LHI EXPO

UNIVERSITY OF
CAMBRIDGE

DEPARTMENT OF
COMPUTER SCIENCE
AND TECHNOLOGY

2 JULY 2024

TABLE OF CONTENTS

WELCOME

Welcome to the first Learning & Human Intelligence Poster Exposition, the LHI Expo, at the Department of Computer Science & Technology. We have more than thirty posters and four talks detailing the work of Ada Computer Science, the Institute for Automated Language Teaching & Assessment (ALTA), the Raspberry Pi Computing Education Research Centre, and other independent researchers.

We hope you enjoy the event and welcome feedback at lhi-admins@cst.cam.ac.uk.

The Learning & Human Intelligence Group, Department of Computer Science & Technology, University of Cambridge.

www.cst.cam.ac.uk/research/lhi

PROGRAMME

11:00	**Start**
11:15-11:30	**Talk** by *Dr Sue Sentance* RPCERC - ”An Overview of Current Research in RPCERC“
11:45-12:00	**Talk** by *Jacob Brown & Skye Purchase* Ada Computer Science - "Isaac Physics & Ada Computer Science in 2024"
12:00-13:00	**Lunch** is served
13:15-13:30	**Talk** by *Dr Luca Benedetto* ALTA - "Using GPT to simulate students' response patterns"
13:45-14:00	**Talk** by *Prof Paula Buttery* ALTA - “What is the role of universities in Language Learning Research?”

POSTER SESSION

1. LCT Semantics and multiple-choice questions

Authors: Jane Waite, Eirini Kolaiti, Meurig Thomas and Karl Maton

This small-scale student-focused study explored how the sociological framework of Legitimation Code Theory's (LCT) Semantics dimension was used to improve multiple-choice question feedback to wrong (and right) answers. In the poster, we explain the process of Semantic Profiling. This process can be used to analyse any learning event and we suggest that profiling may be particularly useful to consider when designing applications that include explanations created by Generative AI models.

2. Teaching Computer Science with and through other forms of knowledge

Authors: Paul Curzon, Jane Waite and Karl Maton

This small-scale case study explored how the sociological framework of Legitimation Code Theory's (LCT) Autonomy dimension was used to improve lesson design where more than one subject/ topic is being taught in the same lesson. In the poster, we explain the process of creating a translation device to help review "what is being taught" (the content) and "why" (the purpose), how to use this to analyse a lesson, how to draw an autonomy pathway and how to suggest lesson improvements based on whether a "tour" or "one-way trip" is seen. This process can be used to analyse any learning event. Do you research contexts where more than one discipline or topic is taught at the same time? If so, come and learn about LCT Autonomy as a way to explore why interdisciplinary teaching may sometimes fall short.

3. Feedback Literacy: Holistic analysis of secondary educators' views of LLM explanations of Program Error Messages

Authors: Veronica Cucuiat and Jane Waite

This poster describes an academic study in which we investigated secondary educators' views of LLM explanations of programming error messages for classroom use. The data analysis is contextualised in Feedback Literacy theory, and the poster details how such a theory might contribute to teaching and learning of programming using LLMs.

4. Experience AI: Introducing AI/ML to Grade 6–8 students in the UK

Authors: Jane Waite, Ben Garside, Robert Whyte, Diana Kirby and Sue Sentance

In this poster, we detail the design of Experience AI, a curriculum designed to introduce AI and machine learning concepts to students aged 11 to 14 in the UK. We outline several of the key design principles: avoiding anthropomorphization to prevent misconceptions about AI, promoting awareness of AI-related careers to foster interest and aspiration, and providing extensive support for teachers who may lack prior AI knowledge. The curriculum includes six comprehensive lessons covering foundational AI concepts, ethical considerations, and real-world applications, supported by career examples and expert interviews. An independent evaluation showed positive impacts on both student and teacher awareness of AI careers, increased teacher confidence in teaching AI, and a reduction in anthropomorphic descriptions of AI by students.

5. Introducing AI concepts in Junior High Schools in Ghana

Authors: Salomey Afua Addo and Sue Sentance

The Ministry of Education and the Ghana Education Service (GES) have introduced Artificial Intelligence (AI) into Junior High Schools in Ghana. They accomplished this by introducing an official computing curriculum into the education system. This new curriculum differs from the old one by new computing concepts such as programming, algorithms, robotics and AI. So far, there have not been any assessments to evaluate its effectiveness in the classroom. We use Bloom's taxonomy as a preliminary framework to assess which cognitive expectations are placed on students as they take lessons in AI in schools. Understanding which cognitive levels students will be functioning at would help teachers design lessons to develop students' higher-order thinking skills as they learn concepts about AI.

