BEA 2025

The 20th Workshop on Innovative Use of NLP for Building Educational Applications

Proceedings of the Workshop

July 31 - August 1, 2025

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

Gold Level

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

ISBN 979-8-89176-270-1

Introduction

This year marks the 20th edition of the *Workshop on Innovative Use of NLP for Building Educational Applications*. As in previous years, we are happy to welcome a plethora of work on various aspects and types of educational applications – from traditionally popular tasks around language learning to novel applications related to teaching math and programming languages. This year, we have also extended BEA to a 2-day event, which allowed us to accept more valuable work from our authors: in total, we received a record number of 169 submissions, and from these, we have accepted 12 papers as talks and 63 as poster and demo presentations, for an overall acceptance rate of 44 percent. As in previous years, we have put the main emphasis on the high quality of research when selecting the papers to be accepted, but we also hope that we have managed to bring together a diverse program. One aspect in which BEA continues to excel is the range of languages that are covered by the work submitted and presented at our workshop: this year, accepted papers feature work on educational applications developed for Arabic, English, Estonian, Finnish, Germanic languages, Indian languages, Italian, Romanian, Russian, and Spanish.

In addition to the diverse oral, poster and demo presentations, this year, Kostiantyn Omelianchuk from Grammarly will give a keynote on *How LLMs Are Reshaping GEC: Training, Evaluation, and Task Framing*. BEA 2025 will also incude, for the first time, a half-day tutorial on *LLMs for Education: Understanding the Needs of Stakeholders, Current Capabilities and the Path Forward*. Finally, BEA 2025 has hosted a shared task on *Pedagogical Ability Assessment of AI-powered Tutors*, which attracted a large number of participants, and the program includes an oral presentation on the shared task from the organizers as well as extended poster sessions for shared tasks participants presenting their systems.

Last but not least, we would like to thank everyone who has been involved in organizing the BEA workshop this year. We are particularly grateful to our sponsors who keep providing their support to BEA: this year, our sponsors include Cambridge University Press & Assessment, Duolingo English Test, Grammarly, National Board of Medical Examiners, SigIQ.ai, and Squirrel Ai Learning.

BEA 2025 Organizing Committee

Organizing Committee

General Chair

Ekaterina Kochmar, MBZUAI

Program Chairs

Andrea Horbach, Hildesheim University
Ronja Laarmann-Quante, Ruhr University Bochum
Marie Bexte, FernUniversität in Hagen

Publication Chair

Anaïs Tack, KU Leuven, imec

Shared Tasks Chairs

Victoria Yaneva, National Board of Medical Examiners
Bashar Alhafni, MBZUAI

Sponsorship Chair

Zheng Yuan, University of Sheffield
Jill Burstein, Duolingo

Program Committee

Chairs

Bashar Alhafni, MBZUAI
Marie Bexte, FernUniversität in Hagen
Jill Burstein, Duolingo
Andrea Horbach, CAU Kiel
Ekaterina Kochmar, MBZUAI
Ronja Laarmann-Quante, Ruhr University Bochum
Anaïs Tack, KU Leuven; imec; UCLouvain
Victoria Yaneva, National Board of Medical Examiners
Zheng Yuan, University of Sheffield

Program Committee

Rania Abdelghani, Hector Institute of Educational Sciences and Psychology, University of Tübingen
Tazin Afrin, NBME
Syeda Sabrina Akter, George Mason University
Ali Al-Laith, University of Copenhagen
Giora Alexandron, Weizmann Institute of Science
David Alfter, Gothenburg University
Jatin Ambasana, Indian Institute of Technology Bombay
Jiyuan An, Beijing Language and Culture University
Antonios Anastasopoulos, George Mason University
Nico Andersen, DIPF | Leibniz Institute for Research and Information in Education
Aitor Arronte Alvarez, University of Hawaii at Manoa
Yuya Asano, University of Pittsburgh
Nischal Ashok Kumar, University of Massachusetts Amherst
Berk Atil, Pennsylvania State University
Shiva Baghel, Extramarks
Xiaoyu Bai, University of Potsdam
Jinhyun Bang, Samsung Research
Stefano Banno, University of Cambridge
Mohmaed Basem, MSA University
Michael Gringo Angelo Bayona, Trinity College Dublin
Lee Becker, Pearson
Beata Beigman Klebanov, Educational Testing Service
Milena Belosevic, Bielefeld University
Enrico Benedetti, Utrecht University
Luca Benedetto, University of Cambridge
Maryam Berijanian, Michigan State University
Kay Berkling, Cooperative State University, Karlsruhe
Ummugul Bezirhan, Boston College, TIMSS and PIRLS International Study Center
Krishnakant Bhatt, IIT Bombay
Souvik Bhattacharyya, Lowe's
Abhidip Bhattacharyya, University of Massachusetts, Amherst
Serge Bibauw, Universite catholique de Louvain
Louise Bloch, University of Applied Sciences and Arts Dortmund
Allison Bradford, University of California, Berkeley

