DEGREE PROJECT IN TECHNOLOGY AND LEARNING

SECOND CYCLE, 30 CREDITS

# Exploring the possibility and implementation of AI-supported online math coaching

**KRISTIN KARLSTRÖM**

KTH Industrial Engineering and Management

**Master of Science Thesis TRITA-ITM-EX 2024:515**

# Exploring the possibility and implementation of AI in online math coaching

Kristin Karlström

| | | |
|---|---|---|
| Approved<br>2024-09-10 | Examiner<br>Stefan Hrastinski | Supervisor<br>Malin Jansson |
| | Commissioner<br>Stefan Hrastinski | Contact Person |

# Abstract

Recently, people have tested using large language models (LLM) such as ChatGPT in education. Studies and tests have been done to see where and how LLMs can be used in education, for example, when generating learning material. However, most studies focused on their use in a classroom setting or the students' use of LLMs to learn outside the classroom. But education reaches outside the school. One example of this is Mattecoach.se, where teacher students help students in primary and secondary school in Sweden with problems regarding their mathematics education. The Department of Digital Learning at KTH has developed an LLM prototype to help the math coaches at Mattecoach.se with their work. This study aimed to evaluate this LLM and discuss what can be done with it in the future. This was done by testing the LLM and ChatGPT on math coaches by having them act as math coaches and students to simulate a conversation while using an LLM tool to help them. The study showed that if the LLM used when teaching or coaching a student in mathematics does not have good mathematical capabilities it is likely to accept incorrect answers as correct from the student and give the student the wrong answers. As such, if an LLM is to be used to teach a student, it needs to either have excellent mathematical capabilities to be able to teach a student reliably or it needs to be able to teach without doing any calculations.

Keywords: Large Language Model, Online Coaching, AI in education, Math Coaching

# Sammanfattning

Inom de senaste åren har personer testat att använda stora språkmodeller (eng. large language models) likt ChatGPT i utbildning. Det har genomförts studier och tester för att undersöka möjligheten att använda stora språkmodeller inom utbildning i form av att exempelvis skapa lärandematerial. Dock har mycket av detta handlat om undervisning i klassrummet eller om hur elever själva använder stora språkmodeller för att lära sig. Men undervisningen sträcker sig även utanför klassrummet. Ett exempel av detta är Mattecoach.se där lärarstudenter hjälper elever i grundskolan och gymnasiet i Sverige med problem de har i matematikundervisningen. Avdelningen för Digitalt Lärande på KTH har utvecklat en prototyp av en stor språkmodell för att hjälpa mattecoacherna på Mattecoach.se med deras arbete. Syftet med denna studie är att utvärdera denna stora språkmodell och diskutera hur den kan utvecklas i framtiden. Detta skedde genom att göra tester med språkmodellen och ChatGPT där mattecoacher fick agera mattecoach och elever och efterlikna en konversation medan de använder en språkmodell som hjälpmedel. Studien visade att om språkmodellen som används vid utlärning av matematik inte har tillräckliga färdigheter inom matematik så riskerar den att acceptera inkorrekta svar som korrekta samt att ge fel svar på uppgifter. För att en stor språkmodell ska kunna användas för att lära ut matematik krävs det då att den antagligen har väldigt goda färdigheter inom matematik eller vara menad att lära ut matematik utan att göra några beräkningar.

Nyckelord: Stora Språkmodeller, Online Coachning, AI i Undervisning, Matte Coachning

# Acknowledgements

I want to thank all who have helped me during this journey, especially the math coaches who participated in this study. You helped with this work, and I was happy to see your interest. I want to thank those who worked on the Math Coach LLM before me and who will work on it in the future for the exciting use and application of AI.

A big thank you to my supervisor, Malin Jansson, for helping me during this long process. You gave me support and ideas when, for example, I was figuring out how to test the LLM.

Lastly, I want to thank my commissioner and examinator, Stefan Hrastinski, for giving me this opportunity and helping me during the process. It has been a fantastic opportunity, and I have enjoyed working on this study.