6. Investigating the use of generative AI in school computing education

Authors: Sue Sentance, Steve Watson, Salomey Afua Addo, Veronica Cucuiat, Shengpeng Shi, Jane Waite and Bo Yu

This poster describes a short project that has been seed-funded by the School of Technology, whereby the Raspberry Pi Computing Education has worked in collaboration with the Faculty of Education. Amongst a lot of top-down advice for schools and teachers on how to use generative AI, particularly LLMs, in the computing classroom, we set up a working group, made up primarily of computing teachers, to collaboratively develop guidance for schools, and to gain a joint understanding of the issues faced by schools around this area of technology. As a result of this project, we will be publishing a short guidance document for computing educators in July.

7. Funds of identity and culturally responsive computing: K–5 teachers' adaptations to computing resources

Authors: Katharine Childs and Jane Waite, Raspberry Pi Computing Education Research Centre

This study aimed to explore how to broaden participation in computing through the lens of culturally responsive computing and sociocultural identity theory. Our results from working with a small group of primary teachers show how they collaboratively adapted computing resources to be culturally relevant and use student identities to foster inclusive and engaging computing lessons.

8. Measuring Teacher Self-Efficacy and Careers Awareness in K-12 AI Education

Authors: Robert Whyte, Diana Kirby and Sue Sentance

In this poster, we detail work that investigates teachers' self-efficacy and awareness of AI careers in K–12 education. Using data from 63 teachers in England, we adopted a validated survey instrument to measure these aspects, finding strong internal reliability and construct validity. This work indicates promising utility for these scales, which will be further tested with a larger sample to observe changes in teachers' self-efficacy and career awareness over time.

9. K-12 Student Early conceptions of AI: Understanding AI Applications, Models, Engines and Implications

Authors: Robert Whyte, Diana Kirby and Sue Sentance

In this poster, we detail work that explores students' conceptions of AI and frame these using an appropriate framework. Analysing responses from 474 students aged 11–18, we identified both accurate conceptions, such as AI's reliance on large datasets, and naïve ones, like attributing human-like qualities to AI. These insights highlight the importance of addressing misconceptions by challenging emergent ideas.

10. EPICS – Exploring Physical Computing in Schools

Authors: Jessie Durk and Sue Sentance

This poster describes a new five-year project we are undertaking in the Raspberry Pi Computing Education Research Centre to investigate how engagement with physical computing can impact primary school children's confidence, creativity, and agency. The main part of the study is qualitative, whereby we will follow a group of children from the ages of 8-9 to when they are 13-14, engaging with them, their parents and their teachers annually. This project is just starting but we have recruited a set of partner schools to work with us over the next five years and are in the middle of a UK tour collecting data. The project has been funded by the Micro:bit Educational Foundation, the BBC, and Nominet, and has been timed to coincide with a roll-out whereby the BBC has donated a class set of BBC micro:bits to all primary schools who requested them.

11. Using GPT to simulate students' response patterns

Authors: Luca Benedetto, Giovanni Aradelli, Antonia Donvito, Alberto Lucchetti, Andrea Cappelli and Paula Buttery

We study whether GPT can be used for simulating the response patterns of students to exam questions, and find that while this is indeed possible, it must be done with great caution as it is not precisely controllable.

12. Measuring learner perception of difficulty in AI-generated text

Authors: Gladys Tyen, Andrew Caines and Paula Buttery

How can we measure the linguistic difficulty of AI-generated text from the perspective of a language learner? We invited 160 language learners to interact with our chatbot, and used 4 different metrics to measure the difficulty of the text. In this poster, we show results comparing the linguistic difficulty of an off-the-shelf AI model, versus one which uses our adaptive decoding method. Come try out our chatbot for yourself at our demo today!

13. Semantic Error Prediction: Estimating Lexical Complexity in Production

Author: David Strohmaier

Estimating word complexity is essential for many computer-assisted language learning technologies. We introduce semantic error prediction (SEP) as a novel task that assesses the production complexity of words. In SEP, a system has to predict which word tokens are replacements, that is corrections, of tokens from the original text. We use LLMs for this task and establish their practical relevance for scoring the vocabulary usage of learner essays, thereby providing a finer-grained assessment of learner skills.