Ted Briscoe, MBZUAI
Dominique Brunato, Institute of Computational Linguistics A. Zampolli"(ILC-CNR), Pisa
Ana-Maria Bucur, Interdisciplinary School of Doctoral Studies
Luciano Cabral, IFPE
Andrew Caines, University of Cambridge
Chris Callison-Burch, University of Pennsylvania
Jie Cao, University of Oklahoma
Dan Carpenter, North Carolina State University
Dumitru-Clementin Cercel, University Politehnica of Bucharest
Sophia Chan, Educational Testing Service Canada
Ignatios Charalampidis, University of Tuebingen
Andreas Chari, University of Glasgow
Danqing Chen, Technical University of Munich
Lei Chen, Jinan University
Mei-Hua Chen, Department of Foreign Languages and Literature, Tunghai University
Longfeng Chen, South China University of Technology
Artem Chernodub, ZenDesk
Mihail Chifligarov, Ruhr University Bochum
Luis Chiruzzo, Universidad de la Republica
Hyundong Cho, USC, Information Sciences Institute
Jinho D. Choi, Emory University
Evgeny Chukharev, Iowa State University
Yan Cong, Purdue University
Mark Core, University of Southern California
Sofía Correa Busquets, Pontificia Universidad Católica de Chile, National Center for Artificial Intelligence Chile, Foundational Research on Data Millenium Institute
Steven Coyne, Tohoku University / RIKEN
Scott Crossley, Georgia State University
Syaamantak Das, Indian Institute of Technology Bombay
Mihai Dascalu, University Politehnica of Bucharest
Tirthankar Dasgupta, Tata Consultancy Services Ltd.
Orphee De Clercq, LT3, Ghent University
Kordula De Kuthy, Universität Tübingen
Michiel De Vrindt, KU Leuven
Jasper Degraeuwe, Ghent University
FATIMA DEKMAK, American University of Beirut
Carrie Demmans Epp, University of Alberta
Dorottya Demszky, Stanford University
Aniket Deroy, IIT Kharagpur
Chris Develder, Ghent University
Srijita Dhar, Chittagong University of Engineering & Technology
Yuning Ding, FernUniversität in Hagen
Rahul Divekar, Educational Testing Service
George Duenas, Universidad Pedagogica Nacional
Marius Dumitran, University of Bucharest
Yo Ehara, Tokyo Gakugei University
Walid El Hefny, Leibniz-Institut für Wissensmedien (IWM)
Mohamed Elaraby, University of Pittsburgh
Ron Eliav, Bar-Ilan University
Jordan Esiason, North Carolina State University
Yao-Chung Fan, National Chung Hsing University

