali Ebrahimpourlighvani
Iowa State University

Abstract Reviewers

Dr. Alif Silpachai

Dr. Agata Guskaroska

Dr. Angela Hakim

Dr. Bethany Gray

Dr. Erik Voss

Dr. Ewa Kusz

Dr. John Levis

Dr. Junghee Byun

Dr. Katherine Burns

Dr. Kristin Terrill

Dr. Luke Plonsky

Dr. Nazlınur Göktürk

Dr. Sarah Huffman

Dr. Shireen Baghestani

Dr. Sinem Sonsaat-Hegelheimer

Dr. Zoë Zawadzki

Ali Ebrahempour

Amy Walton

Andrias Susanto

Beata Walesiak

Burak Şenel

Cindy Lim

Danilo Calle Londono

Droste Hennings

Duong Nguyen

Erika Latham

Febriana Lestari

Gi Jung Kim

Hannah Reischauer

Junghun Yang

Kate Challis

Khoa Do

MacKenzie Novotny

Monique Yoder

Qi Qi

Quy Pham

Sara Afifi

Sara Lastine

Sarah Davis

Şebnem Kurt

Shuhui Yin

Stacy Amling

Trang Ho

Wren Bouwman

Thank You

Conference Information

Break and Social Time

The Break/Social time is designed for in-person and remote attendees. This time will be reserved for exchanging and enjoying some academic conversations. Please refer to the day's schedule for specific room locations.

Dinner and Social Time

All in-person presenters will be invited to complimentary lunch and dinner during the conference. Please refer to the schedule for locations.

## Wifi information

- The wifi is available for all participants: IASTATE-Guest
- No password is required. However, you will be asked to sign in to use the wifi. Please follow the pop-up instruction.
- Click "Accept" and you can connect to the Internet

Online participation

All online registered participants will receive an email with Zoom instructions and passcodes to access the conference.

Session Guidelines
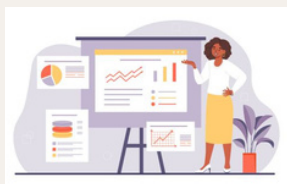
Individual paper

- **In-person**

Paper presentations delivered in person are 20 minutes with five minutes for Q&A, followed by a five-minute transition. We will provide a laptop with audio and visual connections. Presenters may also choose to come with their laptop or similar device with the ability to connect via HDMI. (We will have a dongle ready if your computer needs an adapter). **Please note** that you must use our conference-provided laptop if you prefer to present with a 'Presenter View' mode. In this case, please submit your PowerPoint file named with your last name **48 hours** in advance to this online repository:https://iastate.app.box.com/f/1496e9121f4f4c82abeaa614f90bc3c5

- **Remote**

Paper presentations presented remotely are 20 minutes with five minutes for Q&A, followed by a five-minute transition. Specific Zoom links and passcodes will be sent before the conference starts.

Session Guidelines

Colloquia

- **Colloquia** are organized by **collaborators** who present **four to five short presentations** around a specific theme.

- Each colloquium is allocated a **90-minute session**. Organizers should plan for an **introduction presentation** (15 minutes) that situates the theme of the presentations, followed by **four oral presentations of 20 minutes each**. A specific group of authors should lead each oral presentation. A **concluding presentation** of no more than 15 minutes can open for Q&A and future research implications.

- All presenters (in-person or remote) must be present for the full 90 minutes of the colloquium. In-person presenters will be provided a laptop with audio and visual connections. Presenters may also choose to come with their laptop or similar device with the ability to connect via HDMI. (We will have a dongle ready if your computer needs an adapter). Remote presenters should be in their designated Zoom room five minutes before the start of their presentation. Specific Zoom links and passcodes will be sent before the conference starts.

Session Guidelines

Poster

- **For presenters**
  - **Poster** in PDF format (48" H x 36" W)
  - Sans Serif fonts are strongly recommended
  - Use a minimum 40-point font size
  - **Video presentation of your poster** should be at least 3 mins and not over 7 mins
- **For attendees**
  - All poster presentations will be available to participants asynchronously throughout the conference using a Moodle platform accessible here: https://bit.ly/TSLL24MOODLE.
  - **Registration is required** for all participants, and you must use the following enrollment key to access the Moodle page: **TSLL24poster!**
  - Once the TSLL Moodle becomes available to you, select a poster from the menu on the left side of the page. If you do not see a menu, click on the collapsible menu icon on the top left side of the page. Each poster page includes a title, author(s), abstract, poster PDF, and a brief video presentation of the poster.

- Attendees can either start or contribute to a conversation with the author(s) or other conference participants. To start a conversation, select the "add discussion topic." To contribute to an existing thread, type your comment using the "Reply" option.

- More detailed instructions can be found on our conference website, which includes written instructions and a walkthrough video. **Look for the "Poster (Moodle)" page**.

SCHEDULE

Thursday, October 24, 2024

Time (CDT)	Events	Ross
7:15 – 7.45 a.m.	**Registration & Coffee + Snacks**	212
7:45 a.m.	**Welcome:** Dr. Carol Chapelle Department of English, Iowa State University	212
Posters		
8:00 a.m. – 5:30 p.m.	Poster Presentations (available all day)	Moodle
	Individual Paper	
8:00 – 8:30 a.m.	*Exploring ChatGPT3.5 and ChatGPT4 for AI-enhanced Move Analysis* **Elena Cotos & Burak Senel**	212
8:30 – 9:00 a.m.	*Comparing Human and Generative AI Analysis of Emotions in U.S. Abortion Tweets* **Şebnem Kurt & Mahdi Duris**	212
9:00 – 9:30 a.m.	*Subjectivity Taxonomy for AI-generated Content (STAC): The "Humanness" of Artificial Intelligence* **Danjie Su & Kevin Goslar**	212
9:30 – 10:00 a.m.	***Coffee Break***	215
10:00 – 10:30 a.m.	*ChatGPT, L2 pragmatics, and critical AI literacy* **Robert Godwin-Jones**	212
10:00 – 10:30 a.m.		412

SCHEDULE

Thursday, October 24, 2024

Time (CDT)	Events	Ross
10:30 – 11:00 a.m.	*Framework for Evaluating Large Language Models' Capabilities for Teachers of Low-Resourced Languages* **Nicholas Swinehart**	212
10:30 – 11:00 a.m.		412
11:00 – noon	Plenary 1: **Larissa Goulart** GenAI or Student Writing? Taking a register approach to the human/machine language variation	212
noon – 1:00 p.m.	*Lunch (provided)*	
1:00 – 1:30 p.m.	EMI University Students' Experiences with Using ChatGPT for Language and Literacy Practices: A Qualitative Study **Beyza Aypay & Michelle Bedeker**	212
1:00 – 1:30 p.m.	AI support and coding: Automated corpus analysis for language learners **Robin Couture-Matte & Maura Cruz Enriquez**	412
1:30 – 2:00 p.m.	The Secret Life of International Graduate Students' Use of ChatGPT: Identity and (Dis)empowerment Impacts **Mobina Hosseini**	212
1:30 – 2:00 p.m.	Exploring the Use of ChatGPT in Arabic Language Education: Evaluating Effectiveness and Optimizing Integration **Kamilia Rahmouni**	412

SCHEDULE

Thursday, October 24, 2024

Time (CDT)	Events	Ross
2:00 – 2:30 p.m.	Reliability of Large Language Models at Identifying and Classifying Information in Research Articles **Kristin Terrill & Elena Cotos**	212
2:00 – 2:30 p.m.	Exploring EFL Teachers' Training Readiness to Integrate AI Tools into Their Teaching Practice **Maria Perifanou & Alla Krasulia**	412
2:30 – 3:00 p.m.	***Coffee Break***	215
3:00 – 3:30 p.m.	Developing ArgCoach: Discourse Analysis to Support Specialized Natural Language Understanding of AI **Droste Hennings, MacKenzie Novotny, Samantha Semelroth, Alina Reznitskaya, & Evgeny Chukharev**	212
3:00 – 3:30 p.m.		412
3:30 – 4:00 p.m.	Tracking the Use of Generative AI in Second Language Research **Ling Ding & Cecilia Guanfang Zhao**	212
3:30 – 4:00 p.m.	ChatGPT in ESL Writing: L2 Learners' Perspectives **In Young Na, Mahdi Duris, & Volker Hegelheimer**	412
4:00 – 4:30 p.m.	Leveraging ChatGPT for Language Learning: A Study on Foreign Teachers' Swedish Acquisition **Sofie Johansson & Lina Larsson**	212

SCHEDULE

Thursday, October 24, 2024

Time (CDT)	Events	Ross
4:00 – 4:30 p.m.	Investigating the Influence of ChatGPT Writing Guidance on Peer Review Assessments **Danilo Calle**	412
4:30 – 5:30 a.m.	***Dinner (provided)***	215
5:30 – 6:30 p.m.	Plenary 2: **Eric York** Artificial Intelligence and the Technology of Humankind: A roadmap for critical AI integration	212

SCHEDULE

Friday, October 25, 2024

Time (CDT)	Events	Ross
8:00–8:30 a.m.	**Coffe & Snacks**	**212**
	Individual Paper	
8:30 – 9:00 a.m.	Learners' dialogic interaction with ChatGPT: Could it be an indicator for their proximal development? **Tuba Özturan & Prithvi Shrestha**	212
8:30 – 9:00 a.m.	Learner Agency in the Age of Generative AI: A Study of L2 Writing Skills in Higher Education **Sibel Söğü**	412
9:00 – 9:30 a.m.	Developing a GPT-powered dynamic pragmatics assessment for EFL learners **Gi Jung Kim**	212
9:00 – 9:30 a.m.	ChatGPT as a metalinguistic awareness raising tool for supporting ESL students' academic summary writing skills. **Marcin Kleban**	412
9:30 – 10:00 a.m.	Evaluating the Suitability of AI-Generated Spoken Texts for Listening Assessments: A Comparative Corpus-Based Analysis **Nazlınur Göktürk**	212
9:30 – 10:00 a.m.	The effects of generative AI in scaffolding L2 writing processes **Jianling Liao**	412
10:00 – 10:30 a.m.	*Coffee Break*	215

SCHEDULE

Friday, October 25, 2024

Time (CDT)	Events	Ross
10:30 – 11.00 a.m.	Register-driven prompt design and evaluation of GPT generated stimuli for assessment **Geoff LaFlair, Andrew Runge, Jesse Egbert, & Yağmur Demi**	212
10:30 – 11:00 a.m.	Exploring the Efficacy of ChatGPT Integration on EFL Writing: A Study with Vietnamese Undergraduate Students **Duong Nguyen & Trang Ho**	412
11:00 – noon	Plenary 3: **Ron Darvin** Generative AI, digital literacies and language learning online	212
noon – 1:00 p.m.	*Lunch (provided)*	215
1:00 – 1:30 p.m.	Evaluating AI-Generated Prompts in CILS DUE–B2 exams: A Comparative Validation Study **Sabrina Machetti, Giulia Peri, & Paola Masillo**	212
1:00 – 1:30 p.m.	Can Chat-GPT Produce Corrective Feedback Learners Can Use to Self-Correct Their L2 Texts? **Susanne Rott**	412
1:30 – 2:00 p.m.	Can ChatGPT Evaluate Reading Item Difficulty and Item Discrimination without Response Data? **Hongli Li & Chad Marchong**	212
1:30 – 2:00 p.m.	ChatGPT for Interactive Written Corrective Feedback in French as Second Language Learning and Teaching **Taegan Holmes**	412

SCHEDULE

Friday, October 25, 2024

Time (CDT)	Events	Ross
2:00 – 2:30 p.m.	Exploring ChatGPT's Ability to Recognize and Generate Language Variations: North vs. South Korean Dialects **Jean Young Chun**	212
2:00 – 2:30 p.m.	Enhancing Language Proficiency with GenAI: Investigating the Impact of Corrective Feedback in Academic writing **Ali Ebrahimpourlighvani**	412
2:30 – 3:00 p.m.	*Coffee Break*	215
3:00 – 3:30 p.m.	The potential of GPT-4 in language test preparation: Investigating reading passage generation with zero-shot prompting **Zhengqing Luo**	212
3:00 – 3:30 p.m.	Navigating the Frontiers: Unveiling the Impact of Generative AI on Self-Assessment in College Writing **Hamidreza Moeiniasl**	412
3:30 – 4:00 p.m.	Exploring ChatGPT's potential as automated essay scoring with many-faceted Rasch measurement analysis **Taichi Yamashita**	212
3:30 – 4:00 p.m.	ChatGPT as a Resource for Academic Writing: Students' and an Instructor's Perceptions **Susan Parks**	412

SCHEDULE

Friday, October 25, 2024

Time (CDT)	Events	Ross
4:00 – 4:30 p.m.	Using an AI-chatbot for teaching Korean grammar: A comparison between proficiency levels and task types **Ji-young Shin & Yujeong Choi**	212
4:00 – 4:30 p.m.	Artificial Intelligence and Second Language Writing: Undergraduate Writers' Usage and Perceptions of GenAI Tools' Ethicality **Kübra Çekmegelii, Juan M. Rostrán Valle, & Matt Kessler**	412
4:30 – 5:30 p.m.	*Dinner (provided)*	215
5:30 – 6:30 p.m.	**Plenary:** **Gary Ockey** Harnessing power of AI, the "most destructive technology since the atomic bomb was unleashed"	212

SCHEDULE

Saturday, October 26, 2024

Time (CDT)	Events	Ross
8:30 – 9:00 a.m.	***Coffee & Snacks***	215
	Colloquium	
9:00 – 10:30 a.m.	Practicing and assessing interactional competence using AI-powered agents **Veronika Timpe-Laughlin, Tetyana Sydorenko, Larry Davis, Judit Dombi, Rahul Divekar, Saerhim Oh, Jung In Lee, Reza Neiriz, & Evgeny Chukharev**	212
	(1) Introduction: Interactional competence in an age of AI (2) Language learners' perceptions of written interactions with ChatGPT for practicing English (3) Spoken dialogue technology versus ChatGPT: Benefits and challenges for practicing and assessing oral interaction (4) Elicitation of IC in tests of oral communication: Humans and SDSs (5) Commentary on the studies and implications for future research	212
	Individual Paper	
9:30 – 10:00 a.m.	Leveraging ChatGPT-4 for Enhanced Spanish Language Learning: Insights from the Interactive Narrative 'Escape the Haunt Celia Bravo	412

SCHEDULE

Saturday, October 26, 2024

Time (CDT)	Events	Ross
10:00 – 10:30 a.m.	/ˈfaʊntən/ or */ˈfaʊntaɪn/: Preservice English teacher engagement with AI and TTS for pronunciation self-learning **Alba Paz-López, Kevin Randall Steil, & Boris Vazquez-Calvo**	412
10:30 – 10:55 a.m.	***Coffee Break***	215
11.00 – noon	**Plenary:** **Melinda Dooly** Generative AI in Language Learning: Genie in the Bottle or Devil in Disguise?	212
1:00 – 1:30 p.m.	Teaching Global Englishes with ChatGPT: Modeling and analyzing language varieties with generative AI **Peter Sayer & Ivan Stefano**	212
1:00 – 1:30 p.m.	Empowering Youth Through Language, AI, and Sustainable Advocacy: A Telecollaborative Approach **Maria Mont Algamasilla, Onyemowo Ejeme, Ese Emmanuel Uwosomah, & Melinda Dooly**	412
1:30 – 2:00 p.m.	A Generative AI tool for Dialogue Data Collection in Online Environments **Emma Caputo**	212
1:30 – 2:00 p.m.	Beyond Borders: Integrating Local and Global Perspectives in Project-Based Language Learning **Elif Kemaloglu-Er**	412

SCHEDULE

Saturday, October 26, 2024

Time (CDT)	Events	Ross
2:00 – 3:00 p.m.	**Plenary:** **Kimberly Vinall & Emily Hellmich** Research, Theory, and Practice: Imagining AI technologies in language education	212
3:00 – 3:30 p.m.	***Coffee Break***	215
3:30 – 4:00 p.m.	A Semi-systematic Review of Methodological Aspects in Early GenAI Research for Language Learning: Focus on Studies Published in 2023 **Anne-Marie Sénécal**	212
3:30 – 4:00 p.m.	Exploring L1 Bias in the Automatic Speech Recognition Ability of AI Technology Across Diverse English Accents **Yuna Bae & Okim Kang**	412
4:00 – 4:30 p.m.	Comparing ChatGPT and human rater scores in measuring content in CLIL presentations **Rie Koizumi, Saki Suemori, & Yusuke Kubo**	212
4:00 – 4:30 p.m.	Improving Speaking Skills of EFL Learners through AI-supported activities **Ayse Kizildag**	412
4:30 – 5:00 p.m.	Unlocking the Potential of Generative AI: Enhancing Spanish Language Learning through Humor Translation Strategies **Cristina Pardo-Ballester**	212
4:30 – 5:00 p.m.	Algerian EFL Learners' Views on Replika GenAI Chatbot for Speaking Skills Self-assessment **Aissa Berregui & Mohammed Naoua**	412
5:00 – 5:15 p.m.	**Closing remarks and TSLL 2025 reveal**	212

Plenary Sessions

GENERATIVE AI, DIGITIAL LITERACIES AND LANGUAGE LEARNING ONLINE

Ron Darvin, *University of British Columbia, Canada*

Recognizing digital literacy as a social practice, this presentation examines how the use of generative AI (GenAI) technologies for language learning is contextual and situated, developed through interactions with devices, platforms, and cultures-of-use (Thorne, 2016) and shaped by issues of power. As learners navigate AI platforms, construct prompts and evaluate generated responses, how they are able to achieve their intentions online determines the affordances and constraints of these tools. GenAI however is a black box, and the processes through which output is produced are opaque and complex, dependent on large datasets hidden from view, and governed by the interests of their developers. The extent to which learners are able to exercise agency depends on their capacity to assemble and interpret linguistic and semiotic resources online, while negotiating platform designs, algorithms, norms and conventions.

