Empowering Education with LLMs - the Next-Gen Interface and Content Generation

Steven Moore[1], Richard Tong[2], Anjali Singh[3], Zitao Liu[4], Xiangen Hu[5], Yu Lu[6], Joleen Liang[7], Chen Cao[8], Hassan Khosravi[9], Paul Denny[10], Chris Brooks[3], John Stamper[1]

[1] Carnegie Mellon University, Pittsburgh PA, USA
[2] Carnegie Learning, Pittsburgh PA, USA
[3] University of Michigan, Ann Arbor MI, USA
[4] Jinan University, Guangdong Province 510632, China
[5] The University of Memphis, Memphis TN, USA
[6] Beijing Normal University, Beijing 100875, China
[7] Squirrel AI Learning, Shanghai 200030, China
[8] University of Sheffield, Sheffield, United Kingdom
[9] University of Queensland, St Lucia QLD 4072, Australia
[10] University of Auckland, Auckland 1010, New Zealand
StevenJamesMoore@gmail.com

Abstract. We propose the first annual workshop on Empowering Education with LLMs - the Next-Gen Interface and Content Generation. This full-day workshop explores ample opportunities in leveraging humans, AI, and learning analytics to generate content, particularly appealing to instructors, researchers, learning engineers, and many other roles. The process of humans and AI cocreating educational content involves many stakeholders (students, instructors, researchers, instructional designers, etc.), thus multiple viewpoints can help to inform what future generated content might be useful, new and better ways to assess the quality of such content and to spark potential collaborative efforts between attendees. Ultimately, we want to illustrate how everyone can leverage recent advancements in AI, making use of the latest machine learning methods and large language models (LLMs), and engage all participants in shaping the landscape of challenges and opportunities in this space. We wish to attract attendees interested in scaling the generation of instructional and assessment content and those interested in the use of online learning platforms.

Keywords: Large language models, Educational content creation, human-AI partnerships, learnersourcing, robosourcing

1 Motivation and Theme

Language models that leverage Generated Pre-trained Transformers (GPT) have significantly advanced the field of natural language processing (NLP) and led to the development of various language-based applications [7]. One potential application of language models is as communication interfaces in human-in-the-loop systems, where the model serves as a mediator among the teacher, students and the machine

capabilities including its own. This approach has several benefits, including the ability to personalize interactions, allow unprecedented flexibility and adaptivity for human-AI collaboration and improve the user experience. However, several challenges still exist in implementing this approach, including the need for more robust models, designing effective user interfaces, and ensuring ethical considerations are addressed.

Additionally, instructors often find themselves having to generate large banks of resources such as practice and assessment questions to accommodate the growing shift towards online learning formats. The continual creation and improvement of assessment items allows for a greater breadth of topic coverage, helps to identify well-constructed and valid assessments, and as a result, enables improved learning analytics [6]. However, instructors and teaching staff rarely have the time or incentive to develop large numbers of high-quality questions for formative assessments that are often used for personalization and adaptive learning; instead their efforts tend to be focused on creating high-stakes assessments such as quiz or exam questions. To address this challenge, there is great potential in exploring educational content creation via partnerships that involve pairings of instructors, students, and AI.

Partnerships for cocreating educational content often involve four distinct and iterative phases: creation, evaluation, utilization, and instructor/expert oversight. These four phases are also utilized by *learnersourcing*, which involves students generating their own educational resources and content that can be leveraged by future learners [4, 8]. This offers a domain agnostic way to help scale the creation of high-quality assessments, while also helping students learn the course content. Leveraging advances in large language models (LLMs), learnersourcing can be combined with AI to provide students with near instant feedback on the educational content they create, yielding higher quality contributions [2].

Partnerships between student-AI and instructor-AI present ample opportunity for content creation and evaluation [11]. Advances in natural language processing (NLP) and generative models provide space for AI to play a fundamental role in the co-creation of content with humans or to assist with the automated evaluation of its quality. The quality evaluation of this content can be further supported by learning analytics related to how students perform on these human-AI cocreated questions, compared to traditional assessments. Related work has explored the quality and novelty of AI-generated learning resources [10], and leveraged NLP [5], trust-based networks [1], and deep learning methods [9] to assist students in the evaluation of both student- and AI-generated content. While human input remains critical in this creation and evaluation process, more work needs to look at leveraging AI to further support students and instructors as they create educational content.