Effat Farhana, Auburn University
Mariano Felice, British Council
Nigel Fernandez, University of Massachusetts Amherst
Michael Flor, Educational Testing Service
Jennifer-Carmen Frey, EURAC Research
Benjamin Gagl, University of Cologne
Thomas Gaillat, Rennes 2 university
Martina Galletti, Sony Computer Science Laboratories - Paris | Sapienza University of Rome
Diana Galvan-Sosa, University of Cambridge
Ashwinkumar Ganesan, Amazon Alexa AI
Rujun Gao, Texas A&M University
Lingyu Gao, Toyota Technological Institute at Chicago
Ritik Garg, Extramarks Education Pvt. Ltd.
Voula Giouli, Aristotle University of Thessaloniki / ILSP, ATHENA RC
Sebastian Gombert, DIPF | Leibniz Institute for Research and Information in Education
Kiel Gonzales, University of the Philippines Diliman
Mark Edward Gonzales, De La Salle University
Cyril Goutte, National Research Council Canada
Pranav Gupta, Lowe's
Abigail Gurin Schleifer, Weizmann Institute of Science
Eleonora Guzzi, Universidade da Coruña
Ching Nam Hang, Assistant Professor, Yam Pak Charitable Foundation School of Computing and Information Sciences, Saint Francis University, Hong Kong
Ikhlasul Hanif, Universitas Indonesia
Jiangang Hao, Educational Testing Service
Ahatsham Hayat, University of Nebraska-Lincoln
Ping He, Northeastern University
Nicolas Hernandez, Nantes University
Michael Holcomb, University of Texas Southwestern Medical Center
Matias Hoyl, Stanford University
Chieh-Yang Huang, MetaMetrics Inc
Aiden Huang, Acton-Boxborough Regional High School
Chung-Chi Huang, Frostburg State University
Anna Huelsing, CAU
Leo Huovinen, Metropolia University of Applied Sciences
Catherine Ikae, Applied Machine Intelligence, Bern University of Applied Sciences, Switzerland
Fareya Ikram, University of Massachusetts Amherst
Joseph Marvin Imperial, University of Bath
Radu Tudor Ionescu, University of Bucharest
Raunak Jain, Intuit
Suriya Prakash Jambunathan, New York University
Qinjin Jia, North Carolina State University
Helen Jin, University of Pennsylvania
Abel John, Stanford University
Douglas Jones, MIT Lincoln Laboratory
Edmund Jones, Cambridge University Press & Assessment
Léane Jourdan, Nantes University
Samarth Kadaba, Stanford University
Indika Kahanda, University of North Florida
Tomoyuki Kajiwara, Ehime University
Honeiah Karimi, Cambium Assessment

Anisia Katinskaia, University of Helsinki
Fatemeh Kazemi Vanhari, McMaster University
Elma Kerz, Exaia Technologies
Fazel Keshtkar, St. John's University
Samin Khan, Stanford University
Darya Kharlamova, National Research University Higher School of Economics
Harksoo Kim, Konkuk University
Han Kyul Kim, University of Southern California
Levi King, Indiana University
Kasper Knudsen, ITU
David Kogan, Google
Mamoru Komachi, Hitotsubashi University
Charles Koutcheme, Aalto University
Joni Kruijsbergen, LT3, Ghent University
Andrei Kucharavy, HES-SO Valais-Wallis
Aayush Kucheria, Aalto University
Roland Kuhn, National Research Council of Canada
Gaurav Kumar, University of California San Diego
Murathan Kurfali, RISE Research Institutes of Sweden
Alexander Kwako, University of California, Los Angeles
Kristopher Kyle, University of Oregon
Yunshi Lan, East China Normal University
Antonio Laverghetta Jr., Pennsylvania State University
Jaewook Lee, UMass Amherst
Celine Lee, Cornell University
Seolhwa Lee, Technical University of Darmstadt
Travis Lee, Tennessee Tech University
Bernardo Leite, Faculty of Engineering - University of Porto
Arun Balajiee Lekshmi Narayanan, University of Pittsburgh
Xu Li, Zhejiang University
Hariz Liew, Singapore University of Social Sciences
Chuan-Jie Lin, National Taiwan Ocean University
Yudong Liu, Western Washington University
Naiming Liu, Rice University
Zhexiong Liu, University of Pittsburgh
Julian Lohmann, Christian Albrechts UniversitÃ¤t Kiel
Benny Longwill, Educational Testing Service
Anastassia Loukina, Grammarly Inc
Crisron Rudolf Lucas, UniversityCollegeDublin
Zhihao Lyu, CU Boulder
Sarah Löber, University of Tübingen
Denise Löfflad, Leibniz-Institut für Wissensmedien Tübingen
Wanjing (Anya) Ma, Stanford University
Jakub Macina, ETH Zurich
Lieve Macken, Ghent University
Nitin Madnani, Duolingo
Hang Man, The University of Hong Kong
Zhenjiang Mao, University of Florida
Jacek Marciniak, Adam Mickiewicz University
Arianna Masciolini, University of Gothenburg
Sandeep Mathias, Presidency University