Informed by theories of platformization (Poell, Nieborg and van Dijck, 2019), programmed sociality (Bucher, 2018), and materialist semiotics (Blommaert, 2013), this presentation argues that the pathway to agentive GenAI use is critical digital literacy. Drawing on insights from a study of the GenAI practices of secondary school students in Canada, it demonstrates how learners develop contrasting dispositions towards GenAI tools, and how the designs of GenAI platforms and their sociotechnical structures (Darvin, 2023) can steer learners towards specific digital literacy practices. Various inequalities circumscribe the use of these tools: from the disparities between free and premium platform versions and mobile vs. laptop access, to the unequal ways these tools recognize low-resource vs. high-resource languages. To address these issues, learners need a critical lens to develop an awareness of how platform designs and output index specific ideologies and interests, and to determine the place of these technologies in the production, consumption, and legitimation of knowledge.

Plenary Sessions

GENERATIVE AI IN LANGUAGE LEARNING: GENIE IN THE BOTTLE OR DEVIL IN DISGUISE

Melinda Dooly, *Universitat Autònoma de Barcelona*

In the last year or so, the publicity surrounding AI in education has centered on its potential to personalize learning, automate administrative tasks, and provide intelligent tutoring systems that enhance student engagement and outcomes. However, fears have also emerged, including concerns about data privacy, the potential for increased inequality, the risk of over-reliance on technology and of course, Vox Populi also has AI 'taking over' the role of teaching and leaving millions of qualified teachers unemployed.

This talk aims to explore whether AI is the genie in the bottle or the devil in disguise by first considering, historically, the delicate balance between educational approaches and technology use, such as Project-Based Language Learning (PBLL) in Virtual Exchange (VE), before considering how we might position ourselves towards vanguard technology, including AI, within current and future learning ecologies.

It will also provide examples of how AI can be used to enhance PBLL and VE as an innovative approach within current and future learning ecologies.

Plenary Sessions

GENAI OR STUDENT WRITING? TAKING A REGISTER APPROACH TO THE HUMAN/MACHINE LANGUAGE VARIATION

Larissa Goulart, *Montclair State University*

Since the public release of OpenAI's ChatGPT in November 2022, teachers and writing researchers have grappled with a central question: How will ChatGPT influence the writing classroom? While some scholars have explored its constructive applications, such as helping with brainstorming and providing feedback, others have scrutinized its potential to disrupt writing instruction, citing concerns such as plagiarism and hallucination. A key concern raised by skeptics is whether tools like ChatGPT, Copilot, and Gemini can be utilized (and to what extent) to complete writing assignments. This presentation adopts a register approach to address this question. I will present findings from two projects examining (1) situational and linguistic variation between student-generated and AI-generated responses to identical prompts, and (2) instructors' perceptions of the distinctions between these two sources of writing.

The first project utilizes two distinct corpora: one composed of assignments authored by undergraduate linguistics majors, and the other composed of texts generated by ChatGPT in response to identical prompts provided to the students.

Each text in both corpora was annotated for situational characteristics, including communicative purpose, setting, presence of abstracts, lists, headings, among others. Both corpora were annotated for lexico-grammatical features with the Biber Tagger and an additive multidimensional analysis was conducted to examine variation between AI-generated and student-generated assignments. The second project adopts a triangulation approach, where the perceptions of ESL teachers regarding both AI-generated and student-generated assignments was explored. This investigation involves comparing teachers' perceptions of AI-generated texts with the findings of the linguistic analysis conducted in the first project.

24

Plenary Sessions

HARNESSING POWER OF AI, THE "MOST DESTRUCTIVE TECHNOLOGY SINCE THE ATOMIC BOMB WAS UNLEASHED"

Gary Ockey, *Iowa State University*

In May of 2024, in his John Hopkins commencement speech, Senator Mitt Romney stated that, AI is "the most destructive technology since the atom bomb was unleashed" and will be "100 times more powerful" in the coming ten years. Romney is not alone in his concerns. Some have suggested that we completely ban the use of (generative) AI in schools and even in other areas of society. However, other researchers and practitioners have largely ignored these concerns, jumping on the AI bandwagon, hoping to not be left behind. Some have suggested that because AI can produce language, including traditional academic essays, we no longer need to teach essay writing or even have writing classes. They argue that we should just teach students to use AI to produce the desired language.

In this talk, I hope to encourage researchers and practitioners, who are on both sides of the debate, to take a step back and think about what it is we want language learners to know and be able to do and how we can best assess these abilities. We then use this understanding as our guide for how to use AI. My main message will be that we should not be letting what generative AI can do guide what we actually do. Instead, we should be harnessing it to help us teach and assess language that is needed to be successful in particular target language use contexts.

My talk will be presented through the lens of construct-based language learning and assessment. I will discuss both traditional and developing language constructs and how we can use AI to teach and assess them. I will describe a few research projects, some of which are grounded in a Dynamic Assessment framework, with the aim of using generative AI to help seamlessly connect language learning and assessment to the abilities that language learners need to effectively communicate in particular target language use situations.

Plenary Sessions

RESEARCH, THEORY, AND PRACTICE: IMAGINING AI TECHNOLOGIES IN LANGUAGE EDUCATION

Kimberly Vinall & Emily Hellmich
University of California, Berkeley

AI tools prompt language teachers to imagine new ways of teaching and require students to imagine new ways of learning. These imaginings open new possibilities, but they can also be anxiety-provoking, as they are still largely unknown. In this presentation, we give initial shape to these imaginings by combining practice with current research and theory.

More specifically, we chronicle the path from empirical research studies to tangible pedagogical materials, with AI language technologies at the center. In the first section, we report on key takeaways from several of our research studies on neural machine translation and generative AI tools (e.g., ChatGPT), including: a) the diverse teacher positionalities emerging technologies open and constrain, drawn from a mixed methods survey (n=165) and follow up interviews (n=11) with US-based university language instructors; b) the complexity of student use of neural machine translation tools, drawn from a computer-tracking study (n=74); and c) learner mediating beliefs that influence (non)use of generative AI tools to support language learning, drawn from a mixed method survey (n=175) of university language learners.

In the second section, we discuss how we used these takeaways, combined with applied language studies theory (i.e., digital literacies, ecological CALL), to imagine instructional materials that support critical and meaningful engagement with AI technologies in the language classroom. Specifically, we showcase a series of activities that target developing language learner understanding of tool use, tool functionality, and tool ethicality across proficiency levels.

Plenary Sessions

ARTIFICIAL INTELLIGENCE AND THE TECHNOLOGY OF HUMANKIND: A ROADMAP FOR CRITICAL AI INTEGRATION

Eric York, *Iowa State University*

The sudden advent and rapid commercialization of large language models (LLMs) represents an inflection point in the progress of humankind, for artificial intelligence (AI) is not mere technological invention but a substantive upgrade in human capability. Ahead of us lies a steepening curve toward ever-more-generalized automation, and, more imminently, a communication revolution of historical proportion. As postsecondary educators and researchers in language-related fields, we are among those most exposed to the disruptive impact of AI, and prominent economists have warned we must "augment [ourselves] or be automated away" (Felten et al., 2023). At the same time, we find language to be more central to more conversations than ever before. If we must grow and change, and indeed we must, we cannot risk our values or our identity—so that is precisely the work that critical AI should be about. It must help us navigate the "jagged frontier" (Dell'Aqua et al., 2023) of AI advancement and chart a course into the uncertain future.

This talk presents a roadmap for critical AI integration in language studies and the humanities more broadly. Framing the development of AI as one of the most significant advancements in human history, this talk provides an update on the current state of research, focusing on the major points of concern for language studies in the near term and highlighting important ways that AI is already transforming how we live, work, and study. This roadmap details an agenda for incorporating AI technologies by fostering new literacies, developing new pedagogies and research methods, facilitating interdisciplinary collaboration, and revisiting the relevance of our work. Through these efforts, we can make meaningful contributions to the most consequential language technology since writing itself.

Poster Abstracts

AI-FACILITATED LITERATURE REVIEW: TESTING A PROPOSED FRAMEWORK WITH GRADUATE STUDENT WRITERS

Kristin Terrill, Lily Compton, Elena Cotos, Sarah Huffman, Maryam Saneie Moghadam & Shangyu Jiang
Iowa State University

International graduate students pursuing degrees in a second-language medium confront a daunting challenge when it comes to preparing scholarly literature reviews for a research article, thesis, or dissertation. Activating advanced reading, writing, and critical thinking skills, academic literature review demands that these language learners integrate knowledge from multiple domains, including fundamental principles of research ethics and integrity. Can generative artificial intelligence (GenAI) technologies ethically, effectively, and efficiently facilitate literature review for graduate students without presenting a threat to the integrity of their research? In this paper, we examine how the literature review task can be deconstructed, and how our 3E heuristic tool for assessing ethicality, efficacy, and efficiency of GenAI use can help determine the appropriateness of GenAI facilitation of process components. We also describe an AI-facilitated literature review (AI-FLR) workshop to be piloted with Iowa State University graduate students in November 2024. The workshop development included establishing GenAI-facilitated task approaches within several process components, including classifying and mapping ideas from scholarly literature; comparing and contrasting viewpoints, findings, and interpretations; synthesizing meaning; and identifying problems, needs or gaps. Additionally, we describe assessment methods for seven learning objectives: * Compare and contrast open and closed GenAI applications * Hypothesize affordances of GenAI for different stages of a literature review * Recall procedural steps to use a GenAI tool to conduct specific tasks * Implement suggested prompts with GenAI tool * Evaluate and critique retrieved results based on prompts * Construct an original argument in response to the literature review prompt outputs of GenAI tool * Critically reflect on own experience with GenAI tool These learning objectives aim to strengthen knowledge in two domains: conducting academic literature reviews and using GenAI tools to facilitate complex processes ethically, effectively, and efficiently. Our presentation will conclude with anticipated findings from our workshop pilot.

Poster Abstracts

BREAKING THE ENGLISH LANGUAGE BARRIER BY NON–NATIVE SPEAKERS AT THE WORKPLACE

Doaa Hamam

Higher Colleges of Technology, Dubai, UAE

The presentation explores one of the most important employability issues faced by graduates at their workplace, which is the role of the English language as an increasingly essential skill for employability and career advancement, particularly for non–native speakers. Some non–native speakers face unique challenges in learning and using English effectively, which can impact their ability to secure and thrive in job opportunities. The research study employs a mixed–methods approach, combining qualitative interviews with faculty members and students alongside quantitative assessments of language competency for a sample of 47 students who graduated from a higher education institution and joined the workforce. Through this multifaceted analysis, the study seeks to identify the key challenges that face graduates at their workplace when using the English Language. The findings of this study indicate that several issues were identified, such as accent and pronunciation, limited vocabulary and fluency, and cultural communication disparities. These issues contribute to challenges where non–native speakers may need help to meet the linguistic and cultural demands of English–speaking workplaces, impeding their career progression and job prospects. The study sheds light on strategies and interventions to support graduates and improve their English language skills.

Poster Abstracts

EFFECTIVE INTEGRATION GENAI CHATBOTS IN SPEAKING ACTIVITIES

Kassandra Sharren, *University of Calgary*

By exploring knowledge that language instructors and learners need to effectively use generative artificial intelligence (AI) chatbots in speaking activities, this literature review aims to identify best practices for supplementing oral language learning with artificial AI chatbots and to identify research gaps. Search terms, including but not limited to, AI chatbots and second language/L2 speaking, generative AI and speaking, and AI chatbots and oral production, were entered into databases, such as Google Scholar, ERIC, and Science Direct, as well as prominent journals in technology, including ReCALL and Language Learning and Technology. Eighty-two articles published in English between 2019 and 2024 that discussed using AI in speaking language learning activities and/or ethical considerations of using AI in language education in the abstract or title were found. Empirical studies focused heavily on learning English as a foreign or second language. Findings include that instructors and learners must possess AI literacy, work effectively with and create with AI, and be aware of meaning negotiation strategies and learning objectives (Chu and Min, 2019; Godwin-Jones, 2024; Han and Lee, 2024; Long and Magerko, 2020; Ng et al., 2021; Vazhayil et al., 2019). These findings can guide the implementation of AI chatbots into oral language activities.

Poster Abstracts

EFFECTIVENESS OF AI-ASSISTED INSTRUCTION AND RELATIONSHIP BETWEEN GENDER AND ACHIEVEMENT IN ACADEMIC WRITING WORKSHOPS

**Mohammad Aliakbari &
Mohammad Mahdi Maadikhah**
Ilam University

Recently popularized AI-based solutions are increasingly attracting attention in educational and instructional contexts. This study aimed to (1) investigate the effectiveness of AI-assisted instruction of academic writing skills, and (2) the relationship between gender and achievement in AI-assisted and non-AI academic writing workshops. To do this, following a placement test and random selection of 36 learners at Ilam University, Iran, two workshops of equal size (18 learners, 9 female and 9 male), length, sessions frequency and proficiency composition were held, one a traditional workshop where instruction of outlining, moves, paragraph organization, feedback and revision guidance were provided solely by the teacher, the other involving use of ChatGPT and Microsoft Copilot for the aforementioned steps. At the completion of workshops, participants were asked to submit three papers rated by six external raters in a double-blind manner. The results showed that, (1) as participants of the AI-assisted workshop outperformed the other group, AI-assisted instruction seems effective, (2) more proficient learners outperformed the less proficient peers, and (3) no statistically significant relationship between gender and achievement in either groups was observed. Further studies on effectiveness of AI-powered solutions, as well as effects of prior training, proficiency, gender and individual differences are called for.

Poster Abstracts

EXPLAINABLE GENERATIVE AI IN HIGHER EDUCATION: A CONTRARIAN POSITION ON BLACK BOXES AND GLASS BOXES

Shahneela Tasmin Sharmi & M. Gregory Tweedie
University of Calgary

This poster presentation will argue that the black box problem of generative artificial intelligence (GAI) in assessment in higher education (HE), while certainly a challenging issue to solve, does not present an insurmountable one. Rather, this presentation takes the contrarian position that assessment in HE itself has been plagued by the black box problem, perhaps since inception, and it may even be that GAI represents a unique and timely opportunity to right historical wrongs. The researchers consider via case study student essay admission processes deployed by a large university in Bangladesh, illustrating how a decidedly analog procedure still represents a black box approach to high stakes assessment. The presentation then contextualizes six elements of an Explainable AI in Education framework (Khosravi et al, 2022) to this particular setting, and proposes a novel technical solution whereby Explainable GAI might serve to democratize and demystify processes currently inscrutable and top-down. The researchers discuss how GAI might assist in moving HE admissions assessment from black box to glass box, in keeping with Stein's (2016) conception of just educational measurement.

