An initial exploration of semi-automated tutoring: how AI could be used as support for online human tutors

Malin Jansson, KTH Royal Institute of Technology, Sweden, maljan@kth.se
Kathy Tian, KTH Royal Institute of Technology, Sweden, kathyt@kth.se
Stefan Hrastinski, KTH Royal Institute of Technology, Sweden, stefanhr@kth.se
Olov Engwall, KTH Royal Institute of Technology, Sweden, engwall@kth.se

Abstract

In this paper, we begin our process of incorporating an AI bot in an online chat tutoring setting as a support for the tutor. We explore how an AI bot could give suggestions for tutor messages, although the human tutor will control how to communicate with the student. Tutoring, an important dimension of networked learning, has long been seen as a beneficial approach to students' learning. An AI bot has the potential to aid tutors in the tutoring process and contribute to the scalability. The present pilot study was conducted in the tutoring setting of the Math Coach program. In the program, teacher students aid students from upper primary school to upper secondary school in mathematics through an online text-based chat system. Llama2 was used as a large language model (LLM), fine-tuned for Swedish comprehension utilizing the Math Coach system's chat logs. Four coaches, teacher students at a technical Swedish university and active in the Math Coach program, were invited to interact with the AI bot and participate in a group discussion. The coaches interacted individually with the AI bot while the chat conversation was displayed on a monitor so all participants could discuss the interaction while it took place. A semi-structured interview approach was taken and the participants were also encouraged to 'think aloud' about their experience. In the discussions, the coaches expressed surprise by the AI's social aspect. They perceived the AI bot as friendly with a positive attitude and were especially surprised by its ability to correctly place appropriate emojis. The coaches agreed that the AI was able to ask both appropriate and helpful questions and share some good guidance for how to proceed in the problem-solving process. However, they felt that the AI bot was not able to offer sufficient mathematical guidance, oftentimes the AI bot was confidently wrong. It also wrote too long messages, which humans would typically separate into several chat messages, and did not wait for a response but instead moved too quickly towards the solution. Moving forward we plan to address the effects of improved prompts on the AI bot and continue finetuning the LLM. We will continue to conduct pilot studies and eventually conduct more large-scale empirical studies.

Research context

Tutoring, described as "the means whereby an adult or 'expert' helps somebody who is less adult or less expert" (Wood et al., 1976), has long been seen as a beneficial approach for students' learning in, for example, mathematics (Graesser & Person, 1994; Pellegrini et al., 2021). It is an important dimension of networked learning, which may be defined as involving processes of collaboration and knowledge-creation enabled by technologies (NLEC, 2021), promoting connections; between learners, between learners and tutors, and in learning communities (Goodyear et al., 2004). Through online tutoring, students may receive help with their problems from a tutor, which can be a teacher or a more knowledgeable peer, but it can also be a non-human tutor, such as an Intelligent Tutoring System (ITS). In this paper, we begin to study the inclusion of an ITS in online tutoring sessions as support for the tutor.

Tutoring can trigger students' problem-solving and reasoning, as deeper-level questions may be addressed and misconceptions revealed (Graesser & Person, 1994), as well as promote their ability to understand, analyse, and apply conceptual knowledge (Hanham et al., 2021). Online tutoring has been found to improve academic outcomes (Chappell et al., 2015), but online tutors may face challenges guiding the students and not revealing the solution, finding the right level of tutoring (Hrastinski et al., 2014), keeping the tutoring conversation on track and on the course content (Cifuentes & Lents, 2010), as well as finding the correct level of tutor participation (Mazzolini & Maddison, 2003; 2007). Additionally, tutoring suffers the challenge of the 2 sigma problem, i.e. one-to-one tutoring is effective in reaching a high level of learning but expensive to implement on a large scale (Bloom, 1984). In this aspect, ITS has the potential to aid coaches in the tutoring conversation and contribute to the scalability, as AI-supported tutors may be able to aid more students at the same time. Furthermore, online text-based systems offer a good opportunity for tutoring, as tutors may assist multiple participants at once.

Proceedings of the Fourteenth International Conference on Networked Learning 2024, Edited by: Cutajar, M., Borg, C., De Laat, M., Dohn, N.B., Ryberg, T.