Kaushal Maurya, MBZUAI
Hunter McNichols, University of Massachusetts Amherst
Detmar Meurers, Leibniz-Institut für Wissensmedien (IWM)
Noah-Manuel Michael, Kiel University
Amit Mishra, AmityUniversityMadhyaPradesh
Daniel Mora Melanchthon, Leibniz Institute for Science and Mathematics Education
Sai Sathvik Motamarri, PES University
Phoebe Mulcaire, Duolingo
Laura Musto, Universidad de la Republica
Karthika N J, Indian Institute of Technology Bombay
Farah Nadeem, LUMS
Numaan Naeem, MBZUAI
Ryo Nagata, Konan University
Sungjin Nam, ACT, Inc
Diane Napolitano, The Washington Post
Aneet Narendranath, Michigan Technological University
Léo Nebel, LIP6 - Sorbonne Université
Kamel Nebhi, Education First
Seyed Parsa Neshaei, EPFL
Huy Nguyen, Amazon
Gebregziabihier Nigusie, Mizan-Tepi University
S Jaya Nirmala, National Institute of Technology Tiruchirappalli
Sergiu Nisioi, Human Language Technologies Research Center, University of Bucharest
Adam Nohejl, Nara Institute of Science and Technology
Eda Okur, Intel Labs
Kostiantyn Omelianchuk, Grammarly
Amin Omidvar, PhD student at the Department of Electrical Engineering and Computer Science, York University
Joshua Otten, GeorgeMasonUniversity
Daniel Oyeniran, University of Alabama
Ulrike Pado, HFT Stuttgart
Sankalan Pal Chowdhury, ETH Zurich
Nisarg Parikh, University of Massachussetts, Amherst
Jeiyoon Park, SOOP
Manooshree Patel, University of California, Berkeley
Kaushal Patil, University of Southern California
Kseniia Petukhova, MBZUAI
Henry Pit, University of Melbourne
Long Qin, Alibaba
Mengyang Qiu, Trent University
Marti Quixal, University of Tuebingen
Chatrine Qwaider, MBZUAI
Md. Abdur Rahman, Southeast University
Vatsal Raina, University of Cambridge
Sparsh Rastogi, Thapar Institute of Engineering and Technology
pranshu rastogi, Independent Researcher
Manav Rathod, University of California, Berkeley
Hanumant Redkar, Goa University, Goa
Robert Reynolds, Brigham Young University
Saed Rezayi, National Board of Medical Examiners
Luisa Ribeiro-Flucht, University of Tuebingen

Frankie Robertson, University of Jyväskylä
Shadman Rohan, Center for Computational & Data Sciences, IUB
Donya Rooein, Bocconi University
Aiala Rosá, Instituto de Computación, Facultad de Ingeniería, Universidad de la RepÃºblica
Allen Roush, University of Oregon
Alla Rozovskaya, Queens College, City University of New York
Josef Ruppenhofer, Fernuniviersität in Hagen
Stefan Ruseti, University Politehnica of Bucharest
Johannes Rückert, University of Applied Sciences and Arts Dortmund
Mariam Saeed, Applied Innovation Center
Trishita Saha, IIT Hyderabad
Jonathan Sakunkoo, Stanford University
Annabella Sakunkoo, Stanford University OHS
Omer Salem, Cairo University
Nicy Scaria, Indian Institute of Science
Nils-Jonathan Schaller, Leibniz Institute for Science and Mathematics Education
Veronica Schmalz, KULeuven
Stephanie Schoch, University of Virginia
Matthew Shardlow, Manchester Metropolitan University
Mayank Sharma, Graduate Student, Stanford University
Kevin Shi, University of California, Berkeley
Mariana Shimabukuro, Ontario Tech University
Hyo Jeong Shin, Sogang University
Gyu-Ho Shin, University of Illinois Chicago
Astha Singh, Iowa State University
Li Siyan, Columbia University
Lucy Skidmore, British Council
Anastasia Smirnova, San Francisco State University
Mariia Soliar, Leibniz-Institut für Wissensmedien (IWM)
Mayank Soni, ADAPT Centre, Trinity College Dublin
Alexey Sorokin, Moscow State University
Anna Sotnikova, EPFL
KV Aditya Srivatsa, MBZUAI
Maja Stahl, Leibniz University Hannover
Felix Stahlberg, Google Research
Katherine Stasaski, Salesforce Research
Helmer Strik, Centre for Language and Speech Technology (CLST), Centre for Language Studies (CLS), Radboud University Nijmegen
David Strohmaier, University of Cambridge
Hakyung Sung, University of Oregon
Abhijit Suresh, Graduate Student
Andreas Säuberli, LMU Munich
Chuangchuang Tan, Beijing Jiaotong University
CheeWei Tan, NanyangTechnologicalUniversity
Wenjia Tan, University of Macau
Nhat Tran, University of Pittsburgh
Felipe Urrutia, Center for Advanced Research in Education
Masaki Uto, The University of Electro-Communications
Takehito Utsuro, University of Tsukuba
Martin Vainikko, University of Tartu
Sowmya Vajjala, National Research Council