Poster Abstracts

EXPLORING CHATGPT AS A ROLE–PLAY PARTNER IN IMPROVING INTERACTIONAL COMPETENCE: IMPLICATIONS FOR LANGUAGE TEACHERS AND LEARNERS

Sun-Kwang Bae,
Defense Language Institute Foreign Language Center
Myoyoung Kim & Jee Eun Gaetz,
DLIFLC

Computer partners in speaking tests have the limited ability to co-construct interaction to achieve targeted social actions compared to human partners (Ockey and Chukharev–Hudilainen, 2021). This study examined ChatGPT's capability in engaging in role plays as a conversation partner in Korean in order to explore how the learners can compensate for limitations and benefit (from) AI assistance in improving interactional competence (IC). Thirteen role plays in Korean were generated using ChatGPT 3.5 under diverse conditions: settings, prompt languages, speaker roles, and proficiency levels. Conversation analysis results of the 13 role play scripts showed ChatGPT's strengths in relation to IC features: it understood pragmatic meanings, was able to correct the course of interaction accordingly, and disagreed politely. However, areas for improvement were identified as well: ChatGPT was unable to continue the interaction in the absence of explicit stage directions, and switched roles occasionally. ChatGPT was also inconsistent in selecting register at the morpho–lexical level. It was observed that quality of role plays depends on the prompt factors such as prompt language and specifications for instructions, as well as role play setting.

Poster Abstracts

EXPLORING LANGUAGE IDEOLOGIES IN GENAI CHATBOTS: A MATCHED–GUISE QUALITATIVE INQUIRY

Zeynep Arslan & Peter Sayer,
The Ohio State University

As generative artificial intelligence has blossomed, GenAI chatbots have proliferated. By harnessing large language models (LLMs), GenAIs are engineered to replicate human intelligence, and these tools offer a versatile approach to language processing tasks (Hadi et al., 2023). However, one issue that has arisen is the recognition that the LLMs that GenAIs use for training data manifest all the same biases that humans do. Therefore, the GenAIs interactions with users may embody the prevailing language ideologies within society. This qualitative inquiry employs a modified matched–guise technique (Lambert, 1960) to explore the inherent language ideologies expressed by four widely–used chatbots: ChatGPT, Perplexity, CoPilot, and Gemini. The transcribed speech excerpt, featuring linguistic markers of African American Vernacular English as the guise, was focused on an academic subject. Within the prompts provided to each chatbot, we tasked them with assessing the speaker's educational background, socio–economic status, English proficiency, and potential racial or ethnic identity by interpreting linguistic markers. Employing thematic analysis of the gathered data (Braun and Clarke, 2022), the study reveals various language ideological structures characterizing each AI, shedding light on the connection between AI technology and societal linguistic frameworks.

Poster Abstracts

EXPLORING THE IMPACT OF AI-BASED TOOLS IN IELTS SPEAKING SIMULATIONS

Andrias Susanto, *Iowa State University*

The IELTS Speaking test, administered by a single human examiner who prompts questions and rates test-takers' speaking performance, faces reliability challenges in score assignment. This study aims to address this issue by investigating the efficacy of artificial intelligence (AI) tools in both question prompting and response scoring. The AI tools encompass AI-generated audio and video recordings for online IELTS Speaking test simulations, an AI-generated non-player character (NPC) projected using virtual and augmented reality technologies for in-person test simulations, and an automated scoring system (ASS) for response analysis and scoring. Through a mixed-methods approach, the research endeavors to explore the impact of these tools on the IELTS speaking holistic and analytic scores of 300 students. These scores encompass fluency and coherence, grammar, vocabulary, and pronunciation. The study also seeks to compare the rating performance of human raters and the ASS, examine potential biases, and document the perceptions of IELTS test preparation tutors and students regarding AI-generated modalities for test procedures and evaluations. Quantitative methods will primarily be employed, with the exception of identifying stakeholders' perceptions. The objectives of this study include investigating the similarities and differences in (1) test administrations using AI tools and human examiners, (2) speaking scoring between human raters and the ASS, and (3) stakeholders' perceptions of using AI tools in both administering and rating IELTS speaking test candidates' performances.

Poster Abstracts

EXPLORING THE RELATIONSHIP BETWEEN EAP LEARNERS' PROFICIENCY, PERSONALITY AND ACHIEVEMENT IN AN AI-ASSISTED WRITING COURSE

Reza Khany & Mohammad Mahdi Maadikhah,
Ilam University

With recent advancements in generative AI, the application of AI chatbots in language instruction has gained attention. Identifying the factors influencing, or related to the effectiveness and usefulness of AI-based solutions in language teaching is of special interest to the ELT community. This study aimed to investigate the relationship between EAP learners' proficiency level, personality traits and final achievements in an AI-assisted academic writing course at Ilam University in Iran. To this end, a general proficiency test was given to 42 MA students enrolling in an academic writing course. Based on the test results, two homogeneous classes of 21 students with equal number of students of the three proficiency levels lower intermediate, upper intermediate and advanced were formed. Students received AI-assisted instruction using ChatGPT and Microsoft Copilot for outlining, revision guidance and feedback. Students were asked to complete NEO-PI-R inventory (short form) along with submission of three essays rated by three raters in a double-blind manner. The results of the data analysis revealed that more proficient students achieved higher, and the traits conscientiousness, extraversion and openness to experience had direct positive relationship with achievement, while neuroticism had direct negative relationship, and no significant relationship was observed between agreeableness and achievement.

Poster Abstracts

FROM "ASR–BASED" TO "AI–BASED": UPGRADING PRONUNCIATION ACTIVITIES

William Gottardi & Rosane Silveira,
Federal University of Santa Catarina (Brazil)

This poster presentation demonstrates the use of three AI-based Pronunciation Activities adapted from a prior study that explored teachers' assessment of ASR-based pronunciation activities intended for use in second language (L2) English classes (Gottardi, 2023). The original activities were designed in accordance with the six criteria for CALL task appropriateness (Chapelle, 2001) and rated by 12 in-service L2 English teachers. The three ASR-based pronunciation activities which participant-teachers identified as more likely to be used in L2 English classes were selected to be redesigned. These activities were adjusted to incorporate an AI-powered assistant, Microsoft Copilot, replacing the ASR program (Google Translate's speech feature). This modification enhanced the quality of the instant feedback during the pronunciation practice, which may help learners and teachers to interpret ASR orthographic feedback. Lastly, the prompts used for the AI-based Pronunciation Activities will be shared with the audience to stimulate discussion on the topic.

Poster Abstracts

INTERACTIVE L2 READING ENHANCEMENT: COMBINING AI CAPABILITIES WITH TEACHER INSIGHTS

Mihwa Lee, Björn Rudzewitz & Xiaobin Chen,
University of Tübingen

Acquiring a second language (L2) significantly benefits from rich language input, with reading serving as a vital conduit for such enrichment (VanPatten and Williams, 2015). Recognizing linguistic forms and categories during reading activities is crucial for effective L2 learning, as it enhances language awareness among learners (Long and Robinson, 1998). Input enhancement (Sharwood Smith, 1993), such as visual cues highlighting linguistic forms, has proven to be effective in facilitating this process (Ruiz et al., 2024). Yet, the manual application of these techniques is time-consuming and impractical in everyday ESL/EFL contexts. Existing automatic systems (e.g., Meurers et al., 2010) often lack comprehensive instructional support for a range of linguistic constructs in texts. Addressing these challenges, we present the Annotated Reading Enhancement System (ARES), a web-based platform designed to enrich L2 reading experiences. ARES is capable of autonomously identifying over 650 grammar constructs listed in the English Grammar Profile, offering learners detailed grammar explanation and practical examples. To support practical teaching, ARES implements LLM for tailored reading comprehension questions and answer evaluation, with teachers in the loop. We present the development of ARES and future directions, including our plans to leverage system interaction logs to advance SLA research and refine educational tools.

Poster Abstracts

LITHUANIAN UNIVERSITY STUDENTS ON USING AI TOOLS FOR LEARNING ENGLISH

Aurelija Daukšaitė-Kolpakovienė,
Vytautas Magnus University, Lithuania

This case study aimed to investigate Lithuanian university students' perceptions and experiences of the use of AI-based technologies while learning English. It examined the advantages and disadvantages the students could identify and reflect on (whether AI empowered their learning in any way or not). An online questionnaire with open-ended and closed-ended questions was administered to twenty-eight Lithuanian students at a liberal arts university in Lithuania to gather data. The data was processed using quantitative and qualitative methods. The findings reveal that the study participants used AI tools for a limited number of purposes and functionalities and were rather sceptical about them with regards to improvement of various English skills and competences when using them. Just over a half of all the study participants felt motivated to use AI tools for English learning to achieve the following practical aims: to save time, get ideas, and improve their vocabulary. However, the disadvantages were related to the point of view that AI might not be something to learn or receive information from because of its numerous flaws.

Poster Abstracts

LLM'S PERFORMANCE IN DETERMINING CEFR LEVELS OF TEXTS: CASES OF ENGLISH AND MANDARIN

Daniil M. Ozernyi,
Northwestern University

Recently, it has come to our attention that some classroom practitioners use ChatGPT, Microsoft CoPilot and similar tools in order to determine the level of texts (CEFR or ACTFL) and in so doing ascertain whether texts are appropriate in their classrooms. However, no research has looked at whether LLMs are good tools to determine a level of a text, essentially fulfilling the role of a "text profiler". It is also not clear what mechanism LLMs might be using in order to determine the level of a text. However, it is well-known that LLMs can indeed "tune" their performance if prompted to pose as an "intermediate" or "advanced" speaker of English, etc. We investigate the problem described above by asking ChatGPT and Microsoft Copilot to determine the level of several texts, across the following variables: CEFR level (A2, B1, B2) and language (English, Mandarin-Simplified/簡體字, Mandarin-Traditional/繁體字). Traditional vs. simplified manipulation was chosen because LLMs have fewer available corpora in traditional Mandarin, though the language is the same. We therefore predict asymmetry in LLM's performance. We compare LLM's results against our own profiling: for Mandarin we determine level by percent of vocabulary by HSK level, for English we take texts already validated for international exams like Cambridge B2 First, etc. The number of texts we used was 30 for each language for each level. Our results indicate that neither ChatGPT nor Microsoft Copilot can reliably detect the CEFR level of a text, usually giving an approximate result (e.g., "between B1 and B2") even when the text is C1 or A2. Further studies are necessary to fine-tune our understanding of how LMMs deal with this type of lexical /grammar profiling.

Poster Abstracts

O USO DE IA GENERATIVA PARA O AGRUPAMENTO E CATEGORIZAÇÃO DAS ESTRUTURAS LEXICAIS (THE USE OF GENERATIVE AI FOR CLUSTERING AND CATEGORIZING LEXICAL STRUCTURES)

Simone de Oliveira,
Pontifícia Universidade Católica de São Paulo
Júlia Tamagno
Universidade Federal do Rio Grande do Sul
Ana Bocorny,
Universidade Federal do Rio Grande do Sul
Tony Berber Sardinha,
Pontifícia Universidade Católica de São Paulo
Deise Prina Dutra,
Universidade Federal de Minas Gerais
& Gisele Rotta
Universidade Federal do Rio Grande do Sul

The project aims to develop software integrating Artificial Intelligence, Natural Language Processing, pre-trained Generative AI Transformers, and Machine Learning to streamline the extraction, clustering, and categorization of phraseological linguistic data for corpus analysis. Its objective is to examine English-language articles published between 2013 and 2023 in international journals across various disciplines, including Humanities, Applied Social Sciences, Linguistics, Literature, and Arts, with the aim of characterizing conventional language use. Drawing upon methodologies by Biber (2009) and Gray and Biber (2013) for Lexical Bundles extraction, the research employs both qualitative and quantitative approaches. This validation of our automated approach examines the structural logic of Efficiency, composed of the criteria of Execution Speed and Quality. This will allow observation in percentages of what has improved, in which criteria, and by how much it has improved from the perspective of Efficiency. Notably, preliminary findings suggest that automated and manual processes yield comparable outcomes, underscoring the efficacy of automated methodologies. The research aims to deliver an AI-powered tool to expedite the analysis of extensive corpora, thereby enhancing scholarly inquiry.

Poster Abstracts

PEDAGOGICAL PARADIGMS REIMAGINED: GENERATIVE AI'S INFLUENCE ON L2 WRITING INSTRUCTION

Solbee Kim, *The Ohio State University*
Youngjoo Yi, *The Ohio State University*
& Jinsil Jang, *Dongshin University*

As artificial intelligence (AI) advances, particularly with generative AI tools like OpenAI's ChatGPT, there's increasing interest in their impact on writing development among multilingual students. As educators, our responsibility is to facilitate students' adaptation to this evolving technological landscape, optimizing the educational benefits while mitigating associated challenges. To this end, our research adopted a systematic synthesis approach to analyze empirical studies that explore the role of generative AI in improving writing skills among multilingual writers. This comprehensive analysis covers various dimensions, including the types of technologies employed, their direct and indirect effects on writing proficiency, the contexts in which these technologies are integrated (e.g., research designs, participant demographics), and a detailed evaluation of each technology's advantages and limitations. This presentation aims to demonstrate the key themes emerging from our research and discuss practical pedagogical strategies for integrating generative AI tools into writing instruction. By highlighting successful practices and common pitfalls, we provide educators with a nuanced understanding of how generative AI can be leveraged to enhance multilingual writing development. Our findings suggest that when implemented strategically, generative AI can significantly enrich the learning experience of multilingual writers by providing tailored feedback, enhancing engagement, and offering diverse linguistic models. The presentation will also address future research directions that promise to deepen our understanding of the intersections between generative AI and language education.

Poster Abstracts

RESPONDING TO MULTIPLE-CHOICE ITEMS WITHOUT ACCESSING PASSAGES— IMPLICATION OF USING CHATGPT FOR TEST VALIDATION

Hongli Li, Roula Aldib & Chad Marchong
Georgia State University

There have been concerns regarding the construct validity of reading comprehension tests, particularly when students achieve correct responses to multiple-choice items without reading the accompanying passages. For instance, previous research has indicated correct response rates exceeded chance levels when students did not read the passages (Coleman et al., 2009). However, validating the construct validity of reading comprehension tests through traditional piloting methods requires substantial costs for test administration, data collection, and data analysis. In this study, we aim to address this challenge by utilizing ChatGPT, an advanced generative AI model, to provide responses to multiple-choice reading items without access to the passages. We will request both ChatGPT 3.5 and ChatGPT 4 to generate responses to reading comprehension items without access to the passages and ask for detailed rationales for each response. We hypothesize that given the expansive background knowledge and analytical capabilities of ChatGPT, its correct response rate would represent the upper limit of a typical test taker's ability in responding to multiple-choice items without accessing the passages. The results will provide insights about the construct validity of the reading comprehension test, and the rationales provided by ChatGPT will inform the revision of the test to enhance its construct validity.

Poster Abstracts

STRATEGIES EMPLOYED BY STUDENTS IN EMI PROGRAMS TO ENHANCE SECOND LANGUAGE LEARNING FOR ACADEMIC PURPOSES

Duygu Ispinar Akcayoglu & Omer Ozer,
Adana Alparslan Turkes Science and Technology University

The proliferation of freely accessible generative AI has garnered significant attention worldwide, promising to reshape language-related practices in diverse university disciplines. Lee et al. (2023) highlight the potential of generative AI to positively impact non-native English speakers by providing tools for effective English language learning. Within the context of English medium undergraduate programs, generative AI stands as a potentially valuable resource for supporting students' language learning journeys, offering personalized feedback, practice opportunities, and supplementary resources to enhance both writing and speaking skills. In this study, we interviewed non-native English-speaking students enrolled in English-Medium Instruction (EMI) undergraduate programs at a Turkish state university. Participants were selected based on their enrollment in English for Academic Purposes (EAP) courses and active use of generative AI tools. Twelve students from various academic programs took part in the interviews, and their responses were analyzed using content analysis. The initial findings provide insight into the learning strategies and styles that students use to effectively utilize generative AI tools. Additionally, students reported using generative AI for purposes such as confidence building, accessing open educational resources, and clarification. These insights are crucial for researchers and content lecturers in higher education as they navigate an environment where machines generate contextually relevant, grammatically correct responses to prompts. The imperatives and possibilities presented by generative AI necessitate immediate research that guides EMI content lecturers and language instructors in adapting to their evolving roles and enhances our comprehension of students' language use and learning in this technological era.