Kristin Karlström

Stockholm, August 2024

# Table of Contents

# 1. Introduction

One subject many students struggle with is mathematics, as seen by Skolverkets (n.d) statistics from 2018 to 2023. This statistic shows that among the three core subjects in Swedish schools, mathematics is the subject where most students cannot fulfil the criteria for grade 9. A way to support students with difficulties with their studies is through tutoring, which provides a more personalised experience (Kraft & Lovison, 2024). However, it may not be possible for students to have an in-person tutor. However, tutoring does not only happen in person. As such, students who cannot get an in-person tutor or prefer not to have one can get an online tutor. Studies have shown that online tutoring helps students learn mathematics (Beal et al., 2007; Kraft & Lovison, 2024; Nguyen & Kulm, 2005). Through an online tutor, students can get tutoring even if, for example, there is a lack of tutors in their area. In addition, there are positives with online tutoring that do not appear in in-person tutoring. One example of this is anonymity. Online interaction lets people be anonymous; some are more at ease and open with others if they can be anonymous (Lee & Wagner, 2002).

There are other ways to help a student with their education besides tutoring. Studies have been done regarding the use of large language models (LLMs) in education, such as generating lesson material or using LLMs to grade students' work (Yan et al., 2024). Some studies have been done specifically regarding different subjects in school and how LLMs can be used in them, such as the study done by Bonner et al. (2023) regarding language education. Another study focused on students using LLMs to supplement their calculus and statistics lessons in mathematics education (Calonge et al., 2023). This study focused on the potential usage of different LLMs as a tool that students can use to help them when they cannot ask teachers for help.

Similar to the use of LLMs in education, there are intelligent tutoring systems (ITS). There have been many studies regarding the use of ITS in education. One example was the Nihongo Tutorial System (Maciejewski & Leung, 1992), which was used to help people learn and understand more technical Japanese literature. Another example is an app that teaches children Korean, which builds on an AI tutor (Kim & Kim, 2020). Both examples are of a tutor meant to work without a course connected with a teacher to help and guide the student. However, intelligent tutoring systems can be used in combination with classes. One example comes from a study regarding how different Intelligent Tutoring Systems can help students with reading comprehension and writing (Jacovina & McNamara, 2017).

Studies have been done regarding the use of LLMs in education, but the use of LLMs in mathematics education may come with some difficulties. This is because LLMs are generally not very good at mathematics due to their understanding of mathematics as the statistical correlation between words (Yousefzadeh & Cao, 2023). They typically repeat what they have learned without understanding the underlying concepts. LLMs are continually evolving and getting better at different things. For example, there has been talk about OpenAIs GPT-4 model and how it may have this understanding. However, older models work with statistical correlation between words where the parts of mathematics are seen as different words.

However, there has not been much focus on using LLMs in a learning environment outside of the classroom or in math education, particularly on the tutor side rather than students' usage of LLMs. This study focuses on a learning environment outside the school, Math Coach, where students can ask questions when they do not have a teacher. Furthermore, the study focuses on how LLMs may be usable in mathematics education.

While students may need help with their mathematics education outside school, they may not have the opportunity or want to have an in-person or online tutor. Those students may need help with specific exercises and may be unable to ask their parents for help. In Sweden, there is a website called mattecoach.se or Math Coach where students can chat with engineering or teacher students online through text and ask for help concerning mathematics (*Mattecoach På Nätet*, n.d.). However, there is a limited number of students, and one math coach can help at a time.

There has been an interest in automating some parts of the process to give the math coaches the ability to help more students at a time without sacrificing the quality of the help and help the math coaches with their work in general. As such, an LLM is being developed to help math coaches by giving suggestions on what the math coaches can send to students. This is to help the math coaches by giving them an idea of what they can say to the student if they have difficulty formulating a response and guarantee that the response the math coach sends is high quality and fitting for the situation. The LLM can also decrease the response time of the math coaches. Furthermore, the LLM should help when a math coach is talking to two students simultaneously, as it will keep proper track of what has been said and what level the student is at. However, one crucial aspect is understanding what the math coaches who will use the LLM want and need. The math coaches have been part of developing the LLM from the beginning, where they got to test a simple version of an LLM to use in their work (Jansson et al., 2024). As such, they must continue to be a part of the LLM's development.

## 1.1 Aim

This study aims to evaluate the Math Coach LLM and what the math coaches want from it. The evaluation of the Math Coach LLM is based on difficulties such as giving wrong suggestions or answers or if the math coaches have any problems while using it. What are some areas of improvement in the Math Coach LLM, and how can it be improved.