Piper Vasicek, Brigham Young University
Justin Vasselli, Nara Institute of Science and Technology
Giulia Venturi, Institute of Computational Linguistics Antonio Zampolli"(ILC-CNR)
Anthony Verardi, Duolingo
Amit Verma, Guvi Geek Network
Elena Volodina, University of Gothenburg
Anh-Duc Vu, University of Helsinki
Deliang Wang, The University of Hong Kong
Nikhil Wani, University of Southern California
Taro Watanabe, Nara Institute of Science and Technology
Yuchen Wei, Pennsylvania State University
Alistair Willis, The Open University
Steven Wilson, University of Michigan-Flint
Anna Winklerova, Faculty of Informatics Masaryk University
Hanna Woloszyn, University of Cologne
Simon Woodhead, Eedi
Anna Wroblewska, Faculty of Mathematics and Information Science, Warsaw University of Technology
Changrong Xiao, Tsinghua University
Hiroaki Yamada, Institute of Science Tokyo
Haiyin Yang, University of Florida
Roman Yangarber, University of Helsinki
Sahar Yarmohammadtoosky, NBME
Hanling Yi, Intellifusion, Inc.
Su-Youn Yoon, EduLab
Marcos Zampieri, George Mason University
Alessandra Zarcone, Technische Hochschule Augsburg
Fabian Zehner, DIPF | Leibniz Institute for Research and Information in Education
Kamyar Zeinalipour, University of Siena
Torsten Zesch, Computational Linguistics, FernUniversität in Hagen
Franklin Zhang, Bellevue College
Mike Zhang, Aalborg University
Jing Zhang, Emory University
Yiling Zhao, Stanford University
Yang Zhong, University of Pittsburgh
Yiyun Zhou, NBME
Ej Zhou, University of Cambridge
Jessica Zipf, University of Konstanz
Michael Zock, CNRS-LIS
Leonidas Zotos, University of Groningen
Bowei Zou, Institute for Infocomm Research
Robert Östling, Department of Linguistics, Stockholm University

Keynote Talk

How LLMs Are Reshaping GEC: Training, Evaluation, and Task Framing

Kostiantyn Omelianchuk
Grammarly

Abstract: This keynote will explore the evolving role of Large Language Models (LLMs) in training and evaluating Grammatical Error Correction (GEC) systems, using Grammarly as a case study. It will cover the shift from primarily using human-annotated corpora to semi-synthetic data generation approaches, examining its impact on model training, evaluation practices, and overall task definition. Key topics include task definition challenges, trade-offs between data types, observed biases in models, and recent advances in LLM-based evaluation techniques. The talk will also explore scalable approaches for multilingual GEC and outline implications for future research.

Bio: Kostiantyn Omelianchuk is an Applied Research Scientist and Area Tech Lead at Grammarly, where he works on practical applications of NLP, with a primary interest in Grammatical Error Correction (GEC). He has over nine years of experience in the field and has co-authored several papers, including GECToR: Grammatical Error Correction – Tag, Not Rewrite, a widely used approach in the GEC community. His research explores edit-based modeling, the use of large language models for text correction and simplification, and the transition from human-annotated to synthetic data for training and evaluation. His recent work focuses on multilingual GEC, LLM-based evaluation methods, and synthetic data generation.