Presentation Abstracts

COMPARING HUMAN AND GENERATIVE AI ANALYSIS OF EMOTIONS IN U.S. ABORTION TWEETS

Şebnem Kurt & Mahdi Duris,
Iowa State University

This study examines Twitter user emotions and stances as expressed in U.S. abortion-related tweets. The study utilized data from a subset of the Twitter Abortion Corpus (TAC), which was compiled using the Twitter API and a Python program that extracted publicly available Tweets with the keyword "abortion" from 50 US state capitals and seven large cities. A manual analysis was conducted on a dataset of 100 tweets selected from a total of 2565 tweets retrieved from the Boston_TAC dataset. Employing Martin and White's Appraisal Framework (2005), two human coders assessed the AFFECT subcategory within the Attitude domain to identify positive, negative, and neutral emotions in the tweets. Additionally, Du Bois' stance-taking analysis (2007) was applied to determine pro-abortion, anti-abortion, or neutral stances within the dataset. In a secondary analysis, we used generative AI tools ChatGPT 3.5, ChatGPT 4, and BARD to compare the effectiveness of human and AI coding in detecting emotional content using the two frameworks in the same dataset. The human and AI coding results were compared using a chi-square test of independence to assess the similarity and statistical significance of their associations with emotional categories. This study further informs about the LLM and human-related challenges associated with qualitative data coding, speed of analysis, and rater reliability. This research contributes to understanding the efficacy of human and Generative AI in analyzing emotional content on social media platforms. The findings have implications for social media research methodologies, particularly in sentiment analysis and computational social science.

Presentation Abstracts

AI FEEDBACK TO SUPPORT THE FUTURE FORESIGHT WRITING PROCESS: OBSERVATIONS FROM THE MYESSAYFEEDBACK.AI TOOL

Lana Hiasat, *Higher Colleges of Technology-Dubai*

The purpose of this presentation is to share how AI feedback tool was used to support second language writing process in future foresight class in the United Arab Emirates (UAE). As part of elective courses, Emirati undergraduate students take a future foresight class in English. However, since English is their second language and the content can be challenging, frequent and prompt feedback helps students with their grasp and application of the future foresight tools. Teachers face an additional challenge to teaching content in second language which is having large classes that make personalized and timely feedback difficult to achieve. UAE has fully supported the integration of AI to improve productivity and processes and therefore incorporating AI into teaching is recommended and helps promote AI literacy. Findings of focus group discussions will be shared to shed light on Emirati students' experiences as they interacted with the AI feedback tools in comparison to human feedback. In addition, this presentation will be addressing the key question of whether AI feedback was able to address the limitations of traditional human feedback and support for second language learners and identify the shortcomings. The recommendations will be focused on the best approach to balance human and AI support into second language classes based on the observations in teaching a content class to non–English language speakers. An important recommendation is to continuously seek feedback from students regarding their own experiences and interactions with AI tools.

Presentation Abstracts

EXPLORING CHATGPT3.5 AND CHATGPT4 FOR AI-ENHANCED MOVE ANALYSIS

Elena Cotos & Burak Şenel,
Iowa State University

Generative AI marks a captivating frontier where Large Language Models (LLMs) like ChatGPT promise to redefine methodological possibilities and deepen our understanding of linguistic phenomena by enhancing and possibly automating procedures for linguistic data analyses (Yu et al., 2023). Moving towards automation is particularly important for methods requiring manual annotation, such as Move Analysis (MA) (Swales, 1990) of genres. While researchers began exploring ChatGPT use for automatic genre identification (Kuzman et al., 2023) and for comparison with manual annotation (Huang et al., 2023), the ability of ChatGPT to identify rhetorical genre traits remains uncharted. This study examines the capabilities of ChatGPT3.5 and ChatGPT4 to classify text into rhetorical moves. Focusing on the moves of research article introductions, the procedure involved an initial MA domain knowledge check for both ChatGPT3.5 and ChatGPT4, which demonstrated that they lack appropriate knowledge for identifying moves/steps. Next, following OpenAI strategies for prompt engineering, we designed 5 prompts, each containing different MA domain knowledge criteria for the move/step categories: Prompt1-move-definitions, Prompt2-step-definitions, Prompt3-step-sentences, Prompt4-step-expressions, Prompt5-all-integrated. This allowed for exploratory prompt-based zero, single, and few-shot classification experiments (Chae and Davidson, 2023) with both LMMs. The experiments with all five prompts used the same text and entailed 10-fold cross-validation at three temperature parameters, applied to control the randomness of LLMs' predictions. Per-fold accuracy, precision, recall, and F-1 scores, as well as overall reliability metrics were calculated for each experiment and then used for comparisons of classification performance within and across ChatGPT3.5 and ChatGPT4 models. Noticeable differences were detected. ChatGPT4 demonstrated more consistent performance across different prompts and temperature values, while ChatGPT3.5 overall outperformed ChatGPT4 on all prompts except Prompt1. Considering these encouraging results, we will further replicate these experiments on a large dataset, the ISURA-Introduction corpus, evaluating LMMs' performance using the move/step annotated version of this corpus.

Presentation Abstracts

FRAMEWORK FOR EVALUATING LARGE LANGUAGE MODELS' CAPABILITIES FOR TEACHERS OF LOW-RESOURCED LANGUAGES

Nicholas Swinehart, Phuong Nguyen & Ellen Yeh
University of Chicago

One constraint on the potential of generative AI in language teaching is large language models' (LLMs) performance in low-resourced languages, due to the heavy bias toward English and other Western European languages in the models' training (Johnson et al., 2022). Methods that are currently used by AI researchers and developers to evaluate LLMs multilingual performance, like machine translation benchmarks and the multiple-choice Massive Multitask Language Understanding test, do little to help teachers understand how useful an LLM may be in generating text and responding to queries in their target language. This paper presents a framework that language instructors can use to determine an LLM chatbot's ability to generate and analyze the target language in ways related to language teaching, along with results from piloting this framework with 30 instructors of 20 high- and low-resourced languages (Besacier et al., 2014). Within this framework, instructors elicit a series of ten tasks related to language teaching of increasing complexity (e.g., summarizing a text, generating a short dialogue, and generating genre-specific texts) from an AI chatbot, then evaluate the responses' accuracy and usefulness for their teaching using a four-point Likert scale. Descriptive analysis was performed on the responses from the pilot study to examine the interactions among the type of language, the complexity of the task, and the performance of the LLM. This framework helps language instructors have a better understanding of precisely what LLMs can and cannot do in their language, which languages are currently being left out of the "AI revolution," and how to monitor improvements over time.

Presentation Abstracts

SUBJECTIVITY TAXONOMY FOR AI-GENERATED CONTENT (STAC): THE "HUMANNESS" OF ARTIFICIAL INTELLIGENCE

Danjie Su & Kevin Goslar

University of Arkansas

AI content quality is crucial for applications in all fields. To be compatible with humans, AI content must exhibit subjectivity. Whenever humans speak or act, they convey subjectivity—the expression of self. AI research has addressed some subjective aspects like certainty, emotion, sentiment, and bias. The lack of a classification restricts AI evaluation and cross-field research due to inconsistent terms and incomplete understanding. This study proposes an initial taxonomy for evaluating the subjectivity of AI-generated content—the Subjectivity Taxonomy for AI-generated Content (STAC). Subjectivity in AI-generated content refers to simulated human perspectives, emotions, beliefs, intentions, empathy, and alignment with others. STAC applies to language and multimodality. We identify major areas of human subjectivity on cognitive, linguistic, psychological, and social dimensions, such as affective stance, agency, alignment, deontic modality, discourse grounding, egocentrism, epistemic stance, euphemism, framing, humor, irony, identity, ideology, bias, implicature, lens, linguistic construal, metalinguistic awareness, metaphor, morality, politeness, speech acts, and theory of mind. These terms are drawn from Linguistics, Sociology, Communication, Anthropology, Philosophy, and Psychology. We outline STAC to match human subjectivity, drawing from Computational Linguistics, Computer Science, Applied Linguistics, and the industry, as well as machine learning datasets. We retain and consolidate existing terms in AI research and develop straightforward new terms. I. Capability domains in STAC include Attitude, Agency, Certainty, Cognitive construal, Emotion, Euphemism, Framing, Grounding, Humor, Identity, Implicature, Intention, Lens, Meta-awareness, Metaphor, Morality, Obligation, Politeness, Positioning, Sentiment, Speech act, and Theory of mind. II. Risk domains in STAC include Bias, Divisiveness, Egocentrism, Narcissism, and Toxicity. We define each domain, link it with human subjectivity, and demonstrate its use in real-world scenarios like language education, marketing, customer service, and political analysis. We also provide examples of benchmarks and datasets when relevant. An open and living system, STAC invites additions. Its unexplored domains call for future research.

Presentation Abstracts

Priya Prithiviraj,
International Institute of Information Technology, Hyderabad

Not presented

Presentation Abstracts

Jinming Du, *University of Otago*

Not presented

Presentation Abstracts

EXPLORING CHATGPT'S POTENTIAL AS AUTOMATED ESSAY SCORING WITH MANY-FACETED RASCH MEASUREMENT ANALYSIS

Taichi Yamashita, *Kansai University*

Automated essay scoring has the great potential to reduce the workload of human raters while also providing second language writers with informative feedback. Thus, continuous evaluation is needed to provide evidence that supports such uses. As an alternative method to provide such evidence, the present study investigated the potential of many-faceted Rasch measurement (MFRM) analysis to evaluate ChatGPT-3.5 and ChatGPT-4.0 as automated essay scoring. The study used 80 human raters' ratings for 120 argumentative essays written by Asian English language learners, available in the International Corpus Network of Asian Learners of English (ICNALE). Additional data were collected by asking ChatGPT-3.5 and ChatGPT-4.0 to assign scores based on the rubric used in the ICNALE for complexity, accuracy, and fluency. MFRM analysis suggested that both ChatGPT-3.5 and ChatGPT-4.0 exercised the average severity, not lenient nor severe. It was also found that ChatGPT's ratings were too consistent and did not produce the moderate variance in ratings nor exercise the rater bias that is naturally assumed for human ratings. This limited human touch was slightly more noticeable for ChatGPT-4.0 than ChatGPT-3.5. These findings indicate the potential risk of using ChatGPT as automated essay scoring while also highlighting the benefits of MFRM as an alternative statistical analysis to evaluate automated essay scoring.

Presentation Abstracts

NAVIGATING THE FRONTIERS: UNVEILING THE IMPACT OF GENERATIVE AI ON SELF-ASSESSMENT IN COLLEGE WRITING

Hamidreza Moeiniasl, *University of Toronto*

This study employs a mixed-methods approach to investigate the integration of ChatGPT within college communication courses. Engaging 213 students from various language backgrounds and English proficiency levels, the research evaluates the tool's effectiveness in providing personalized, adaptive feedback on writing manuscripts, enhancing learner engagement, and meeting the diverse needs of English language learners. Participants utilized ChatGPT for grammar and vocabulary assistance, essay structuring, paraphrasing, summarizing, cultural insights, and assignment evaluation using rubrics. Comparison of initial and subsequent writing samples over 12 weeks revealed quantitative enhancements in overall writing scores by 87percent, particularly in the areas of word choice, grammar, paragraph and argument development. Qualitative analysis of the data collected through one-on-one interviews with students at the end of the semester illuminated enhanced user engagement, motivation, and personalized learning facilitated by ChatGPT. The study contributes insights into the benefits and challenges of integrating ChatGPT and its responsible integration into instructional and assessment writing practices. Specific findings demonstrate significant positive impacts on writing skills, potentially revolutionizing instructional and assessment methodologies. These insights offer valuable guidance for educators, policymakers, and stakeholders. Moreover, the findings highlight the potential of AI-driven technologies to promote inclusivity by catering to the diverse linguistic and cultural backgrounds of learners, thereby fostering a more equitable learning environment. Furthermore, the study sheds light on the evolving role of educators as facilitators of technology-mediated learning experiences and underscores the importance of fostering collaboration between educators, technologists, and educational researchers to harness the full potential of AI in education.

Presentation Abstracts

COMPARING CHATGPT AND HUMAN RATER SCORES IN MEASURING CONTENT IN CLIL PRESENTATIONS

Rie Koizumi, Saki Suemori & Yusuke Kubo,
University of Tsukuba

Considering the improvement of artificial intelligence (AI) in accessibility and response quality, this study examines the degree to which generative AI ChatGPT 4.0 (ChatGPT, hereafter) can be used in comparison with human raters to score content aspects in Content and Language Integrated Learning (CLIL) presentations. Previous studies (e.g., Mizumoto and Eguchi, 2023) compared ChatGPT with human raters and reported relatively comparable holistic scores. However, assessment of content aspects is challenging even for humans (Sato, 2024). Thus, the current study investigated the utility of ChatGPT for measuring content and compared it to that of human raters to explore the extent to which ChatGPT can be employed as a supplement for human raters. We obtained 52 presentation scripts from 40 university students who completed a CLIL course. Based on Sato (2024), we developed a content rubric with five criteria scored from 0 to 4: Task achievement, Clarity of content, Elaboration, Logicality, and Accuracy of content. We then adopted benchmarking scores based on the rubric for 12 scripts. We inputted the rubric, its explanation, example scripts, and the benchmarking scores in the ChatGPT before asking it to generate scores for 52 scripts and an explanation for each score. This procedure was repeated three times. Seven human raters also evaluated the scripts and reported when they were unsure of the scoring. We then compared the ChatGPT scores with human–rated scores using many–facet Rasch measurement with a rating scale model. The results showed that the global model fit to the data and other statistics were satisfactory and that human raters and ChatGPT produced consistent scores. However, ChatGPT tended to produce slightly lenient ratings, especially regarding the Logicality and Accuracy of content. Almost all rater–related biases were from humans. The implications for ChatGPT as a second rater for assessment of content are also provided.

Presentation Abstracts

DEVELOPMENT OF CHATGPT–POWERED DYNAMIC PRAGMATICS ASSESSMENT TASKS

Gi Jung Kim, *Iowa State University*

With a growing body of research on L2 pragmatics over decades, pragmatic competence has been positioned as an integral component in various language ability models (e.g., Bachman and Palmer, 1996, 2010; Canale and Swain, 1980; Celce–Murcia, Dörnyei, and Thurrell, 1995). This emphasis on the functional and social use of language also aligns with Korea's recently revised national English curricula (Ministry of Education, 2015, 2022), which list communicative functions and corresponding expressions to be taught. Despite its recognized importance, pragmatic competence has not been systematically assessed in large language tests or classroom assessments (Roever, 2011, 2022; Youn, 2015). To bridge this gap, this study aims to develop dynamic pragmatics assessment tasks in the Korean middle school classroom context. Given the dual purpose of classroom–based assessment for both learning and certification (Carless, 2007), this study adopts Arieli–Attali et al.'s (2019) expanded evidence–centered design framework, an assessment design framework in which learning and assessment are blended. Drawing upon this framework, the domain of assessment will be analyzed through a questionnaire distributed to Korean secondary English teachers and an analysis of Korean middle school English textbooks. In addition to the domain analysis results, Timpe Laughlin, Wain, and Schmidgall's (2015) pragmatic competence model will be used to define the construct of pragmatic ability in terms of pragmatic–functional knowledge, vocabulary, and grammar. To develop and prototype tasks to measure this construct, this study will also explore the potential of artificial intelligence algorithms for performance–based testing of second language pragmatics with back–and–forth real–time exchanges with learners (Van Moere and Downey, 2016). Consequently, the study will develop two role–play tasks focused on suggesting and requesting, wherein ChatGPT 4.0 serves as an interlocutor, feedback provider, and rater using its voice chat feature.