2   Objectives

This workshop aims to bring together researchers and practitioners from academia and industry to explore the potential of LLMs as the communication and collaboration interfaces in human-in- the-loop systems. The objectives of the workshop are to:

- The application of Large Language Models (LLMs) in educational settings
- Generation and evaluation of educational content with the help of LLMs

- Co-creation of educational partnerships, where the human or AI might benefit the most
- Ethical considerations in the use of LLMs as communication interfaces in educational settings
- Designing effective and standardized user interfaces for LLM-based educational systems
- Crowdsourcing & Learnersourcing in conjunction with LLMs

3   Call for Submission

While no submission is required to participate in the workshop, we encourage 4-6 page submissions of work-in-progress or position papers that are related to partnerships for co-creating educational content. The call for papers will open on March 27, 2023, and the submission deadline will be May 27, 2023. The submitted papers will be reviewed by three members of the program committee and the authors will be notified of their paper acceptance on June 03, 2023.

Papers can target a range of topics related to human-AI partnerships for educational content creation. When it comes to the evaluation process of having students or AI review and revise student-generated content, there is a challenge regarding how we can assist students in optimally acting on the suggestions and revisions. Further work remains to investigate how we might leverage AI to assist students in making consistently high-quality learnersourced contributions [3]. Among these challenges with humans and AI cocreating educational content lie many opportunities to explore ways of making it more accessible and beneficial to student learning. A clear opportunity regarding the creation of student-generated content is the different ways we can encourage students to make high-quality contributions, such as leveraging learnersourcing interventions [12]. There are limitless activity types that can be created and evaluated using a plethora of techniques. For instance, students could work in conjunction with a large language model, like ChatGPT or GPT-3, to develop and refine assessment questions or explanations of learning content [10]. This can help them quickly improve the content they generate, while also engaging them in critical thinking as they review the model's suggestions, such as recommended distractors, and explore the limitations of the models. LLMs are trained on broad data produced by humans, and thus are known to suffer from biases like humans. Using automatically generated content without human oversight for educational content generation runs the risk of perpetuating some of these biases. We see a human-in-the-loop approach, involving both students and instructors, as essential for moderating such biases and for improving and tailoring the performance of the underlying generative models for suitability in educational contexts.

4   Workshop Format and Activities

The workshop will run as an interactive full-day session with mini-presentations and round-table discussions on the theme. The provisional schedule is given below:

- **Introductions**: Introductions of workshop organizers and participants, and a background to the focus of the workshop.
- **Invited Guest Speaker**: A leading researcher not on the organizing committee will give a presentation regarding the use of AI in educational content creation
- **Short Presentations**: Authors of accepted submissions present their work which would be followed by a Q&A session
- **Round-table discussion**: Participants will move around specific topics of interest related to various types of human-AI partnerships for educational content including creating, evaluating, utilizing and overseeing.
- **Mini-Hackathon**: Participants will work alongside the organizers to build out prototypes of educational applications that leverage LLMs.
- **Concluding remarks and community engagement**: Closing remarks on the workshop will be made with future steps and opportunities for collaboration between participants.

The main goal of this workshop is to explore how partnerships between students, instructors, and AI can be leveraged towards educational content. We believe participants from a wide range of backgrounds and prior knowledge on learning sciences, machine learning, natural language processing, and learning analytics can both benefit and contribute to this workshop. As the creation of educational content involves many stakeholders, multiple viewpoints can help to inform what future student and AI-generated educational content might be useful, better ways to assess the quality of the content and spark collaboration efforts between attendees. The accepted submissions will be published as part of a CEUR proceedings. We will also provide participants with a Slack channel and mailing list dedicated to sharing related work, along with updating our workshop website.

5 Organizers

- **Steven Moore** is a PhD student at Carnegie Mellon University and is advised by Dr. John Stamper. His research is focused on leveraging NLP with learnersourcing, finding ways to assess the quality of student-generated content.
- **Richard Tong** is an experienced technologist, executive, entrepreneur and one of the leading evangelists for global standardization efforts for learning technologies and artificial intelligence. He serves as the current chair of IEEE Artificial Intelligence Standards Committee.
- **Anjali Singh** is a PhD candidate at the University of Michigan and is advised by Dr. Christopher Brooks. Her research uses learnersourcing and AI to improve Data Science education, in formal and informal learning settings like MOOCs.
- **Zitao Liu** is the Dean of Guangdong Institute of Smart Education, Jinan University, Guangzhou, China. His research is in the area of machine learning, and includes contributions in the areas of artificial intelligence in education and educational data mining.
- **Xiangen Hu** is a professor in the Department of Psychology, Department of Electrical and Computer Engineering and Computer Science Department at The University of Memphis (UofM) and senior researcher at the Institute for

Intelligent Systems (IIS) at the UofM and is professor and Dean of the School of Psychology at Central China Normal University (CCNU).