Various studies have explored ITS as support for students, by acting as human tutors, monitoring students' progress and suggesting learning content, as well as the design of ITS (see for example Chen et al. (2022), Kim and Kim (2020), Sharples, (2023), VanLehn (2011), and Yang and Zhang, (2019)). ITS have rarely been applied in courses including problem-solving and mainly focused on students' knowledge and performance (Mousavinasab et al., 2021). These systems have been found to foster and encourage deep learning, improving student's learning experience as well as teacher efficiency and instructional quality (Chen et al., 2020). They have also been suggested to promote equal access to education, though potential harms and bias must be considered (Lin et al., 2023). While AI tutoring has shown promising results, Afzal et al. (2019) argue that technology cannot completely replace human tutors with technology as one-on-one tutoring requires human intuition, expertise, and communication skills.

This paper begins our process of incorporating an ITS, or AI-bot, in an online tutoring setting, The Math Coach program. In the Math Coach program (https://www.mattecoach.se) students from upper primary school to upper secondary school are offered help in mathematics from teacher-students through an online text-based chat system. As the platform was built to support students based in Sweden, the interactions between tutors and students are predominantly in Swedish, although occasional interactions take place in English. The online tutors in Math Coach (from here on referred to as coaches) must aid the students in their mathematic problem-solving by supporting their cognitive processes as well as creating a safe and secure environment for students to communicate meaningfully (Garrison et al., 2000; Garrison, 2009) and be able to express emotions connected to the problem-solving process (De Corte et al., 2011; Op't Eynde et al., 2006; Stenbom et al., 2016). The social aspect is necessary for cognitive development and learning (Bozkurt, 2017; Ernest, 1996), and an AI bot intended to support the tutors must also be able to consider the context.

Using an AI bot as support for tutors in their tutoring of students has, to the best of our knowledge, been overlooked in the research field. We call this approach semi-automated tutoring to describe tutoring where human and automated tutoring are combined. This leads to many important questions, such as what the role is of the human vs. automated tutor, and how they will work together. In this paper, we begin to investigate how online tutoring conversations are affected by the use of an AI bot to support tutors. Through a focus group interview, we aim to receive initial feedback from coaches in the Math Coach program on including an AI bot in the network of tutors and learners.

Research design

Large Language Model (LLM)

Our study uses pre-trained transformers, notably large language models (LLMs), guided by the generative pre-training principles as described by Radford et al. (2018). We explored several open-source LLMs, our primary concerns including (1) safety of use, (2) capacity for Swedish comprehension, (3) completeness and accuracy, (4) model size, and (5) context length. Our choice converged on Meta's 13-billion parameter Llama2, which stood out due to its performance and capacity to manage extended conversations. The base Llama model had been pre-trained on data including Swedish texts and initial interactions with the base model revealing its capacity to understand the Swedish language. However, for the model to be able to respond consistently in Swedish and capable of suggesting posts to our human coaches, fine-tuning was necessary. For this, we utilized the Math Coach system's chat logs and applied the Low-Rank Adaptation technique (LORA, Hu et al., 2021). We deployed the model via the Google Cloud Platform, utilizing two Nvidia L4 GPUs. Finally, an interface was built, allowing tutors to interact with the model through API calls, which was used for the purposes of our qualitative research on the Math Coach coaches' experiences with the LLM.

Focus group interview

Participants for the focus group interview were selected from the Math Coach program, as it is the coaches the AI bot is meant to support. Coaches, who are also teacher-students at a technical university in Sweden, were invited to test an early version of the supportive AI bot and discuss their interactions with the chatbot in real time with the first and second authors. Four coaches (two female, two male; ages 20-30 years), all of whom had been working as coaches for nearly one year as well as had practical experience in classroom teaching, chose to participate. In the session, they engaged with the model, simulating the experience of students using the Math Coach system, and offering insights from a maths education perspective. Specifically, tutors were instructed to engage with the chatbot as though they were students using the Math Coach system. Their experience as coaches made them uniquely positioned to understand the common needs and questions of the platform's users, as well as their own needs for support as coaches. Their insights provided the researchers with a valuable perspective on the potential benefits and areas of improvement for the model before its integration into the Math Coach site.

Each of the four tutors individually interacted with the AI chatbot and the chat conversation was displayed on a large monitor within the interview room such that all participants, including the first and second authors, were able to view the dialogue exchanges. Since this was a focus group interview, every participant and both authors were simultaneously present in the room. The authors acted as facilitators and moderators following a semi-structured interview approach. To foster organic conversation and reflections, we implemented a popular methodology used in the field of user interaction and design research (UI/UX) called the "Think-Aloud Protocol" which is often employed in usability testing (Nielsen, 1993). With the principles of the aforementioned method in mind, all four participants were asked to verbalise their thoughts, feelings, and decisions in real time for every interaction. Participants were also encouraged to share their thoughts, both when actively leading the chat interaction and when observing passively. This approach allowed the researchers to gain insights into the participants' cognitive processes, understand how they interact with the LLM system and pinpoint moments that elicited positive or negative feedback. The focus group interview, with the informed consent of all participants, was audio recorded so that the interview process could be revisited.
After all students had completed their sessions, they came together to discuss their experiences collectively. However, following a semi-structured approach, a few guiding questions were prepared to ensure that the conversations stayed on the topic (e.g., "What did you find (un)helpful about this response?", "What are your thoughts on this response?", etc.). The first author served as the primary moderator for the discussion.