Presentation Abstracts

INVESTIGATING THE INFLUENCE OF CHATGPT WRITING GUIDANCE ON PEER REVIEW ASSESSMENTS

Danilo Calle, *Iowa State University*

In undergraduate composition courses, students often participate in peer review to enhance their writing skills. However, the impact of AI tools like ChatGPT on this process has not been fully explored. While previous studies have investigated the effect of AI on writing quality and revision strategies, very few have examined how AI feedback influences students' peer review assessments. Particularly, the influence of AI writing tools such as ChatGPT on students' criteria and language in peer review assessments has not been adequately studied, especially in first-year composition courses. To bridge this gap, an empirical study was conducted to investigate how ChatGPT's writing guidance influences students' peer review assessments in undergraduate composition courses. In this study, students were engaged in a feedback loop, wherein they drafted papers, received ChatGPT feedback, conducted peer reviews, and revised their papers accordingly. The data collected was analyzed to determine how their assessment criteria was influenced by their interactions with ChatGPT. In this session, the presenter will share the findings and insights from the study, discussing the implications of the results for teaching methods and the integration of AI tools in writing assessment. The study proposes that generative AI tools can enhance writing instruction by promoting critical thinking and corrective feedback in assessment-oriented collaborative tasks. This adds to discussions on the ethical use of AI in education. Understanding AI's impact on informing peer review can lead to improved teaching methods, curriculum design, and student preparation for the modern world where technology plays a significant role. Ultimately, attendees will gain insights into how AI tools such as ChatGPT can shape peer review assessments, informing their approaches to teaching and integrating AI in writing education.

Presentation Abstracts

EVALUATING THE SUITABILITY OF AI-GENERATED SPOKEN TEXTS FOR LISTENING ASSESSMENTS: A COMPARATIVE CORPUS-BASED ANALYSIS

Nazlınur Göktürk,

Republic of Türkiye Ministry of National Education

Despite the importance of using authentic texts in listening assessments for the validity of the interpretation and use of the assessment scores (Wagner, 2016), practical issues hinder the widespread use of these texts in many assessment contexts (Rossi and Brunfaut, 2021). One potential solution to this problem is to use large language models to generate listening texts that closely resemble authentic spoken texts (e.g., LaFlair et al., 2023). However, little research has been conducted to evaluate the appropriateness of using AI-generated spoken texts in listening assessments. The current research aims to address this gap by examining the comparability of AI-generated academic spoken texts with those from real-world academic settings and standardized language proficiency tests, with a focus on language use. Three distinct corpora will be created for a corpus-based register analysis (Biber and Conrad, 2019). The first corpus will comprise academic spoken texts in English generated by a GPT model, while the second corpus will be derived from an existing collection of academic spoken English (e.g., MICASE). The final corpus will consist of academic spoken texts sourced from official preparation materials of a standardized English proficiency test (e.g., TOEFL iBT). The linguistic features of the three corpora will be analyzed using multi-dimensional analysis (Biber, 1988). Results will reveal the extent to which the language of AI-generated academic spoken texts corresponds to the language of spoken texts produced in real-life academic settings and those used in standardized English proficiency tests. The results will have implications for the development of listening tests.

Presentation Abstracts

EXPLORING CHATGPT'S ABILITY TO RECOGNIZE AND GENERATE LANGUAGE VARIATIONS: NORTH VS. SOUTH KOREAN DIALECTS

Jean Young Chun,
Defense Language Institute Foreign Language Center (DLIFLC)

Recent advances in AI have sparked growing interest in language teaching and testing while raising the question of whether its outputs closely resemble human performance, providing appropriate inputs for language learners. Given the various dialects of the target language, ChatGPT's task of comprehending and producing multiple language variations across different regions can be challenging, which warrants empirical scrutiny regarding its accuracy. To address this question, this study explores ChatGPT's ability to recognize and produce North Korean dialect, crucial for the military Korean language program. In the study, ChatGPT-4 was initially tasked with distinguishing between North and South Korean passages to examine its recognition of North Korean words and expressions and then with providing their meanings. The recognition task included 100 North Korean and 100 South Korean passages, the latter serving as a baseline for comparison. These passages were extracted from textbooks spanning four terms, each containing increasingly difficult passages. The production task prompted ChatGPT-4 to convert South Korean passages used in the recognition task into North Korean passages using North Korean words and expressions. ChatGPT's passages were then compared to those rewritten by North Korean teachers in the program. For the analysis, a multiple logistic regression model was initially conducted to compare ChatGPT-4's ability to distinguish between North and South Korean passages, while controlling for passage difficulty. Subsequently, accuracy rates for word and expression meanings were calculated. The data from the production tasks were analyzed by tallying the occasions where ChatGPT generated target North Korean words and expressions. The study findings offer valuable insights for language teachers and testers on how to leverage ChatGPT in developing materials that incorporate dialects.

Presentation Abstracts

LEARNERS' DIALOGIC INTERACTION WITH CHATGPT: COULD IT BE AN INDICATOR FOR THEIR PROXIMAL DEVELOPMENT?

Tuba Özturan, *Erzincan University*
& Prithvi Shrestha, *The Open University*

Dynamic Assessment (DA) is predicated on the dialogic interaction between teachers and learners, emphasizing the fusion of assessment and instruction through dialogic feedback. This process-based approach aims to uncover learners' Zone of Proximal Development by establishing a personalized learning-assessment setting (Poehner, 2008). Despite its promising contributions, DA in actual classrooms can be time-consuming and overwhelming for most teachers, especially in L2 writing classes, given that writing inherently involves a multifaceted process. To address this issue, this presentation aims to showcase the interaction between learners and ChatGPT 3.5, an underexplored generative AI tool in DA. Participants are L2 writers in an EFL setting, and the targeted genre is argumentative writing, which has been deliberately chosen due to the existing body of research in this area, enabling comparisons to be made with previous findings. For four weeks, the learners produced argumentative texts on different topics (one text per week), and they used ChatGPT as a personalized feedback and instruction provider. They initially used the same prompt, and following the ChatGPT-generated feedback, participants were free to seek further assistance. After each session, they shared the chat-log with the researchers, and stimulated recall sessions were conducted to delve deeper into the learners' interaction, their underlying reasons for seeking additional help, and potential changes in their prompts over time (such as a growing preference for more explicit or detailed prompts, or the emergence of advanced abilities after receiving personalized feedback and instruction from ChatGPT). We will report on the impact of learner inteeractions with ChatGPT on the participants' writing development, the nature of dialogic feedback (examining the learners' requests/prompts for additional assistance) and learner perceptions. The results will shed light on the potential of generative AI as a valuable tool to assist teachers in DA and provide insights for future studies.

Presentation Abstracts

EVALUATING AI-GENERATED PROMPTS IN CILS DUE-B2 EXAMS: A COMPARATIVE VALIDATION STUDY

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

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 – University for Foreigners of Siena) 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 examining 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, methodology and the findings from the pilot administration conducted, promoting a discourse that will pave the way for more extensive future research.

Presentation Abstracts

CAN CHATGPT EVALUATE READING ITEM DIFFICULTY AND ITEM DISCRIMINATION WITHOUT RESPONSE DATA?

Hongli Li & Chad Marchong,
Georgia State University

ChatGPT, a leading generative AI model introduced in November 2022, has gained widespread recognition for its capabilities in developing test items (Bezirhan and von Davier, 2023; Kiyak et al., 2004). This study aims to assess the effectiveness of ChatGPT-4 in evaluating item characteristics of a reading comprehension test without response data. This test is tailored to assesses the advanced-level English-language competence of nonnative speakers of English. It includes four passages, each accompanied by five multiple-choice items, totaling 20 items. Our approach involves soliciting ChatGPT's assessment of item difficulty and item discrimination for each question, along with the rationale behind its judgments. We also varied the prompts provided to ChatGPT to explore how different inputs influence its output. Simultaneously, we conducted an item response theory (IRT) analysis to estimate the item difficulty and discrimination parameters for each item based on response data from 2,019 examinees from a previous test administration. Results from the traditional IRT analysis will serve as a benchmark against the ChatGPT evaluations. Our preliminary findings indicate some agreement between ChatGPT's evaluation of item difficulty and the IRT item difficulty parameters. These results offer valuable insights into the generative AI's role in facilitating test development and test evaluation process, highlighting its potential to enhance assessment efficiency.

Presentation Abstracts

REGISTER–DRIVEN PROMPT DESIGN AND EVALUATION OF GPT GENERATED STIMULI FOR ASSESSMENT

Geoff LaFlair, *Duolingo*
Andrew Runge, *Duolingo*
Jesse Egbert, *Northern Arizona University*
& Yağmur Demir, *Northern Arizona University*

The use of GPT in content generation for assessment and learning continues to grow. One of the primary assumptions underlying the use of this tool is its ability to automatically generate texts that are representative of target registers. The degree to which it delivers on this assumption is an empirical question in the areas of language teaching and assessment, where AI–generated language has been proposed as a means of creating text passages for language learning (e.g., Shin and Lee, 2023) and assessment (e.g., Wodzak, 2022). Given the well–established role of register as a predictor of linguistic variation (see Biber and Egbert, 2023), it follows that the success of AI–produced texts depends on the degree to which they produce language in register–appropriate ways. Building on previous work in this area (Berber–Sardinha, 2023), we investigate the degree to which GPT generates texts with register–specific and functionally–appropriate linguistic features. We propose a register–driven approach to prompting GPT models to create listening stimuli for language assessment. This approach is similar to "tree of thought" prompting (Yao et al., 2023) and draws upon register–based corpus linguistics for prompting GPT. Specifically, we vary situational parameters—namely communicative purpose, discipline, social roles, and topic—in order to add domain relevant language variation to the outputs of the GPT model (following Biber and Conrad, 2019). This creates a theoretically motivated generation framework the outputs of which can be quantitatively evaluated for its ability to create stimuli that capture relevant domain variation. We answer the following research question: To what extent does GPT produce functionally–appropriate and register–aligned language? The construct of register–alignment is operationalized and evaluated through multi–dimensional analysis (Biber, 1988). Functional appropriateness is operationalized as human perceptions of the output's apparent effectiveness in achieving its intended goals. Results from this study will inform automated content generation assessment and teaching.

Presentation Abstracts

THE POTENTIAL OF GPT-4 IN LANGUAGE TEST PREPARATION: INVESTIGATING READING PASSAGE GENERATION WITH ZERO-SHOT PROMPTING

Zhengqing Luo, *Beijing Normal University*

This study examines the application and efficacy of using OpenAI's GPT-4 for zero-shot prompting in generating reading passages suitable for language assessments, comparing these AI-generated texts to human-written passages used in educational contexts (taking English reading materials in Zhongkao as an example). Addressing how GPT-4 can create tailored reading materials and the comparative quality of these materials, the research employs a mixed-method approach, including textual analysis, questionnaires, and stimulated recall interviews. In the initial stage, detailed prompts, which incorporate characteristics such as text type, form, multimodality, topics, and readability, were devised and fed into GPT-4, leading to the generation and subsequent selection of texts. The author, together with an in-service junior high English teacher, examined the generated texts one by one to decide whether they were consistent with the required standards. When the retention rate reached 85percent and the prompt produced a reliable text pattern, the study moved on to the second stage of a blinded evaluation where junior high school English teachers assess both AI-generated and human-written passages without knowledge of their origins. Altogether 103 in-service teachers participated and results indicate that AI-generated texts are comparable to human-authored ones, with some even surpassing in engagement and readability. Last, further insights are gained through stimulated recall interviews with a subset of respondents, revealing cautious optimism about the integration of AI in educational settings, alongside concerns about losing the human touch in content creation. The findings illustrate that GPT-4 can effectively produce reading passages that align with traditional educational standards and that these AI-generated passages can potentially match or even exceed the quality of human-written texts. The study highlights the promise and challenges of integrating AI into the fabric of educational testing and content creation, offering crucial insights for future applications of AI in educational assessment.

Presentation Abstracts

CHATGPT FOR INTERACTIVE WRITTEN CORRECTIVE FEEDBACK IN FRENCH AS SECOND LANGUAGE LEARNING AND TEACHING

Taegan Holmes, *University of Ottawa*

AI tools capable of generating original content based on databases are becoming increasingly widespread, and consequently it is imperative to take a critical look at their potential impact in educational settings. The interactive nature of generative artificial intelligence (GAI) tools enables the user to enter into a conversation with a machine (OpenAI, 2023). ChatGPT is a GAI tool that is widely available and easily accessible. This empirical and exploratory study focuses on its use for interactive written corrective feedback (WCF) in French as a second language (FSL) teaching and learning. Students submitted a short text to ChatGPT and then engaged with the tool to obtain WCF. Students were asked to enter into a conversation with ChatGPT in order to understand its proposed corrections and they were encouraged to challenge ChatGPT if they felt that the tool had made an error, thus furthering the interactive nature of the correction process. Transcripts of students' interactions with ChatGPT were analyzed using a taxonomic approach (Bilmes, 2009) situated in a qualitative content analysis (Selvi, 2019) to assess how students interacted with the tool. Students also filled out a pre-task questionnaire about their digital writing and correcting practices as well as a post-task questionnaire pertaining to their experience using ChatGPT for WCF. Findings from this study suggest that the textual revision practices of the participants are heterogeneous, yet a majority of students expressed positive sentiments towards the experience in the post-task questionnaire. This research is of critical importance seeing as the use of GAI for WCF is a burgeoning domain (Fang et al., 2023; Godwin-Jones, 2022; Rudolph et al., 2023; Wu et al., 2023) and preliminary studies published thus far have identified ChatGPT as having much potential in various natural language processing tasks, including grammatical error correction (Fang et al., 2023).

Presentation Abstracts

ARTIFICIAL INTELLIGENCE AND SECOND LANGUAGE WRITING: UNDERGRADUATE WRITERS' USAGE AND PERCEPTIONS OF GENAI TOOLS' ETHICALITY

Kübra Çekmegeli, Juan M. Rostrán Valle & Matt Kessler,
University of South Florida

The use of generative artificial intelligence (GenAI) tools in language teaching is a critical topic that has captured researchers' and teachers' attention (e.g., Kohnke et al., 2023; Pfau et al., 2023). Within second language (L2) writing, in particular, GenAI tools have become increasingly prevalent and valuable, as recent studies have examined the impact of GenAI tools (e.g., ChatGPT) when integrated with learners in the L2 writing classroom (e.g., Guo et al., 2023; Yan, 2023; Zou et al., 2023). However, there is limited research exploring issues involving students' self-initiated uses of GenAI chatbots for L2 writing, especially in languages other than English (LOTEs). Additionally, there is a lack of investigation into these students' perceptions regarding the ethicality of using such tools in their classrooms. Using a survey and semi-structured interviews, this mixed-methods study addresses that gap by examining two research questions: (1) To what extent do L2 writers utilize GenAI for writing tasks in LOTEs (and what are their current practices with such tools)? (2) How do learners of LOTEs perceive the ethicality of using GenAI tools in L2 writing classrooms? To address the research questions, we recruited a total of 287 L2 learners with varying proficiencies from Chinese, French, Italian, Japanese, and Spanish classes offered at a large US university. The findings demonstrate that 24.1percent of students reported actively using GenAI tools for writing in LOTEs. However, such students reported diverse uses of GenAI, including for 12 different purposes (e.g., corrective feedback, writing sentences, and learning to use words/phrases appropriately in context). Additionally, the results suggest that students hold a wide range of views regarding the ethicality of such tools. This paper enriches existing GenAI research in L2 writing, offering insights and recommendations for educators.

Presentation Abstracts

CHATGPT IN ESL WRITING: L2 LEARNERS' PERSPECTIVES

In Young Na, Mahdi Duris & Volker Hegelheimer,
Iowa State University

Automated Writing Evaluation (AWE) systems, such as Grammarly and ETS's Criterion, have proven beneficial for reducing instructors' workload and enhancing L2 learners' writing through immediate and frequent feedback (Warschauer and Grimes, 2008). With the recent release of generative AI chatbots like ChatGPT, many language learners are now able to engage with a simple chat interface for various writing tasks. This development prompts a need to explore how learners interact with ChatGPT and assess its effectiveness as an AWE tool for L2 writers. This exploratory study examined the language learning potential (Chapelle, 2001) of ChatGPT through a case study involving eight international students from two university-level ESL writing courses at different levels. Screen-capture recordings, interviews, and surveys were used to examine the nature of students' interactions with and their views on ChatGPT for their L2 writing across three revision sessions. We specifically looked at learners' prompts, the nature of feedback received, and the extent of the revisions they made — minimal, moderate, and substantial — based on the AI feedback. Findings indicate that the learning potential of ChatGPT greatly depends on the specificity of learner prompts and ChatGPT's feedback. Although students showed varying degrees of engagement and used different prompting strategies, most employed ChatGPT to produce revised drafts, often copying, pasting, and selectively replacing some words from the generated text. The insights from this study will better inform educators on how to integrate ChatGPT more effectively into classrooms by tailoring strategies that maximize the benefits of ChatGPT for language development. Full details of our findings and specific pedagogical recommendations will be presented.