Table of Contents

Program

Thursday, July 31, 2025

09:00 - 10:30 *Tutorial Session A*

10:30 - 11:00 *Coffee Break*

11:00 - 12:30 *Tutorial Session B*

12:30 - 14:00 *Lunch Break / Birds of a Feather*

14:00 - 15:30 *Oral Session A*

A Bayesian Approach to Inferring Prerequisite Structures and Topic Difficulty in Language Learning
Anh-Duc Vu, Jue Hou, Anisia Katinskaia, Ching-Fan Sheu and Roman Yangarber

Enhancing Arabic Automated Essay Scoring with Synthetic Data and Error Injection
Chatrine Qwaider, Bashar Alhafni, Kirill Chirkunov, Nizar Habash and Ted Briscoe

Alignment Drift in CEFR-prompted LLMs for Interactive Spanish Tutoring
Mina Almasi and Ross Kristensen-McLachlan

You Shall Know a Word's Difficulty by the Family It Keeps: Word Family Features in Personalised Word Difficulty Classifiers for L2 Spanish
Jasper Degraeuwe

Assessing Critical Thinking Components in Romanian Secondary School Textbooks: A Data Mining Approach to the ROTEX Corpus
Madalina Chitez, Liviu Dinu, Marius Micluta-Campeanu, Ana-Maria Bucur and Roxana Rogobete

Unsupervised Automatic Short Answer Grading and Essay Scoring: A Weakly Supervised Explainable Approach
Felipe Urrutia, Cristian Buc, Roberto Araya and Valentin Barriere

15:30 - 16:00 *Coffee Break*

16:00 - 17:30 *Poster Session A*

Automatic Generation of Inference Making Questions for Reading Comprehension Assessments
Wanjing (Anya) Ma, Michael Flor and Zuowei Wang

Lessons Learned in Assessing Student Reflections with LLMs
Mohamed Elaraby and Diane Litman

Automated L2 Proficiency Scoring: Weak Supervision, Large Language Models, and Statistical Guarantees
Aitor Arronte Alvarez and Naiyi Xie Fincham

Advances in Auto-Grading with Large Language Models: A Cross-Disciplinary Survey
Tania Amanda Nkoyo Frederick Eneye, Chukwuebuka Fortunate Ijezue, Ahmad Imam Amjad, Maaz Amjad, Sabur Butt and Gerardo Castañeda-Garza

Exploring LLMs for Predicting Tutor Strategy and Student Outcomes in Dialogues
Fareya Ikram, Alexander Scarlatos and Andrew Lan

Temporalizing Confidence: Evaluation of Chain-of-Thought Reasoning with Signal Temporal Logic
Zhenjiang Mao, Artem Bisliouk, Rohith Nama and Ivan Ruchkin

18:00 - 21:00 *Workshop Dinner*

Friday, August 1, 2025

09:00 - 09:45 *Keynote Talk by Kostia Omelianchuk*

09:45 - 10:30 *Oral Session B*

LLMs in alliance with Edit-based models: advancing In-Context Learning for Grammatical Error Correction by Specific Example Selection
Alexey Sorokin and Regina Nasyrova

Findings of the BEA 2025 Shared Task on Pedagogical Ability Assessment of AI-powered Tutors
Ekaterina Kochmar, Kaushal Maurya, Kseniia Petukhova, KV Aditya Srivatsa, Anaïs Tack and Justin Vasselli

MSA at BEA 2025 Shared Task: Disagreement-Aware Instruction Tuning for Multi-Dimensional Evaluation of LLMs as Math Tutors
Baraa Hikal, Mohmaed Basem, Islam Oshallah and Ali Hamdi

10:30 - 11:00 *Coffee Break*

11:00 - 12:30 *Poster Session B*

Leveraging Generative AI for Enhancing Automated Assessment in Programming Education Contests
Stefan Dascalescu, Marius Dumitran and Mihai Alexandru Vasiluta

Is Lunch Free Yet? Overcoming the Cold-Start Problem in Supervised Content Scoring using Zero-Shot LLM-Generated Training Data
Marie Bexte and Torsten Zesch

Towards a Real-time Swedish Speech Analyzer for Language Learning Games: A Hybrid AI Approach to Language Assessment
Tianyi Geng and David Alfter

LEVOS: Leveraging Vocabulary Overlap with Sanskrit to Generate Technical Lexicons in Indian Languages
Karthika N J, Krishnakant Bhatt, Ganesh Ramakrishnan and Preethi Jyothi

The Need for Truly Graded Lexical Complexity Prediction
David Alfter

Educators' Perceptions of Large Language Models as Tutors: Comparing Human and AI Tutors in a Blind Text-only Setting
Sankalan Pal Chowdhury, Terry Jingchen Zhang, Donya Rooein, Dirk Hovy, Tanja Käser and Mrinmaya Sachan