14. Distractor Generation Using Generative and Discriminative Capabilities of Transformers

Authors: Shiva Taslimipoor, Luca Benedetto, Mariano Felice and Paula Buttery

We propose and evaluate a distractor generation approach based on transformer models that perform particularly well in generating i) a diverse set of distractors, ii) distractors that are not too similar to the correct answer option, and iii) that match human-generated distractors better than previously proposed approaches.

15. Grammatical Error Correction for Code-Switched Sentences by Learners of English

Authors: Kelvin Wey Han Chan, Christopher Bryant, Li Nguyen, Andrew Caines and Zheng Yuan

This work presents the first-ever investigation into Grammatical Error Correction (GEC) for code-switching (CSW) text. Specifically, we explore different methods of generating synthetic CSW data, and find that: (i) data augmentation is most effective in the context of multilingual (rather than monolingual) pre-trained models (e.g. XLM-RoBERTa), and (ii) replacing a random noun token in a sentence yielded the best improvement, which is consistent with linguistic insights on code-switching. Our best model shows a 1.57 F0.5 average improvement across three CSW test sets (English-Chinese, English-Korean, and English-Japanese) and generalises well to typologically similar CSW languages without affecting the model's performance on a monolingual dataset.

16. Decoupling language learner text and a personalized learner model

Author: Yuan Gao

Our project models second language learners' text as a blend of their native and target grammar, allowing for individualized learning and precise progress tracking. Our model offers personalized modeling and accurate progress tracking at low computational and resource costs.

17. Etymology Dataset and Its Applications for Vocabulary Learning and Mastery

Author: Yuan Gao

Introducing our latest dataset that builds a structured etymological network of English and other Indo-European words. We believe this resource offers exciting potential for enhancing vocabulary learning and adaptive language education tools. We would love to receive feedback and comments on potential future applications to CUP&A products.

18. Evaluating LLMs' knowledge of the CEFR

Authors: Luca Benedetto, Gabrielle Gaudeau, Andrew Caines and Paula Buttery

We study the actionable knowledge that several LLMs have of the CEFR, by prompting them to perform three CEFR-related tasks (automated scoring, CEFR level classification of reading passages, and simulation of students of different levels), and find that their performance is often not satisfactory.

19. KUPA-KEYS: Logging Keystrokes in Writing by English Learners

Authors: Georgios Velentzas, Andrew Caines, Rita Borgo, Erin Pacquetet, Clive Hamilton, Taylor Arnold, Diane Nicholls, Paula Buttery, Thomas Gaillat, Helen Yannakoudakis and Nicolas Ballier

In this poster, we present an exciting new dataset which includes not only essays written by learners and native speakers of English, but also expert grades and keystroke data. This data enables analysis of the writing process, identification of words and phrases which learners struggle with, and detection of plagiarism from online sources.

20. LLMs "off-the-shelf" or Pretrain-from-Scratch? Recalibrating Biases and Improving Transparency in language teaching and assessment technology using Small-Scale Language Models

Authors: Suchir Salhan, Richard Diehl-Martinez, Zébulon Goriely, Andrew Caines and Paula Buttery

Is it possible to significantly reduce the financial and environmental footprint of deploying NLP systems for specialist domains, while preserving performance? Our research finds that lightweight models, built from scratch and trained on smaller datasets that are carefully selected and ordered, can achieve comparable performance to much larger Language Models, despite being trained on x100-100 less data.

21. Let's Agree to Disagree: Beyond the Golden Standard in Analytic Automated Essay Scoring

Authors: Gabrielle Gaudeau, Paula Buttery, Andrew Caines, Øistein Andersen and Zheng Yuan

The shift from holistic to analytic scoring in essay assessment has the potential of reducing the apparent arbitrariness of marking while also opening new avenues for student feedback. Unfortunately, this practice tends to be less reliable and prone to rater biases. This project explores an innovative approach of leveraging rater disagreements in analytic scoring data for more reliable and nuanced Automated Essay Scoring (AES) of essay traits.

22. How good are LLM explanations? An Education-inspired Rubric

Authors: Diana Galvan-Sosa, Pride Kavumba, Yunmeng Li, Gabrielle Gaudeau, Zheng Yuan, Camélia Guerraoui, Keisuke Sakaguchi, Andrew Caines and Paula Buttery

This paper presents a novel methodology for evaluating large language models (LLMs), focusing specifically on assessing the quality of their generated explanations. Our approach offers a systematic and comprehensive evaluation framework, addressing current limitations in LLM assessment techniques."