Presentation Abstracts

EXPLORING THE EFFICACY OF CHATGPT INTEGRATION ON EFL WRITING: A STUDY WITH VIETNAMESE UNDERGRADUATE STUDENTS

Duong Nguyen, *Iowa State University*
& Trang Ho, *Da Nang University of Economics*

Research exploring the integration of ChatGPT in L2 teaching and learning has proliferated. As technology continues to shape language education, there is a growing need to critically examine how innovative tools like ChatGPT contribute to the linguistic development of second language (L2) learners. While existing studies have offered insights into learners' perceptions of ChatGPT in writing classrooms (e.g., Yan, 2023; Kwon et al., 2023), questions persist regarding concrete assessments of its impact on L2 writing development. To fill this gap, this study aims to investigate how ChatGPT enhances EFL undergraduate students' academic writing abilities at a university in Central Vietnam. A mixed-methods quasi-experimental study (Creswell and Clark, 2011) with the pretest, posttest, and delayed posttest design, involving 60 Vietnamese undergraduate students, was conducted to explore the impact of using ChatGPT 3.5 for feedback and revision on their writing performance. Two intact classes were randomly assigned to a control and a treatment group who received identical writing instructions, with the latter being trained in ChatGPT use for seven weeks, focusing on the feedback they received for vocabulary and grammar of their essays, using the prompt provided. Screen recordings of students interacting with ChatGPT for feedback and revision were also documented. Writing development was measured by syntactic and lexical complexity and fluency in EFL learners' writings over three stages. Repeated measures ANOVA was used for quantitative analysis. Furthermore, a questionnaire examining the participants' perceptions of ChatGPT was conducted. Semi-structured interviews and screen recordings of students' interaction with ChatGPT were analyzed for qualitative data. The preliminary findings suggest an improvement in lexical rather than syntactic complexity and a general positive attitude towards the use of ChatGPT in feedback and revision processes. This study enhances understanding of ChatGPT's efficacy in L2 writing pedagogy, providing clear guidance on using ChatGPT properly for language development.

Presentation Abstracts

CAN CHAT–GPT PRODUCE CORRECTIVE FEEDBACK LEARNERS CAN USE TO SELF–CORRECT THEIR L2 TEXTS?

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 and 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.

Presentation Abstracts

CHATGPT AS A METALINGUISTIC AWARENESS RAISING TOOL FOR SUPPORTING ESL STUDENTS' ACADEMIC SUMMARY WRITING SKILLS

Marcin Kleban,

Jagiellonian University, Krakow, Poland

Writing academic summaries is a fundamental skill for university students, requiring them to synthesize content, structure arguments, and select key information effectively. Comparative synthesis, which involves analysing multiple texts, demands a critical overview of the issues discussed within them. At Jagiellonian University in Krakow, Poland, English department students receive instruction on developing summary writing skills in English as part of their academic writing course, with these skills evaluated in end-of-semester examinations. This study explores the potential of ChatGPT as a tool to enhance students' metalinguistic awareness—defined as the ability to notice, reflect on, and manipulate language to convey intended meanings (Alipour, 2014). While existing research has predominantly focused on ChatGPT's role in providing feedback for second language writing (e.g. Athanassopoulos et al., 2023), this study takes a reverse approach, utilizing ChatGPT as a language input subjected to learner transformation. In a small-scale investigation, advanced MA programme students were tasked with reading and analysing a ChatGPT-generated summary of two articles on a similar topic, reflecting on its language features. Subsequently, they wrote their own summaries based on the AI-generated text, subsequently listing useful phrases, lexical items, and grammatical or structural elements from ChatGPT's output. Data collected from this process was compared with students' unassisted academic summaries produced during end-of-semester examinations.

Presentation Abstracts

LEARNER AGENCY IN THE AGE OF GENERATIVE AI: A STUDY OF L2 WRITING SKILLS IN HIGHER EDUCATION

Sibel Söğüt, *Sinop University*

The rapid emergence and growing prevalence of GenAI tools have provided opportunities and challenges, particularly concerning learner agency in tertiary-level English language learning. The overreliance on GenAI tools specifically in L2 writing skills and text generation has raised ethical concerns and negative impacts on the development of learner agency. This ongoing study explores the intersection of learner agency and GenAI in the context of L2 writing skills development in higher education. This study frames learner agency as a dynamic, complex, and multifaceted element that is influenced by contextual factors and learners' past learning experiences. This study includes 80 L2 learners in an L2 context. A mixed-methods approach is used to gather data on learner demographics, past engagement with GenAI use, current dispositions towards GenAI integration, and the factors impacting learner agency in GenAI-supported L2 writing contexts. Descriptive statistics and thematic analysis are carried out in the data analysis to document a composite picture of different factors that influence learner agency. The findings of the study will address broader instructional implications for L2 learners, specifically the knowledge and skills needed to establish a balance between learner agency and GenAI use. The study will also offer further insights and recommendations for fostering learner autonomy and cultivating the sense of learner agency in the age of GenAI.

Presentation Abstracts

THE EFFECTS OF GENERATIVE AI IN SCAFFOLDING L2 WRITING PROCESSES

Jianling Liao, *Arizona State University*

In L2 writing classrooms, the process writing approach has received insufficient attention due to various reasons. A product–driven writing approach has significantly reduced the communicative nature and social purposes of L2 writing and made the learning of writing less relevant to L2 learners (for example, compared to oral communication). The existence of various generative AI tools may hold the potential for providing interactive intervention to L2 writing processes and making L2 writing practice more dialogic. The current study investigates the effects of generative AI in scaffolding L2 Chinese learners' writing process. Fifteen college advanced learners of Chinese participated in the study. Participants were guided to interact with a generative AI tool to develop an argumentative essay. For each stage of the writing process, including planning, drafting, and writing, students interacted with AI using question prompts with specific specifications to elicit output and guidance on language use, ideas, and discourse structure. Students also resorted to AI for revision and editing suggestions after they completed the first draft. Based on the input from AI, students developed, revised, and finalized their essays. Participants' question prompts, the input generated by the AI, and learners' first and final written drafts were analyzed to understand the interactive processes between the AI and learners and the learners' processing mechanisms of the AI's input. Selected learners were interviewed for their perceptions of the roles AI played in their writing process. Results showed that the interactive processes with AI fostered a dialogic process of L2 writing. Particular interactive strategies with AI elicited more effective guidance and scaffolding than others.

Presentation Abstracts

CHATGPT AS A RESOURCE FOR ACADEMIC WRITING: STUDENTS' AND AN INSTRUCTOR'S PERCEPTIONS

Susan Parks,
Université Laval (Quebec City, Canada)

Although the potential of ChatGPT (and other such generative AI tools) for developing writing skills has frequently been evoked (Goodwin-Jones, 2022), research related as to how such tools might be integrated into classroom instruction is scant. To explore this issue, the present study focused on a group of pre-service TESL students who were enrolled in an academic writing course in a francophone university in Quebec (Canada) and concurrently in their first practicum. Most of these students were advanced ESL learners. The main assignment of the course involved writing up a report on an action research project (ARP) using data collected by the student teachers as they observed and took notes on how their cooperating teachers used the L2 and the L1 with ESL students in elementary and high school settings. As the TESL students progressed through the project, they were invited at various points to use prompts written by the instructor (the presenter of this paper) to engage with ChatGPT for feedback on content/organization and/or language. Throughout the course, students also received peer and instructor feedback; ChatGPT was viewed as one additional resource. All process related feedback, including the ChatGPT conversations, were archived in the students' Google folders. Following the submission of their ARP, students were asked to complete a survey questionnaire in order to determine how they perceived ChatGPT as a source of feedback. Analysis of the feedback provided by ChatGPT was also undertaken to better understand the students' comments. Results of both the students' and the instructor's perceptions of the usefulness and limitations of ChatGPT as a resource for improving writing skills will be discussed.

Presentation Abstracts

ENHANCING LANGUAGE PROFICIENCY WITH GENAI: INVESTIGATING THE IMPACT OF CORRECTIVE FEEDBACK IN ACADEMIC WRITING

Ali Ebrahimpourlighvani, *Iowa State University*

In the evolving landscape of language acquisition, the integration of artificial intelligence (AI) presents a transformative potential for enhancing linguistic competencies. This study delves into the efficacy of GenAI's corrective feedback in prompting the noticing of linguistic forms among university students, irrespective of English being their first (L1) or second language (L2). Previous research has explored AI's role in language learning, yet a comprehensive understanding of its impact on developing linguistic skills remains uncharted territory. Addressing this gap, the research aims to demonstrate how GenAI's feedback fosters students' ability to identify and correct errors, thereby facilitating the internalization and subsequent application of correct linguistic forms. Employing a mixed-methods approach, the study juxtaposes observational insights with quantitative data from pre-test, post-test, and questionnaire. Anticipated outcomes suggest that students will not only improve in linguistic accuracy but also exhibit heightened awareness of language structures. The real-world benefits of such advancements extend to academic and professional settings, where effective communication is paramount. Emphasizing the importance of this research, it contributes to the pedagogical discourse on AI-assisted language learning, offering novel perspectives on educational technology. Attendees of this session can expect to gain valuable insights into the practical applications of AI in language pedagogy, equipping them with knowledge to navigate the future of language instruction.

Presentation Abstracts

LEVERAGING CHATGPT-4 FOR ENHANCED SPANISH LANGUAGE LEARNING: INSIGHTS FROM THE INTERACTIVE NARRATIVE 'ESCAPE THE HAUNT

Celia Bravo, *University of Chicago*

This study explores the application of ChatGPT-4, a generative AI model, in facilitating second language (L2) learning for Spanish learners through the interactive narrative "Escape the Haunt". Utilizing a corpus linguistics framework, the research analyzes a dataset comprising over 500,000 words generated by both the AI and student interactions within this narrative setup. The primary focus is on the grammatical quality and stylistic appropriateness of the AI-generated Spanish as it dynamically responds to learner inputs in a contextually rich, game-based learning environment. "Escape the Haunt" challenges learners to navigate through a series of tasks and puzzles in Spanish, requiring them to interact extensively with ChatGPT-4. This interaction provides a unique dataset for analyzing how generative AI copes with the intricacies of Spanish grammar and lexicon in a constrained thematic setting. The study assesses the language produced by ChatGPT-4 against standard linguistic descriptors of accuracy, syntactic complexity (at the sentential and clausal levels) and lexical complexity (diversity and sophistication) and compares it to learner language development over the course of the narrative. Further, this paper discusses the pedagogical implications of using generative AI in language learning, focusing on the AI's ability to adapt linguistically to the learners' proficiency levels and metalinguistic prompts. By examining the interaction patterns and the quality of the responses from both the AI and the students, the study provides evidence on the potential of generative AI to augment traditional language learning methodologies. It also addresses broader questions about the role of AI in shaping language teaching strategies and the necessary competencies learners must develop to effectively utilize these advanced technological tools in language education.

Presentation Abstracts

ALGERIAN EFL LEARNERS' VIEWS ON REPLIKA GENAI CHATBOT FOR SPEAKING SKILLS SELF-ASSESSMENT

Aissa Berregui & Mohammed Naoua,
University of El Oued, Algeria

Innovations in the field of generative artificial intelligence (GenAI) are leading an unprecedented revolution at all levels, and language learning is no exception. This qualitative study examines the perspectives and experiences of Algerian English as foreign language learners regarding the use of GenAI chatbot Replika for the sake of self-assessment of speaking abilities. Based on semi-structured interviews and focus group discussions, data were gathered from a cohort of 20 Algerian EFL learners who employed Replika to assess their English speaking Proficiency. The thematic analysis of the data yielded several substantial findings. Firstly, most participants showed favourable attitudes towards the Replika application for self-assessment due to its accessibility and convenience. Secondly, learners liked the personalized feedback generated by Replika which enabled them to identify areas of improvement in their speaking skills. Thirdly, participants noticed a positive impact on their motivation and confidence in learning English because of the friendly and non-judgmental nature of the chatbot design. Nevertheless, some learners raised concerns regarding the accuracy and reliability of Replika's assessment mechanisms, which, as per learners' suggestions, require further refinement in feedback provision. Overall, this research accentuates the ability of GenAI tools like Replika to facilitate self-assessment of oral proficiency among English as a Foreign Language learners. Additionally, the study highlights areas for improvement and revision. These findings have many implications and contribute to the continuous debate on incorporating GenAI tools in language learning and assessment.

Presentation Abstracts

A GENERATIVE AI TOOL FOR DIALOGUE DATA COLLECTION IN ONLINE ENVIRONMENTS

Emma Caputo, *University of Barcelona*

This study explores the development and implementation of a generative AI tool for collecting online audio data to assess fluency, intelligibility, and comprehensibility of spoken L2 English among frequent players of online video games. A pilot study revealed that participants experienced foreign language speaking anxiety and demonstrated a strong preference for computer-mediated communication over online interaction with humans. In response to these challenges, we created a conversational agent that imitates dialogue and collects audio data for research purposes. The program is accessible via a webpage, utilizing the browser API to collect audio data. The participant's audio is then transcribed by whisper.cpp, and the resulting text is input into a libre and self-hosted Large Language Model running on llama.cpp. The output of the LLM, which acts as the dialogue partner, is then fed into TTS software, and sent back to the browser. This results in an interactive dialogue with the participant. The program is designed using libre software to ensure independence from commercial interests, compliance with European privacy data laws, and complete control of data throughout the process. Preliminary tests indicate that programs such as this could enhance participants' comfort and willingness to communicate while reducing their speaking anxiety. It also could replace the unknown, variable biases of human conversation partners with a bias that is more stable, observable and reproducible, and thus easier to control for. As an application of a novel technology, the use of libre, self-hosted generative AI in collecting online dialogue data requires future research, particularly in evaluating the reliability compared to actual dialogue situations between humans. However, this study suggests that generative AI can be a valuable tool for collecting audio data from participants in online and particularly closed environments.

Presentation Abstracts

TEACHING GLOBAL ENGLISHES WITH CHATGPT: MODELING AND ANALYZING LANGUAGE VARIETIES WITH GENERATIVE AI

Peter Sayer & Ivan Stefano,
The Ohio State University

An important aspect of the multilingual turn in TESOL (May, 2014) has been to recognize and incorporate in language teaching the wide range of varieties of English that, collectively, are referred to as Global Englishes. The rationale is to diversify the varieties of English (and English speakers) that are held up as linguistic models for students is to prepare them for the "real world" where they are likely to encounter English in all shapes and sizes. Whereas TESOL teachers used to debate the relative merits of teaching British or American English as the norm, scholars (Kachru, 1992; Kirkpatrick, 2007) have argued that English teaching should not be centered around the native speaker model. Instead, TESOL should contribute to broader project of legitimizing global varieties of English. Matsuda (2003) explained that "the exposure to different forms and functions of English is crucial for [global English] learners, who may use the language with speakers of an English variety other than American and British English [and] an awareness of different varieties helps students develop a more comprehensive view of the English language" (p. 721). English teachers nowadays have a plethora of digital resources to draw on to provide students with input in a vast range of different English varieties, including geographical dialects and standard and vernacular varieties. In this chapter, we will present ChatGPT as another excellent tool for modeling for students sociolinguistic differences in Global Englishes. Whereas YouTube videos are invaluable resources for showing real-world language use, we argue that ChatGPT allows the learner to hone in on specific language choices and linguistic features of English varieties. We begin this chapter by looking at what ChatGPT does when prompted to generate several different examples of Global Englishes. We then show some prompt engineering techniques (Brownell, 2023) whereby the particular features that ChatGPT uses to model varieties are made explicit. We conclude by offering some pedagogical ideas for how teachers can use this approach to get their students to use ChatGPT to better understand and explore Global Englishes.