Costs and Benefits of AI-Enabled Topic Modeling in P-20 Research: The Case of School Improvement Plans
Syeda Sabrina Akter, Seth Hunter, David Woo and Antonios Anastasopoulos

Are Large Language Models for Education Reliable Across Languages?
Vansh Gupta, Sankalan Pal Chowdhury, Vilém Zouhar, Donya Rooein and Mrinmaya Sachan

Span Labeling with Large Language Models: Shell vs. Meat
Phoebe Mulcaire and Nitin Madnani

STAIR-AIG: Optimizing the Automated Item Generation Process through Human-AI Collaboration for Critical Thinking Assessment
Euigyum Kim, Seewoo Li, Salah Khalil and Hyo Jeong Shin

End-to-End Automated Item Generation and Scoring for Adaptive English Writing Assessment with Large Language Models
Kamel Nebhi, Amrita Panesar and Hans Bantilan

bea-jh at BEA 2025 Shared Task: Evaluating AI-powered Tutors through Pedagogically-Informed Reasoning
Jihyeon Roh and Jinhyun Bang

K-NLPers at BEA 2025 Shared Task: Evaluating the Quality of AI Tutor Responses with GPT-4.1
Geon Park, Jiwoo Song, Gihyeon Choi, Juoh Sun and Harksoo Kim

IALab UC at BEA 2025 Shared Task: LLM-Powered Expert Pedagogical Feature Extraction
Sofía Correa Busquets, Valentina Córdova Véliz and Jorge Baier

TBA at BEA 2025 Shared Task: Transfer-Learning from DARE-TIES Merged Models for the Pedagogical Ability Assessment of LLM-Powered Math Tutors
Sebastian Gombert, Fabian Zehner and Hendrik Drachsler

COGENT: A Curriculum-oriented Framework for Generating Grade-appropriate Educational Content
Zhengyuan Liu, Stella Xin Yin, Dion Hoe-Lian Goh and Nancy Chen

Friday, August 1, 2025 (continued)

LLM-Assisted, Iterative Curriculum Writing: A Human-Centered AI Approach in Finnish Higher Education
Leo Huovinen and Mika Hämäläinen

12:30 - 14:00 *Lunch Break / Birds of a Feather*

14:00 - 15:30 *Poster Session C*

Can LLMs Effectively Simulate Human Learners? Teachers' Insights from Tutoring LLM Students
Daria Martynova, Jakub Macina, Nico Daheim, Nilay Yalcin, Xiaoyu Zhang and Mrinmaya Sachan

Adapting LLMs for Minimal-edit Grammatical Error Correction
Ryszard Staruch, Filip Gralinski and Daniel Dzienisiewicz

Do LLMs Give Psychometrically Plausible Responses in Educational Assessments?
Andreas Säuberli, Diego Frassinelli and Barbara Plank

Towards Automatic Formal Feedback on Scientific Documents
Louise Bloch, Johannes Rückert and Christoph Friedrich

Transformer-Based Real-Word Spelling Error Feedback with Configurable Confusion Sets
Torsten Zesch, Dominic Gardner and Marie Bexte

Unsupervised Sentence Readability Estimation Based on Parallel Corpora for Text Simplification
Rina Miyata, Toru Urakawa, Hideaki Tamori and Tomoyuki Kajiwara

Exploiting the English Vocabulary Profile for L2 word-level vocabulary assessment with LLMs
Stefano Banno, Kate Knill and Mark Gales

Improving AI assistants embedded in short e-learning courses with limited textual content
Jacek Marciniak, Marek Kubis, Michał Gulczyński, Adam Szpilkowski, Adam Wieczarek and Marcin Szczepański

GermDetect: Verb Placement Error Detection Datasets for Learners of Germanic Languages
Noah-Manuel Michael and Andrea Horbach

Automated Scoring of Communication Skills in Physician-Patient Interaction: Balancing Performance and Scalability
Saed Rezayi, Le An Ha, Yiyun Zhou, Andrew Houriet, Angelo D'Addario, Peter Baldwin, Polina Harik, Ann King and Victoria Yaneva

Can GPTZero's AI Vocabulary Distinguish Between LLM-Generated and Student-Written Essays?
Veronica Schmalz and Anaïs Tack