Preliminary findings and discussion

In this paper, we aimed to receive initial feedback from online coaches on the benefits and challenges of an AI bot as a support during tutoring sessions, by letting the coaches interact with the AI and then discuss. In the focus group interview, the coaches focused preliminary on three aspects of the AI bot: tutoring, mathematical and social aspects.

The tutoring aspect

The coaches thought the AI bot provided some well-formulated tutoring expressions and showed the potential to be able to support the problem-solving process. They found the AI sharing guidance for how to proceed in problem-solving, and having an encouraging and supportive tone: *"It's like really insistent that it's here to help. Yeah. ´I want to help you with that.´ It's very eager."*. The coaches appreciated that the AI would try to explain things with different words when told they did not understand: *"Because he* [participant referring to the AI bot as "he"] *tried to explain the problem with different words. That's pretty good. Yeah. At least the last question is very interesting with emojis that people feel like it's appropriate."*. The coaches also agreed that the AI was able to ask questions that were both appropriate and helpful. The questions were supportive and guiding while not revealing the answer. This showed the possibility for AI to be a part of networked learning through tutoring conversations, helping coaches to give guidance and ask meaningful questions, thereby supporting the students' problem-solving and understanding (Chen et al., 2020; Graesser & Person, 1994; Hanham et al., 2021). However, they also agreed that the AI came across as too much, and included too much information in each message: *"It's good. It has really good questions and good like what it's saying is just saying a lot and the same thing. But it's like good phrases and everything."*. This leads to useful questions or guidance drowning in meaningless information. The coaches reflected that the AI bot did not separate its response in several chat messages meaning that there was a monologue rather than a dialogue. The AI bot suggested information, questions, steps towards solutions and answers all at once. The AI bot must be able to give short concrete suggestions for messages and wait for the student to respond, before giving new suggestions, in order to engage in a dialogue. The responses from the AI bot were connected to the subject and problem-solving, indicating that the AI was able to keep the conversation on track (Cifuentes & Lents, 2010). However, we need to make sure that not too much information is sent out at once, and that the AI bot listens to student responses before deciding what to communicate.

The social aspect

The coaches expressed surprise about how friendly and supportive the AI bot was. Encouraging results as tone of conversation have been found to be an important aspect for patients with the system and any limitations (Afzal et al., 2019) and the necessity of the social aspect for cognitive development (Bozkurt, 2017; Ernest, 1996). The coaches felt that it kept a good tone in its messages and had a positive attitude. "*Friendly, social, positive attitude, good phrases and reasoning."* The guidance was perceived to be in an encouraging and kind tone, and they appreciated how the AI bot mirrored their tone in the messages by connecting to them and repeating what they said. The coaches especially focused on how the AI bot was able to use emojis appropriately, adding, what they deemed, a correct emoji at the correct place. *"It's kind of cool that it uses the appropriate emojis actually. I feel*

like that's what it got spot on". This result is promising regarding the AI bot being able to consider the context and supporting the coach in maintaining social presence, important for learning and cognitive processes (Bozkurt, 2017; Ernest, 1996; Garrison et al., 2000; Garrison, 2009).

The mathematical aspect

The coaches were all in agreement that the AI was not able to offer sufficient mathematical guidance. The AI tried giving solutions or guidance towards solutions that the coaches deemed not mathematically correct. The coaches explained to the AI that it was incorrect, but the AI was not able to give correct answers even with further tries. The coaches reflected that the AI as it was, would not be able to aid and support the coaches with the process of solving mathematical questions and analysing the answers, since the mathematical knowledge is too limited. They were also surprised that the AI's social aspect was better than the mathematical one. *"Yeah, I'm surprised that the AI social competence is better than its mathematical competence. But once again, it's very fast to give the answer"*. Furthermore, the coaches reflected on what they as coaches wanted support with during one-to-one tutoring. They considered the AI bot's limited mathematical knowledge and argued it unnecessary for the AI to be a subject expert, as the coaches themselves are teacher-students and have sufficient knowledge on the subject. They did, however, discuss that it would be useful to receive support on important steps in problem-solving, as well as the solution, so they know where the problem-solving is supposed to end up. Not having to calculate the problem themselves may free up time and make them as coaches more efficient, i.e., be able to help more students. However, the coaches also recognized the value of calculating the problem themselves in order to fully understand where the student might get stuck, have issues, or miscalculate.