- **Yu Lu** is currently an Associate Professor with the School of Educational Technology, Faculty of Education, Beijing Normal University (BNU), where he also serves as the director of the artificial intelligence lab at the advanced innovation center for future education (AICFE) and the director of ICT center at Sino- Finnish Joint Learning Innovation Institute (JoLii).
- **Joleen Liang** is a Ph.D candidate in the Intelligent Science System at Macao University of Science and Technology. She is the executive director of the AI and Robotics Education Committee of the China Education Development Strategy Society, the deputy head of the Technology and Standards Working Group, and the Smart Education Working Committee of the Internet Society of China.
- **Chen Cao** is a PhD candidate at the Information School at the University of Sheffield and is advised by Professor Frank Hopfgartner, Dr Laura Sbaffi and Dr Xin (Skye) Zhao. Her research leverages LLMs to create intelligent tutoring systems with generative content, creative analogies, and adaptive feedback.
- **Hassan Khosravi** is an Associate Professor in the Institute for Teaching and Learning Innovation and an Affiliate Academic in the School of Information Technology and Electrical Engineering at the University of Queensland. He has conducted extensive learnersourcing research and leads the development and dissemination efforts of the RiPPLE system.
- **Christopher Brooks** is an Assistant Professor at the University of Michigan and is an applied Computer Scientist who builds and studies the effects of educational technologies in higher education and informal learning environments. He has led learnersourcing efforts on the Coursera platform, where he investigated student choice in the generation of multiple-choice questions.
- **Paul Denny** is an Associate Professor in Computer Science at the University of Auckland, New Zealand. He leads the PeerWise project, which hosts educational content created by students from 90 countries, and has pioneered work evaluating the implications of LLMs for computing education..
- **John Stamper** is an Associate Professor at the Human-Computer Interaction Institute at Carnegie Mellon University and the Technical Director of the Pittsburgh Science of Learning Center DataShop. His work involves leveraging educational data mining techniques and the creation of data tools.

References

1. Darvishi, A., Khosravi, H., & Sadiq, S. (2021). Employing peer review to evaluate the quality of student generated content at scale: A trust propagation approach. Proceedings of the Eighth ACM Conference on Learning@ Scale, 139–150.
2. Denny, P., Sarsa, S., Hellas, A., & Leinonen, J. (2022). Robosourcing Educational Resources--Leveraging Large Language Models for Learnersourcing. arXiv preprint arXiv:2211.04715.
3. Huang, A., Hancock, D., Clemson, M., Yeo, G., Harney, D., Denny, P., & Denyer, G. (2021). Selecting student-authored questions for summative assessments. Research in Learning Technology, 29.

4. Khosravi, H., Demartini, G., Sadiq, S., & Gasevic, D. (2021). Charting the design and analytics agenda of learnersourcing systems. LAK21: 11th International Learning Analytics and Knowledge Conference, 32–42.

5. Moore, S., Nguyen, H. A., & Stamper, J. (2020). Evaluating Crowdsourcing and Topic Modeling in Generating Knowledge Components from Explanations. International Conference on Artificial Intelligence in Education, 398–410.

6. Moore, S., Nguyen, H. A., & Stamper, J. (2022). Leveraging Students to Generate Skill Tags that Inform Learning Analytics. In Proceedings of the 16th International Conference of the Learning Sciences-ICLS 2022, pp. 791-798.

7. Moore, S., Nguyen, H. A., Bier, N., Domadia, T., & Stamper, J. (2022, September). Assessing the Quality of Student-Generated Short Answer Questions Using GPT-3. In Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption: 17th European Conference on Technology Enhanced Learning, EC-TEL 2022, Toulouse, France, September 12–16, 2022, Proceedings (pp. 243-257). Cham: Springer International Publishing.

8. Moore, S., Stamper, J., Brooks, C., Denny, P., & Khosravi, H. (2022, June). Learnersourcing: Student-generated Content@ Scale. In Proceedings of the Ninth ACM Conference on Learning@ Scale (pp. 259-262).

9. Ni, L., Bao, Q., Li, X., Qi, Q., Denny, P., Warren, J., Witbrock, M., & Liu, J. (2022). Deepqr: Neuralbased quality ratings for learnersourced multiple-choice questions. Proceedings of the AAAI Conference on Artificial Intelligence, 36(11), 12826–12834.

10. Sarsa, S., Denny, P., Hellas, A., & Leinonen, J. (2022). Automatic Generation of Programming Exercises and Code Explanations Using Large Language Models. Proceedings of the 2022 ACM Conference on International Computing Education Research-Volume 1, 27–43.

11. Singh, A., Brooks, C., & Doroudi, S. (2022). Learnersourcing in Theory and Practice: Synthesizing the Literature and Charting the Future. In Proceedings of the Ninth ACM Conference on Learning@ Scale (pp. 234-245).

12. Singh, A., Brooks, C., Lin, Y., & Li, W. (2021, June). What's In It for the Learners? Evidence from a Randomized Field Experiment on Learnersourcing Questions in a MOOC. In Proceedings of the Eighth ACM Conference on Learning@ Scale (pp. 221-233).