Presentation Abstracts

IMPROVING SPEAKING SKILLS OF EFL LEARNERS THROUGH AI-SUPPORTED ACTIVITIES

Ayse Kizildag, *Aksaray University*

The revolution in the sphere of artificial intelligence (AI) through generative modules of large language models (LLM) brought novel and stimulated milieu for instructional processes. Such AI tools that could easily be used for language learning (Baidoo-Anu and Ansah, 2023; Koh and Doroudi, 2023) have also been discovered very recently with the potentials to engage pre-service teachers for employing fresh teaching strategies by integrating them into their lesson plans during the practicum (Luo et al., 2024; van den Berg and du Plessis, 2023). Furthermore, including AI into English as a Foreign Language (EFL) pre-service practice teaching is thought to be improving the quality of training by addressing the constraints of traditional methods to teach English (Mohamed, 2023). Offering a technologically enriched learning environment, this paper attempts to explore how a group of EFL teachers experience the newly marketed AI-tools to improve speaking skills of their students at a practicum school. Designed as a case study, this qualitative research is conducted with six EFL prospective teachers (four females and two males, aged between 23 and 26) in their final year, registered at a Turkish public university, English Language Teaching (ELT) Department. They designed speaking activities with the application of some certain AI tools during their practice teaching within the twelve-week practicum course of spring term in the academic year of 2022/23. Narratives from the weekly mentoring meetings, written reflective reports on the field experiences and the lesson plans of practice teaching formed the data sources; data were examined utilizing thematic analysis framework developed by Braun and Clarke (2006). Preliminary findings are twofold. While participants highly embraced AI tools in utilizing at different stages of their teaching, i.e., warm up, practice, production, and assessment, they reported high rates of learner engagement towards their practice-teaching. Keywords: generative artificial intelligence, large language models, chatbots, EFL, ELT, pre-service teacher education.

Presentation Abstracts

EXPLORING L1 BIAS IN THE AUTOMATED SPEECH RECOGNITION ABILITY OF AI TECHNOLOGY ACROSS DIVERSE ENGLISH ACCENTS

Yuna Bae, *Northern Arizona University*

Artificial intelligence (AI) technology emerges as a promising solution to mitigate the influences affecting listeners' judgments. However, there remains a dearth of empirical evidence verifying the impartiality and fairness of AI. Given that machines have exhibited biases concerning gender (Tatman, 2017), race (Koenecke et al., 2020), and regional dialects (Lima et al., 2019), it is crucial to investigate whether biases related to L2 accent varieties may also manifest in AI systems. Therefore, the present study aims to explore the bias in the automated speech recognition (ASR) ability of AI across diverse English accents. The current study involved two data sets. The first set of speech files (Dataset 1) consisted of 30-second recordings of TOEFL listening passages, with each passage recorded by twelve speakers. Among these speakers, three represented each of four different English varieties: Chinese Mandarin, Indian Hindi, Mexican Spanish, and South African. The second set of speech materials (Dataset 2) was obtained from speech repositories, resulting in 15 speech files for each L1 variety. After transcribing speech samples using two AI devices, Siri and Google Assistant, an expert coder calculated the word error rate (WER) for each transcription result from the two datasets. The results of a 2 (AI) x 4 (L1) factorial ANOVA reported a significant effect of L1 on WER ($F(3, 112) = 25.767$, $p<.0001$). Post hoc tests revealed that, for both Siri and Google Assistant, WER for Chinese-accented speech was significantly higher than for South African ($p<.0001$), Indian ($p<.0001$), and Mexican Spanish ($p<.0001$) accents. Conversely, Indian and Spanish accents showed comparatively lower WER, indicating potential L1 bias in AI transcription. The findings provide implications for the suitability and effectiveness of AI across diverse linguistic contexts, suggesting careful incorporation of AI technology in L2 learning and teaching.

Presentation Abstracts

UNLOCKING THE POTENTIAL OF GENERATIVE AI: ENHANCING SPANISH LANGUAGE LEARNING THROUGH HUMOR TRANSLATION STRATEGIES

Cristina Pardo-Ballester, *Iowa State University*

This study employs a mixed-methods approach (Creswell and Plano Clark, 2011) to examine the potential of integrating generative AI and translation techniques to enhance Spanish language acquisition in a translation course. The research targets intermediate-mid proficiency learners, comprising 25 third-year American participants studying Spanish. It delves into the capabilities of AI tools (ChatGPT, Copilot, and Gemini) in enhancing translation learning. While recognizing the potential limitations of these tools in performing creative tasks, such as humor translation, the study provides valuable insights into effective strategies for nurturing creative translation skills. Quantitative analysis of student translations of humor-rich jokes assesses dynamic equivalence (Nida, 1964) and cultural adaptation (Venuti, 2008) strategies' effectiveness, while qualitative insights from student journals reflect on their experiences (Schön, 1987) with translation tasks and AI tool usage. The assessment focuses on whether these tasks present a sufficient challenge to participants and whether they contribute to enhancing translation skills, the development of cultural sensitivity, and comprehending humor in a second language. Inspired by the General Theory of Verbal Humor (Dore, 2021; Attardo, 1994; Chiaro, 2017), the study analyzes the linguistic capabilities of AI models and their implications for translation and language learning. The investigation into generative AI is crucial as it offers a unique opportunity to explore the boundaries of machine translation and its potential to support language learning (Pardo-Ballester, 2022). By examining how participants interact with AI-generated translations of humor, which often lack creativity, we gain insights into effective strategies for fostering creative translation skills. This research contributes by examining generative AI's role in translation education, investigating humor translation's fostering of creativity, and understanding student experiences with AI tools.

Presentation Abstracts

A SEMI-SYSTEMATIC REVIEW OF METHODOLOGICAL ASPECTS IN EARLY GENAI RESEARCH FOR LANGUAGE LEARNING: FOCUS ON STUDIES PUBLISHED IN 2023

Anne-Marie Sénécal, *Université de Sherbrooke*

Generative Artificial Intelligence (GenAI) has emerged as a groundbreaking innovation, revolutionizing education by offering new perspectives and opportunities. In SLA, GenAI presents valuable language-learning affordances, such as providing opportunities for writing practice and support for aspects such as reading comprehension (Barrot, 2023). While GenAI shows promise in enhancing the learning experience, concerns regarding pedagogy and ethics, particularly academic integrity, persist (Rudolph et al., 2023). Since the introduction of ChatGPT in November 2022, there has been a notable surge of publication in Artificial Intelligence in Education (Van Dis et al., 2023). This project aims to examine the current state of research of the pedagogical implementation of GenAI in SLA by reporting on a semi-systematic review of 12 primary studies published between November 2022 and October 2023. This review focuses on themes, participant characteristics, and methodologies employed. Two main themes emerged: perceptions regarding GenAI's application in SLA and practices associated with learning and assessment. The two key data-collection methods identified, survey and document-based research, align with these themes and reveal insights into current research trends and methodological approaches. This review not only interprets the findings but also anticipates a shift towards investigating specific relationships, such as the impact of GenAI on learning. Thus, this review provides valuable insights into the current state of research in GenAI's role in SLA, paving the way for future investigations.

Presentation Abstracts

/ˈFAʊNTƏN/ OR */ˈFAʊNTAɪN/: PRESERVICE ENGLISH TEACHER ENGAGEMENT WITH AI AND TTS FOR PRONUNCIATION SELF-LEARNING

Alba Paz-López, *University of Málaga*
Boris Vazquez-Calvo, *University of Barcelona*
& Kevin Randall Steil, *University of Barcelona*

As AI becomes increasingly prevalent, studies such as Moorhouse (2024) investigate teacher readiness and adoption of AI in language education. However, the use of text-to-speech (TTS) technology in English as a Second Language (ESL) education, particularly for pronunciation improvement in language teacher education programs, remains underexplored. This study aims to address this gap by examining how preservice English teachers engage with AI and TTS for pronunciation self-learning. Using a multi-site exploratory approach, the study involved two cohorts from Primary Education Teacher Training programs in Catalonia and Andalusia, Spain, comparing a Pronunciation cohort (Catalonia), which received pronunciation instruction besides language pedagogy theory and practice, with a Pedagogy cohort (Andalusia), focusing on language pedagogy theory and practice. The research sought to evaluate TTS technology's effectiveness in enhancing pronunciation skills through questionnaires, journal entries from participants practicing with TTS at home (Pronounce, Microsoft Studio, Speech Ace), and some TTS-evaluated recordings, concluding with a Q-sort. Thirty individuals volunteered for the study. Nonetheless, the data collection process encountered disparities in participant engagement, resulting in partially incomplete datasets from some participants. Despite this, the study offers insights into participants' learning trajectories and views on TTS for pronunciation practice. Preliminary findings show mixed results. While most participants noted improvements in confidence and self-perceived pronunciation skills, the impact of TTS-assisted training, the need for more structured feedback, and the prevalence of rather lenient assessment to ensure user engagement rather than improvement (especially at suprasegmental levels of language) highlight the need for further investigation. The study underscores TTS technology's potential in language (teacher) education, suggesting how integrating AI into pronunciation training could benefit learners, teachers, and teacher educators to complement classroom-based language practice with AI-mediated oral interaction and assessment.

Presentation Abstracts

CHATGPT, L2 PRAGMATICS, AND CRITICAL AI LITERACY

Robert Godwin-Jones,
Virginia Commonwealth University

The ability of generative AI to produce language that closely resembles human-produced speech has led to claims that AI chatbots "facilitate an authentic, interactional language learning environment" (Chiu et al., 2023) or that AI use is "essential for promoting cultural sensitivity, intercultural compe-tency, and global awareness" (Anis, 2023). The assumption is that AI output is linguistically and culturally authentic enough that it could substitute in language learning settings for human interlocutors or for intercultural exchange. Such a view ignores how AI systems reproduce language and the limitations of that process for the linguistic features and the cultural content of the resulting output, as well as for AI's understanding of the world behind the generated language. Acknowledging those inadequacies is key to developing critical AI literacy, an essential component of becoming an in-formed AI user. The mathematical model of language in AI (words transformed into vectors) lacks the sociocultural grounding humans have through sensorimotor interactions and from simply living in the real world. Studies of AI's capabilities to engage in pragmatically effective language use have shown significant limitations (Lee and Wang, 2022; Su and Goslar, 2023). While AI systems can gain pragmalinguistic knowledge and learn appropriate formulaic sequences through the verbal exchanges in their training data (politeness conventions, for example), they have proven to be much less effective in sociopragmatic engagement, that is, in generating contextually acceptable speech reflecting an interlocutor's state of mind (Chen et al., 2024). AI systems will inevitably improve through larger datasets and integrated multimedia (Godwin-Jones, 2024). However, those measures will not substitute for the human experience of negotiating common ground linguistically and culturally in social interactions (Barattieri di San Pietro, 2023) and therefore the ability to deal with nuanced pragmatic scenarios. Through published research on the topic and the presenter's study of ChatGPT's performance on discourse completion tasks, this presentation will examine AI's pragmatic abilities.

Presentation Abstracts

Hung-Yun Liu, Russell Hugo & Rachel Quiles
University of Washington

Not presented

Presentation Abstracts

EXPLORING THE USE OF CHATGPT IN ARABIC LANGUAGE EDUCATION: EVALUATING EFFECTIVENESS AND OPTIMIZING INTEGRATION

Kamilia Rahmouni,

Virginia Commonwealth University

Large language models use vast textual data to mimic human-like responses and engage users in natural conversations. ChatGPT, a prominent example, has sparked both excitement and apprehension among educators since its emergence in November 2022. While some see its potential to revolutionize teaching methods by providing valuable support to both educators and learners, others raise concerns about its impact on academic integrity and scholarly publishing (e.g., Kohnke et al., 2023; Teng, 2023). However, despite these ongoing debates, there is a prevailing consensus that Artificial Intelligence (AI) not only endures but also stands poised for further development and evolution (Kostka and Toncelli, 2023; Warner, 2023). Therefore, a commitment to exploring ChatGPT's applications in education is crucial for adequately preparing students for a future increasingly shaped by AI. This study evaluates ChatGPT's effectiveness in Arabic language learning by analyzing data from learning sessions, quizzes, and surveys with Arabic students. It focuses on assessing Arabic grammar and vocabulary acquisition, while also identifying the strengths and limitations of ChatGPT in the context of Arabic language education. Additionally, this study offers practical suggestions for optimizing the use of ChatGPT in Arabic teaching and learning. Through these findings, this research aims to advance language education and contribute to the broader discourse surrounding the integration of AI in educational contexts.

Presentation Abstracts

EXPLORING EFL TEACHERS' TRAINING READINESS TO INTEGRATE AI TOOLS INTO THEIR TEACHING PRACTICE

Maria Perifanou,
Hellenic Open University
& Alla Krasulia,
Sumy State University, Kyiv National Linguistic University, Riga Technical University

In the dynamic landscape of second language acquisition, the integration of generative Artificial Intelligence (AI) tools into English as a Foreign Language (EFL) instruction presents transformative potential. This study aims to explore the preparedness of master's degree students from the Hellenic Open University, Greece; Riga Technical University, Latvia; and Sumy State University, Ukraine, to utilize AI tools in EFL teaching contexts. This research primarily aims to assess their awareness of existing AI resources, gauge their readiness to employ these tools effectively and identify any educational gaps in their master's programs that might impede their ability to use such technologies upon graduation. The methodology involves an initial survey to measure the participants' familiarity with and attitudes toward AI tools designed for language education. Subsequently, participants will be introduced to a range of AI technologies tailored for teaching the four main linguistic skills – reading, listening, writing, and speaking – alongside essential soft skills such as critical thinking, communication, collaboration, and creativity. Following a practical engagement phase where these future educators will implement AI-driven tasks in real classroom settings, a post-intervention survey will evaluate the effectiveness of the AI tools in enhancing both linguistic and soft skills development. The outcomes of this study indicate a low familiarity among EFL teachers with AI tools in their teaching practice. While they perceive the impact of AI tools in EFL teaching as high, they also express a need for continuous training. These findings are anticipated to provide valuable insights into the ongoing discourse surrounding the utilization of AI in language learning. They will support a more informed approach to the development and implementation of AI resources in educational settings.

Presentation Abstracts

AI SUPPORT AND CODING: AUTOMATED CORPUS ANALYSIS FOR LANGUAGE LEARNERS

Robin Couture-Matte & Maura Cruz Enriquez,
TÉLUQ University

In the field of second language acquisition and learning, the analysis of student production provides valuable insights into how learners acquire and apply language. In this regard, automated analysis, driven by advancements in natural language processing, has revolutionized how corpora are analyzed. More specifically, the use of Python (Welcome to Python.Org, 2024) and its libraries offers a free and open-source option that can be customized to the needs of researchers and practitioners (Kane, 2023). However, the fact that programming skills are necessary to use such technologies can be a significant challenge. The assistance of artificial intelligence models can help overcome this challenge. The present exploratory study sought to leverage artificial intelligence to generate codes that facilitate automated corpus analysis. More specifically, two case studies were carried out using two corpora created with the writing assignments of second language learners enrolled in second language courses (A1 to B1 levels) in a Canadian university. The first case study was carried out with a Spanish corpus (n = 97) and analyzed the predictive variables of the grammatical gender. The second case study drew on an English corpus (n=98) and analyzed the use of the gerund. With regard to the first case study, the statistical analysis identified predictive variables affecting gender assignment in Spanish, such as the absence of gender marks and the presence of non-prototypical marks. Concerning the second case study, greater complexity in gerund use was observed as English proficiency levels increased. These significant findings, consistent with previous research, support the affordances of generative artificial intelligence in corpus linguistic analysis, allowing researchers and practitioners with little coding experience to successfully harness the potential of artificial intelligence.