Conclusion and the next steps

The coaches found the initial guiding responses and questions that the AI bot was able to present to be a good start. They thought many of the AI bot's expressions could be helpful to them in real conversations, to aid in problem-solving and make them able to help more students at the same time. However, the coaches did not find the AI bot suitable to hold conversations by itself, considering the lack of mathematical accuracy, how it gathered too much information in each message and was too quick to give steps toward a solution, or even a solution itself. They argued that a human coach was needed to make sure the conversation focused on aiding the students in their cognitive process in an appropriate way, in line with Afzal et al. (2019) that human tutors can not completely be replaced. Having human tutors between the AI bot and the students may also help minimize potential biases from the system (Lin et al., 2023). Finally, they did recognize that this was an early version of the AI bot and that more work is needed for the bot to be useful in live tutoring sessions. Nevertheless, these initial results are promising as the use of AI could contribute to making tutoring more scalable, which is a key issue identified in previous research (Bloom, 1984).

For the next step, further development of the LLM and prompting is necessary to incorporate a useful supportive AI bot in Math Coach. We will explore the impact of using a larger parameter model, such as Llama2-70B, for the AI bot's responses. Furthermore, we will continue to refine the parameters of the LLM to address issues such as verbosity, accuracy, and repetition. By adjusting variables such as the temperature, which dictates the model's "creativity," we can test and evaluate the performance of the generated text and iteratively improve the language model. The model will receive more finetuning regarding the Swedish language, as it was not always correct with expressions and grammar and sometimes slipped into English during the conversations. We will also work with testing different prompts to see how that will affect the AI suggestions and personality. One thing the coaches emphasised was the supportive, friendly and encouraging tone of the AI, and how well it used emojis. This aspect we want to nourish while still managing how the AI expresses itself, concerning the length of the AI messages and what is included in each message.

As the aim is to have semi-automated tutoring where human and automated tutoring are combined, we will continue to invite coaches from the Math Coach program to test the AI bot as a support for them. From there we can further address the aspect of what the coaches need support on as well as want support on, and if the AI bot's mathematical skills are important or not to be able to give good support.

References

Afzal, S., Dempsey, B., D'Helon, C., Mukhi, N., Pribic, M., Sickler, A., Strong, P., Vanchiswar, M., & Wilde, L. (2019). The Personality of AI Systems in Education: Experiences with the Watson Tutor, a one-on-one virtual tutoring system. *Childhood Education.* 95(1), 44-52. https://doi.org/10.1080/00094056.2019.1565809

Bloom, B. S. (1984). The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educational Researcher*, *13* (6), 4–16.

Bozkurt, G. (2017). Social Constructivism: Does it Succeed in Reconciling Individual Cognition with Social Teaching and Learning Practices in Mathematics? Journal of Education and Practice, 8 (3), 210-218.

Chappell, S., Arnold, P., Nunnery, J., & Grant, M. (2015). An examination of an online tutoring program's impact on low-achieving middle school Students' mathematics achievement. Online Learning, 19(5), 37–53. https://doi.org/10.24059/olj.v19i5.694.

Chen, L., Chen, P., & Lin, Z. (2020). "Artificial Intelligence in Education: A Review," in *IEEE Access*, vol. 8, pp. 75264-75278, https://doi.org/10.1109/ACCESS.2020.2988510.

Chen, M., Liu, F., & Lee, YH. (2022). My Tutor is an AI: The Effects of Involvement and Tutor Type on Perceived Quality, Perceived Credibility, and Use Intention. In: Degen, H., Ntoa, S. (eds) Artificial Intelligence in HCI. HCII 2022. Lecture Notes in Computer Science, vol 13336. Springer, Cham. https://doi.org/10.1007/978-3-031-05643-7_15

Cifuentes, O., & Lents, N. (2010). Increasing Student-Teacher Interactions at an Urban Commuter Campus through Instant Messaging and Online Office Hours. Electronic Journal of Science Education. 14(1).