23. Sentence level construction: distinguishing good from bad sentences

Author: Diana Galvan Sosa

This project aims to address the issue of run-on sentences in student writing, focusing on identifying overly long or unclear sentences. The goal is to provide constructive feedback to students, helping them recognize and transform run-on sentences into clear, concise, and effective "good" sentences.

24. Applications of Code-Switching to Language Learning: Code-Switch Tutor

Authors: Igor Sterner and Simone Teufel

Bilinguals have strong intuitions as to what code-switching sounds natural, and what does not. We develop computational models of this acceptability and apply them to language learning applications!

25. Dynamic Example Selection for Grammatical Error Correction with Large Language Models

Author: Saksham Shah

Investigating the use of in-context learning and few-shot prompting for grammatical error correction, as well as exploring the possibilities for performance improvements using dynamic example selection strategies.

26. Prompting large language models to correct learner text

Authors: Christopher Davis, Andrew Caines, Øistein Andersen, Shiva Taslimipoor, Helen Yannakoudakis, Zheng Yuan, Christopher Bryant, Marek Rei, Paula Buttery

Large language models (LLMs) can perform many NLP tasks through carefully crafted natural language prompts. We examine a wide selection of commercial and open-source LLMs for automatic grammatical error correction of English second language learner text, on four established benchmarks. We find open-source models obtain comparable performance to commercial ones, and observe a bias towards fluency corrections, supporting recent literature.

27. Tending Towards Stability: Convergence Challenges in Small Language Models

Authors: Richard Diehl Martinez, Paula Buttery

We study the learning dynamics of small and large language models and find that layers of larger models converge faster and more smoothly to their final representations.

28. English Grammar Profile tagger

Authors: Øistein Andersen, Andrew Caines, Paula Buttery, Anne O'Keeffe, Geraldine Mark, Diane Nicholls

Automated identification of English Grammar Profile (EGP) constructions in learner text has a number of applications including explainable automated assessment and feedback to learners and teachers on strong and weak areas. This work involves refining the original EGP descriptions so they can be encoded in terms of lexico-semantic patterns, which can then be implemented and tested on a set of example sentences and a larger corpus of learner writing to ensure high precision and recall.

29. Recurrent Neural Collaborative Filtering for Knowledge Tracing

Authors: Russell Moore, Andrew Caines, Paula Buttery

Knowledge Tracing (KT) is the task of inferring a student's knowledge or skill from observing their work. Here we present Recurrent Neural Collaborative Filtering (RNCF) a new KT method that across several data sets achieves more accurate predictions of a student's ability than the related Deep Knowledge Tracing. The internal representations in RNCF also display promising clustering characteristics, potentially useful in grouping students for personalised learning.

30. Extended rater representations: automarker training

Authors: Mark Elliott and Paula Buttery

This work applies many-facet Rasch models to automarker training data; mitigating for smaller, imbalanced data sets, and exploiting customised loss functions.

31. Mitigating Frequency Bias and Anisotropy in Language Model Pre-Training with Syntactic Smoothing

Authors: Richard Diehl Martinez, Zébulon Goriely, Andrew Caines, Paula Buttery, Lisa Beinborn

“Frequency bias” is the tendency of large language models to prefer frequent tokens when making grammatical judgements. We introduce Syntactic Smoothing as a method for improving the representation of rare tokens.

PRESENTERS

ADA COMPUTER SCIENCE

Jacob Brown
Skye Purchase
James Sharkey
Meurig Thomas

ALTA INSTITUTE

Øistein Andersen
Luca Benedetto
Paula Buttery
Andrew Caines
Christopher Davis
Diana Galvan-Sosa
Yuan Gao
Gabrielle Gaudeau
Russell Moore
Li Nguyen
David Strohmaier
Shiva Taslimipoor
Gladys Tyen

COMPUTER SCIENCE & TECHNOLOGY

Richard Diehl Martinez
Mark Elliott
Zeb Goriely
Suchir Salhan
Saksham Shah
Igor Sterner

RPCERC

Salomey Afua Addo
Katharine Childs
Veronica Cucuiat
Jessie Durk
Claire Johnson
Diana Kirby
Carrie Anne Philbin
Sue Sentance
Jane Waite
Robert Whyte

Cambridge ALTA

Institute for Automated Language Teaching and Assessment

RASPBERRY PI
COMPUTING EDUCATION RESEARCH CENTRE

UNIVERSITY OF CAMBRIDGE
DEPARTMENT OF
COMPUTER SCIENCE AND TECHNOLOGY

Edited by Kata Szabo