Presentation Abstracts

USING AN AI-CHATBOT FOR TEACHING KOREAN GRAMMAR: A COMPARISON BETWEEN PROFICIENCY LEVELS AND TASK TYPES

Ji-young Shin & Yujeong Choi,
University of Toronto Mississauga

Generative artificial intelligence (AI) chatbots have been increasingly used in second language (L2) teaching but primarily concerned the impact on English (Hwang et al., 2023; Jeon, 2021, 2022). Moderation of language proficiency and task types also require more investigation. The current study examined the impact of AI chatbot-incorporated tasks regarding teaching a less commonly taught language (LCTL), Korean as a foreign language (KFL) grammar, including the interactions with KFL proficiency and task types. Learners' perceptions were also examined. Data were collected from 66 students in a KFL university course in Canada, which covered lexico-grammar performing everyday communicative functions and topics (e.g., asking future plans). Students' proficiency ranged from high-beginner to high-intermediate levels. Half of the grammar points were taught with communicative tasks using an AI chatbot called Iruda, selected for its superior performance in Korean language processing and cultural inclusion. The other half lexico-grammar was taught conventionally. After the course, all grammar points were tested using selected-response and constructed-response tasks. A paired sample t-test compared the scores of grammar items taught using Iruda with conventionally taught items, revealing significantly higher means of the items related to chatbot practice ($t(64) = -4.29$, $p < .001$). Follow-up regression and repeated ANOVAs were conducted to examine the role of pre-existing KFL proficiency and task types. The results indicated greater changes by chatbot practice among less proficient students and significant moderation of task types with higher selected-response item means. Students' perceptions were also generally positive, attributed to enhanced accessibility, increased practice opportunities, and exposure to contemporary Korean culture. The study provides a practical example of an AI chatbot-incorporated curriculum development in an LCTL context with evidence of its usefulness, mitigating limited interaction opportunities. Implications are also drawn for the role of L2 proficiency and tasks with different difficulty levels regarding technology-enhanced L2 education.

Presentation Abstracts

EMPOWERING YOUTH THROUGH LANGUAGE, AI, AND SUSTAINABLE ADVOCACY: A TELECOLLABORATIVE APPROACH

Maria Mont Algamasilla, *Universitat Autònoma de Barcelona*
Onyemowo Ejeme, *Oakbridge Foundation (Abuja, Nigeria)*
Ese Emmanuel Uwosomah, *Universitat Autònoma de Barcelona*
Melinda Dooly, *Universitat Autònoma de Barcelona*

This paper presents the Young Activists' Adventure, an innovative educational initiative that intertwines language learning, AI utilization, and sustainable advocacy among children from diverse backgrounds. Based on data derived from a telecollaborative project, students from Spain (ages 9-10) and Nigeria (10-14) engaged in a collaborative exploration aimed at championing the UN's Sustainable Development Goals (SDGs), particularly raising awareness about the global water crisis (SDG6). The pedagogical approach to the geographically-distanced collaboration focused on enhancing English language skills, breakdown stereotypes, while instilling empathy, global citizenship, and environmental stewardship. Prompt engineering and AI integration were used to facilitate learning and content creation so that learners were guided through discussions and other language output (e.g. campaign posters, discussion with policy advocates) on current World issues while fostering cultural exchange and environmental education. Qualitative data from in-class video recordings, screen captures, student output, teacher and researchers' diary and students' interviews were analyzed to assess the effectiveness of our approach. Preliminary findings suggest that integrating language learning, aided by AI tools, with real-world issues enhances student engagement and understanding, though it does present teachers with interesting challenges such as classroom management. In our talk we will briefly outline the project design before presenting the main results of the initial analysis of the data. We argue that by empowering children with language skills, AI literacy, and a sense of global responsibility trying to solve real World problems, we can open a path towards a more vibrant and sustainable future for all. The comparative approach of this study highlights the potential of technology to serve as a bridge in educational practices between the Global North and Global South, contributing to more equitable learning opportunities and enhanced global awareness. Through collaborative efforts and boundless enthusiasm, even the smallest hands can make an impact on our planet.

Presentation Abstracts

BEYOND BORDERS: INTEGRATING LOCAL AND GLOBAL PERSPECTIVES IN PROJECT–BASED LANGUAGE LEARNING

Elif Kemaloglu-Er,

Adana Alparslan Turkes Science and Technology University

In today's globalized world, it is crucial to implement English language learning not only in the routine of in–class practices but also by directly experiencing real life skills. This study aims to develop and implement a real–life–focused, project–based language learning model in a tertiary–level context connecting the powers of local and global resources via AI–enhanced learning and investigates the effectiveness of the model through student views. In this context, the students have completed the steps of i) research, ii) translation, iii) creativity, and iv) interaction within the scope of their projects and produced meaningful and purposeful outcomes of their own on a wide variety of glocal themes mainly related to art, history, sociology, psychology, architecture, technology, entertainment, and business life. The real–life based skills the projects aimed to improve included research and critical thinking skills, personal expression and creativity skills, team work and problem solving skills, as well as time management, autonomous decision–making and digital skills. The data were collected through a questionnaire, semi–structured interviews, and learner diaries, and analyzed via statistical and thematic analyses. According to the findings, the model was perceived to positively contribute to the development of several linguistic and non–linguistic skills and help learners become competent citizens of the globalized world. Pedagogical implications will also be presented along with suggestions for further research.

Presentation Abstracts

LEVERAGING CHATGPT FOR LANGUAGE LEARNING: A STUDY ON FOREIGN TEACHERS' SWEDISH ACQUISITION

Sofie Johansson, *Gothenburg University*

This research explores the utilization of ChatGPT as an innovative tool in teaching Swedish to foreign educators, presenting a significant scholarly contribution to the fields of language learning and educational technology. By integrating theoretical frameworks such as Interaction Theory (Mackey and Gass, 2014), the Technology Acceptance Model (TAM) (Liu et al., 2023), and Cognitive Load Theory (CLT) (Sweller et al., 2011), the study provides a multidimensional analysis of AI's role in enhancing linguistic proficiency. A mixed-methods approach evaluates ChatGPT's impact on learners, focusing on cognitive load and language acquisition. Significantly, the findings highlight a potential variation in success linked to the learners' metalinguistic awareness, suggesting that individuals with a higher understanding of language structure and function may derive greater benefit from AI-assisted learning tools like ChatGPT. This insight contributes original knowledge to the field, indicating that metalinguistic awareness can influence the effectiveness of technological interventions in language learning. The research underscores the need for personalized learning approaches that consider individual differences in metalinguistic competence, pointing towards a more nuanced understanding of AI's potential in educational settings. By evidencing the correlation between metalinguistic awareness and learning outcomes with AI tools, this study enriches the academic discourse on technology-enhanced language education, offering valuable implications for the development of pedagogical strategies and AI applications tailored to diverse learner profiles Gass, S. M., and Mackey, A. (2014). Input, interaction, and output in second language acquisition. In Theories in second language acquisition (pp. 194-220). Routledge. Liu, G., and Ma, C. (2023). Measuring EFL learners' use of ChatGPT in informal digital learning of English based on the technology acceptance model.

Presentation Abstracts

THE SECRET LIFE OF INTERNATIONAL GRADUATE STUDENTS' USE OF CHATGPT: IDENTITY AND (DIS)EMPOWERMENT IMPACTS

Seyyedeh Mobina Hosseini,
State University of New York at Buffalo

Technologies like AI introduce new possible constellations of identity, agency, and social practices. Williams (2023) stresses the need for AI literacy at all educational levels so that students become "technosocial change agents" or people who use powerful technologies for personal and societal empowerment and addressing complex problems. Although everyone can use ChatGPT, international students have specific experience about using ChatGPT. While anyone who uses this innovative technology may rely frequently on it in their educational life, this dependence is qualitatively different for nonnative immigrant students in higher education who use it to overcome language barriers while also supporting content knowledge development and sharpening of critical thinking ability. However, heavy use of ChatGPT may have adverse effects on students' identities and may even lead them to question themselves and their abilities. Preliminary inquiries indicate that nonnative immigrant students, perhaps like other learners, often keep their use of ChatGPT a secret, fearing judgement and being stigmatized by their peers and educators. Pilot data also suggests that reliance on ChatGPT for every assignment and class, using it in secrecy, and subsequent feelings of disempowerment may influence international graduate–level students' academic and emerging professional identities, attitudes, and confidence. To that end, this exploratory study is guided by the following research questions: • How do international graduate students in U.S. institutions report using ChatGPT to navigate their academic endeavors? • What attitudes do graduate–level students express when using ChatGPT and about using ChatGPT? To answer the RQs, a qualitative, interview–based study has been designed; findings are likely to contribute to deeper understanding of graduate–level, international students experiences with ChatGPT and to inform rapidly evolving discussion of the role of ChatGPT in higher education among international students.

Presentation Abstracts

DEVELOPING ARGCOACH: DISCOURSE ANALYSIS TO SUPPORT SPECIALIZED NATURAL LANGUAGE UNDERSTANDING OF AI

Droste Hennings, MacKenzie Novotny, Samantha Semelroth, Alina Reznitskaya, Evgeny Chukharev,
Iowa State University

In this presentation, I shall report on an exploratory study that investigated six high school EFL teachers' AI assessment literacy as manifested through their use of ERNIE Bot, a Chinese AI-powered chatbot similar to ChatGPT, as they conducted assessments on and gave feedback to the students' written works. Data were collected via individual interviews and digital protocols, which showed how the teachers used the ERNIE Bot. As teachers' assessment knowledge underpins their assessment literacy, a framework for discerning teachers' knowledge that supported their assessment and feedback practices was employed to enable the categorization of the manifestations of teachers' AI assessment literacy into a three-tier knowledge construct, i.e., technicality, practicality, and conceptuality. Findings show that the layers of practicality and conceptuality, which encompass teacher's practical and conceptual knowledge of assessment and feedback, do not seem to be added with new components due to exposure to AI technologies. The layer of technicality, nonetheless, seems to have developed as a body of knowledge that could operate independently, i.e., it seems to have become less attached to practical and conceptual knowledge. Besides, technicality might not necessarily promote the development of practical and conceptual knowledge, although the lack of technical knowledge obviously posed a hindrance.

This re-conceptualization of teacher's AI assessment literacy may provide an alternative to the more popular view that AI assessment literacy is the enactment of assessment literacy in AI-powered digital environments. This re-conceptualization thus calls for more attention to the indispensability of technical knowledge beyond being epiphenomenal, which co-exists with AI-powered digitality as a permeating culture that characterizes teachers' assessment and feedback practices.

Presentation Abstracts

EMI UNIVERSITY STUDENTS' EXPERIENCES WITH USING CHATGPT FOR LANGUAGE AND LITERACY PRACTICES: A QUALITATIVE STUDY

Beyza Aypay & Michelle Bedeker

Nazarbayev University (Graduate School of Education), M.A. in Multilingual Education

Generative artificial intelligence technologies have gathered attention from various domains of higher education due to their benefits and concerns and their potentials to revolutionize the language learning field (Karataş et al., 2024). This qualitative phenomenology research used four different data collection instruments to obtain rich data from nine students at an EMI university in Kazakhstan. The participants were selected from an EMI university's foundation-year (high school-to-university transition) program to explore how students use ChatGPT as a language learning tool and support for literacy practices. Qualitative questionnaires and focus group interviews were employed to investigate different ways and purposes students use ChatGPT. Two art-based data collection instruments were used, image-based reflections, for student perspectives on the benefits and drawbacks of using ChatGPT for an academic context and a significant circle to capture insights on how students integrated using ChatGPT into their university literacy practices. The results showed that participants used ChatGPT in various ways in the EFL context; drawing from Oxford's (1990) language learning strategies classification, ChatGPT served as different types of language learning strategies. Moreover, drawing from Streets' (1984) literacy as a social practice theory, findings revealed the influence of EMI university literacy practices on students' utilization of ChatGPT. The study provides insights and implications for university instructors and policymakers.

Presentation Abstracts

TRACKING THE USE OF GENERATIVE AI IN SECOND LANGUAGE RESEARCH

Ling Ding & Cecilia Guanfang Zhao,
University of Macau

The increasingly widespread use of generative AI tools in scholarly communication and academic publishing, following the release of ChatGPT and similar Generative AI technologies, has sparked intense interest and debate among researchers (Gray, 2024). Assessing the use and traces of such AI generated materials and contents in academic literature, for instance, becomes particularly interesting and important, as the use of Large Language Models (LLMs) as prose editors has raised concerns about the credibility of academic discourse in general. While academics whose first language is not English are particularly encouraged by the availability of such tools, which they believe could finally help diminish the "linguistic injustice" in academic publishing (Hyland, 2016), and increase the chance of getting their work published in mainstream academic journals, limited research exists to have closely examined the use of AI tools and the potential influence of such uses on academic writing and publishing practices and outcome. To address this gap, our study conducted a systematic analysis of papers published between 2016 and 2023 in the field of Second Language Studies, and developed a means to measure the prevalence of LLM-modified content over time. To detect LLM-modified materials, we employed the distributional LLM quantification framework by Liang et al. (2024) to estimate the extent to which sentences in such academic texts have been significantly altered by generative AI. This large-scale analysis revealed subtle shifts in academic language use and means of knowledge construction, which may otherwise go unnoticed in analysis of individual research articles. Implications for academic publishing and second language writing will be discussed.

Presentation Abstracts

RELIABILITY OF LARGE LANGUAGE MODELS AT IDENTIFYING AND CLASSIFYING INFORMATION IN RESEARCH ARTICLES

Kristin Terrill & Elena Cotos,
Iowa State University

GenAI has demonstrated functionality that seems, uncannily, to parallel reading and writing by identifying/reformulating information from source texts and generating novel content. These skills are essential yet very challenging for students tasked with producing literature reviews. This study investigates the feasibility of a human-in-the-loop (Tang, 2020) GenAI-facilitated literature review writing process. The human-in-the-loop concept posits that complex processes can be deconstructed and compartmentalized, and that component functions needed for these processes can be delegated to machines while humans contribute to, and control, the overall process. This study explores the hypothesis that certain functions of the literature review might be reliably and ethically delegated to GenAI. For that, it will test and evaluate prompts designed to elicit GenAI output that simulates human-like performance indicative of reading comprehension and written composition. Prompts testing is grounded in Kim's (2020) Interactive and Dynamic Literacy Model which deconstructs and compartmentalizes reading and writing skills and theorizes their interaction. The results will inform the development of a practical framework that will highlight the capabilities and limitations of GenAI in the context of teaching novices to leverage technology-supported processes while composing literature reviews. Important implications are foreseen for both writing theory and pedagogy as well as for policy guidance related to ethical uses of GenAI for scholarly communication. This abstract was submitted to the Journal of Academic Writing as a proposal for an empirical study. The reviewers recommended an invitation to submit a paper to a special issue on GenAI, which will be published in November 2024.

Presentation Abstracts

PRACTICING AND ASSESSING INTERACTIONAL COMPETENCE USING AI-POWERED AGENTS

Veronika Timpe-Laughlin, *Educational Testing Service*
Tetyana Sydorenko, *Portland State University*
Larry Davis, *Educational Testing Service*
Judit Dombi, *University of Pécs, Hungary*
Rahul Divekar, *Bentley University*
Saerhim Oh, *Educational Testing Service*
Jung In Lee, *Portland State University*
Reza Neiriz *MetaMetrics***,**
& Evgeny Chukharev, *Iowa State University*

Colloquium Overview

Developing the ability in a second language to co-construct meaning in diverse contexts requires opportunities for communicative interaction; however, such opportunities are often scarce (Timpe-Laughlin et al., 2024). While spoken dialogue systems (SDS) can simulate transactional interactions, recent developments in generative AI raise the promise of automated interlocutors that can engage in collaborative meaning-making. The series of talks in this colloquium will present and discuss the findings and implications from three studies examining the various affordances of interactions with generative AI, SDS, and humans for the practice and assessment of interactional competence (IC).

Talk 1: Introduction: Interactional competence in an age of AI
Larry Davis (Educational Testing Service)

Talk 2: Language learners' perceptions of written interactions with ChatGPT for practicing English
Tetyana Sydorenko (Portland State University), Judit Dombi (University of Pécs, Hungary) & Veronika Timpe-Laughlin (Educational Testing Service)

Talk 3: Spoken dialogue technology versus ChatGPT: Benefits and challenges for practicing and assessing oral interaction
Veronika Timpe-Laughlin (Educational Testing Service), Judit Dombi (University of Pécs, Hungary), Rahul Divekar (Bentley University), Saerhim Oh (Educational Testing Service), Tetyana Sydorenko (Portland State University), & Jung In Lee (Portland State University)

Talk 4: Elicitation of IC in tests of oral communication: Humans and SDSs
Reza Neiriz (MetaMetrics)

Talk 5: Commentary on the studies and implications for future research
Evgeny Chukharev (Iowa State University)