A Framework for Proficiency-Aligned Grammar Practice in LLM-Based Dialogue Systems
Luisa Ribeiro-Flucht, Xiaobin Chen and Detmar Meurers

RETUYT-INCO at BEA 2025 Shared Task: How Far Can Lightweight Models Go in AI-powered Tutor Evaluation?
Santiago Góngora, Ignacio Sastre, Santiago Robaina, Ignacio Remersaro, Luis Chiruzzo and Aiala Rosá

Archaeology at BEA 2025 Shared Task: Are Simple Baselines Good Enough?
Ana Roșu, Jany-Gabriel Ispas and Sergiu Nisioi

NLIP at BEA 2025 Shared Task: Evaluation of Pedagogical Ability of AI Tutors
Trishita Saha, Shrenik Ganguli and Maunendra Sankar Desarkar

LLM-based post-editing as reference-free GEC evaluation
Robert Östling, Murathan Kurfali and Andrew Caines

Estimation of Text Difficulty in the Context of Language Learning
Anisia Katinskaia, Anh-Duc Vu, Jue Hou, Ulla Vanhatalo, Yiheng Wu and Roman Yangarber

Exploring task formulation strategies to evaluate the coherence of classroom discussions with GPT-4o
Yuya Asano, Beata Beigman Klebanov and Jamie Mikeska

EyeLLM: Using Lookback Fixations to Enhance Human-LLM Alignment for Text Completion
Astha Singh, Mark Torrance and Evgeny Chukharev

Decoding Actionability: A Computational Analysis of Teacher Observation Feedback
Mayank Sharma and Jason Zhang

Friday, August 1, 2025 (continued)

Thapar Titan/s : Fine-Tuning Pretrained Language Models with Contextual Augmentation for Mistake Identification in Tutor–Student Dialogues
Harsh Dadwal, Sparsh Rastogi and Jatin Bedi

Wonderland_EDU@HKU at BEA 2025 Shared Task: Fine-tuning Large Language Models to Evaluate the Pedagogical Ability of AI-powered Tutors
Deliang Wang, Chao Yang and Gaowei Chen

BJTU at BEA 2025 Shared Task: Task-Aware Prompt Tuning and Data Augmentation for Evaluating AI Math Tutors
Yuming Fan, Chuangchuang Tan and Wenyu Song

SmolLab_SEU at BEA 2025 Shared Task: A Transformer-Based Framework for Multi-Track Pedagogical Evaluation of AI-Powered Tutors
Md. Abdur Rahman, MD AL AMIN, Sabik Aftahee, Muhammad Junayed and Md Ashiqur Rahman

LexiLogic at BEA 2025 Shared Task: Fine-tuning Transformer Language Models for the Pedagogical Skill Evaluation of LLM-based tutors
Souvik Bhattacharyya, Billodal Roy, Niranjan M and Pranav Gupta

DLSU at BEA 2025 Shared Task: Towards Establishing Baseline Models for Pedagogical Response Evaluation Tasks
Maria Monica Manlises, Mark Edward Gonzales and Lanz Lim

LookAlike: Consistent Distractor Generation in Math MCQs
Nisarg Parikh, Alexander Scarlatos, Nigel Fernandez, Simon Woodhead and Andrew Lan

From End-Users to Co-Designers: Lessons from Teachers
Martina Galletti and Valeria Cesaroni

15:30 - 16:00 *Coffee Break*

16:00 - 17:15 *Oral Session C*

Down the Cascades of Omethi: Hierarchical Automatic Scoring in Large-Scale Assessments
Fabian Zehner, Hyo Jeong Shin, Emily Kerzabi, Andrea Horbach, Sebastian Gombert, Frank Goldhammer, Torsten Zesch and Nico Andersen

Direct Repair Optimization: Training Small Language Models For Educational Program Repair Improves Feedback
Charles Koutcheme, Nicola Dainese and Arto Hellas

Advancing Question Generation with Joint Narrative and Difficulty Control
Bernardo Leite and Henrique Lopes Cardoso

Intent Matters: Enhancing AI Tutoring with Fine-Grained Pedagogical Intent Annotation
Kseniia Petukhova and Ekaterina Kochmar

LLMs Protégés: Tutoring LLMs with Knowledge Gaps Improves Student Learning Outcome
Andrei Kucharavy, Cyril Vallez and Dimitri Percia David

17:15 - 17:30 *Closing Remarks*