De Corte, E., Depaepe, F., Op't Eynde, P., & Verschaffel, L. (2011). Students' self-regulation of emotions in mathematics: An analysis of meta-emotional knowledge and skills. ZDM, 43, 483-495. https://doi.org/10.1007/s11858-011-0333-6

Ernest, P. (1996). Varieties of constructivism: A framework for comparison. In L. Steffe, P. Nesher, P. Cobb, G. Goldin, & B. Greer, (Eds.), Theories of mathematical learning (pp. 335–350). Mahwah, NJ: Lawrence Erlbaum.

Garrison, D.R. (2009). Communities of Inquiry in Online Learning. Encyclopedia of Distance Learning. 352-355. 10.4018/978-1-60566-198-8.ch052.

Garrison, D. R., Anderson, T., & Archer, W., (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. Internet and Higher Education 2(2-3), 87-105. https://doi.org/10.1016/S1096-7516(00)00016-6

Goodyear, P., Banks, S., Hodgson, V., & McConnell, D. (Eds.). (2004). Advances in research on networked learning. Dordrecht: Kluwer Academic Publishers.

Graesser AC., & Person NK. (1994). Question Asking During Tutoring. American Educational Research Journal, 31(1), 104-137. https://doi.org/10.3102/00028312031001104

Hanham, J., Lee, C., & Timothy, T. (2021). The influence of technology acceptance, academic self-efficacy, and gender on academic achievement through online tutoring. Computers & Education. 172. 104252. https://doi.org/10.1016/j.compedu.2021.104252

Hrastinski, S., Edman, A., Andersson, F., Kawnine, T., & Soames, C.-A. (2012). Informal math coaching by instant messaging: Two case studies of how university students coach K-12 students. Interactive Learning Environments, 22(1), 84–96. http://dx.doi.org/10.1080/10494820.2011.641682

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., ... & Chen, W. (2021). Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685.

Kim, W. -H. and Kim, J. -H. (2020) "Individualized AI Tutor Based on Developmental Learning Networks," in *IEEE Access*, vol. 8, pp. 27927-27937. http://doi.org/10.1109/ACCESS.2020.2972167.

Lin, CC., Huang, A.Y.Q. & Lu, O.H.T. Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review. *Smart Learn. Environ.* **10**, 41 (2023). https://doi.org/10.1186/s40561-023-00260-y

Mazzolini, M., & Maddison, S. (2003). Sage, guide or ghost? The effect of instructor intervention on student participation in online discussion forums. Computers & Education, 40(3), 237-253. https://doi.org/10.1016/S0360-1315(02)00129-X

Mazzolini, M., & Maddison, S. (2007). When to jump in: The role of the instructor in online discussion forums. Computers & Education, 49(2), 193-213. https://doi.org/10.1016/j.compedu.2005.06.011

Mousavinasab, E., Zarifsanaiey, N., R. Niakan Kalhori, S., Rakhshan, M., Keikha, L., & Ghazi Saeedi, M. (2021). Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods. *Interactive Learning Environments*, *29*(1), 142–163. https://doi.org/10.1080/10494820.2018.155825

Nielsen, J. (1993). Usability engineering. Boston: Academic Press.

Networked Learning Editorial Collective (NLEC). (2021). Networked learning: Inviting redefinition. *Postdigital Science and Education*, 3(2), 312-325. https://doi.org/10.1007/s42438-020-00167-8

Op't Eynde, P., De Corte, E., & Verschaffel, L. (2006). "Accepting emotional complexity": A socio-constructivist perspective on the role of emotions in the mathematics classroom. *Educational Studies in Mathematics*, 63 (2), 193-207. https://doi.org/10.1007/s10649-006-9034-4

Pellegrini M, Lake C, Neitzel A, & Slavin RE. (2021). Effective Programs in Elementary Mathematics: A Meta-Analysis. AERA Open. https://doi.org/10.1177/2332858420986211

Sharples, M. (2023). Towards social generative AI for education: theory, practices and ethics. *Learning: Research and Practice*. 9:2. 159-167. https://doi.org/10.1080/23735082.2023.2261131

Stenbom, S., Cleveland-Innes, M., & Hrastinski, S. (2016). Emotional Presence in a Relationship of Inquiry: The Case of One-to-One Online Math Coaching. Online Learning, 20(1), 41. https://doi.org/10.24059/olj.v20i1.563

VanLEHN, K. (2011) The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems, Educational Psychologist, 46:4, 197-221, DOI: 10.1080/00461520.2011.611369

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. Journal of Child Psychology and Psychiatry, 17(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Yang J, & Zhang B. (2019). Artificial Intelligence in Intelligent Tutoring Robots: A Systematic Review and Design Guidelines. *Applied Sciences*. 9(10). https://doi.org/10.3390/app9102078