### 1.1.1 Research Questions

This study has three research questions.

1. How do the math coaches use the Math Coach LLM?
2. What difficulties does the Math Coach LLM have?
3. How can the Math Coach LLM be adjusted to match the wants and needs of the math coaches?

# 2. Theoretical background

For this study, two areas need to be defined. One area concerns Math Coach and theories connected to Math Coach, such as tutoring and online tutoring. The other area concerns LLMs, such as what they can and cannot do and the theories regarding why that is.

## 2.1 Online Tutoring

According to Wood et al. (1976), tutoring is when someone more experienced helps someone less experienced. More specifically, the more experienced one helps the less experienced do something they do not have the skill to do. Often, this involves what is called scaffolding. This consists of a task outside the scope of what the less experienced one can do unassisted, where the more experienced one supports the less experienced. Thus, the less experienced can focus on what they can do and get the support they need in areas that are out of their scope.

Online tutors are tutors that are separated from their students (Denis et al., 2004). They work like in-person tutors, except for not meeting the students face to face. There are differences due to this, one of which is the possibility of anonymity on the students' side (McKenna et al., 2002). Anonymity can cause students to be more open to sharing their thoughts and feelings online than in person. Another difference is the absence of certain nonverbal cues in the form of facial expressions (Walther & D'Addario, 2001). In the case of online communication, where facial expressions are unavailable, emoticons have taken the role of facial expressions.

### 2.1.1 Pedagogical Perspective in Online Tutoring

Scaffolding connects to Lev Vygotsky's (1978) idea of the zone of proximal development. The zone of proximal development represents what someone can do on their own and what they can do with support from someone more knowledgeable or experienced. Scaffolding is when a student or less experienced works on a problem that is out of the scope of what they can do on their own but can do with the help of someone more experienced (Wood et al., 1976); it falls in the zone of proximal development (Säljö, 2015). The more the student knows, the more support the teacher or tutor gives can be removed until the student can do the task independently. The ideas of scaffolding and the zone of proximal development connect to the pedagogical theory of social constructionism, which focuses on the idea that knowledge is constructed through communication with others. In social constructionism, one learns to communicate better with others and their surroundings.

### 2.1.2 Math Coach

Math Coach is an example of student-to-student online coaching (Hrastinski & Stenbom, 2013). This happens when a student gets support from a more experienced student. This connects to Vygotsky's (1978) theory of a zone of proximal development. As such, the activities of Math Coach are in the zone of proximal development, which is what the student can do with the support of someone more experienced. The online coaching of Math Coach is inquiry-based, meaning that the math coaches minimise the direct instruction and instead work more with guided exploration (Stenbom et al., 2012). As such, the coaches do not give direct answers to students but instead are to guide the students into discovering the answers themselves.

## 2.2 Large Language Model (LLM)

LLM is short for Large Language Model, a category of AI trained on large amounts of text data to generate human-like responses (Routray et al., 2023). One example of an LLM is ChatGPT. LLMs are trained on massive amounts of text data. They then work by predicting the next word in a sequence using statistics. Once an LLM has been trained, it can be used for various natural language processing tasks, such as translating text and being a chatbot. The LLM's response depends on the text data it is given to learn as it calculates the probability of the next word in a sequence based on the text data it has been trained with.

### 2.2.1 Limitations of LLMs

LLMs have several limitations, one of which is that LLMs can be wrong (Routray et al., 2023; Tamkin et al., 2021). Sometimes, the response the LLM gives is inaccurate or factually wrong. Furthermore, the LLM is unable to evaluate the sources it has. As such, it is vital to control the response and fact-check. Another limitation with LLMs is that the responses they generate are based on the data they learned from and, as such, may be out of date. If a person asks an LLM about something happening in real-time, it cannot answer correctly as there is no data about it. Similarly, what the LLM knows is based on the training data; as such, all biases in the training data will be reflected in the LLM (Routray et al., 2023; Tamkin et al., 2021). Lastly, in several cases, LLMs have problems with reasoning and logic due to their statistical nature.

### 2.2.2 LLMs and Mathematics

LLMs' understanding of mathematics at their core is based on the correlation between words (Yousefzadeh & Cao, 2023). For some LLMs, it has been suggested that they have gained an understanding that reaches past this. One example is OpenAI’s GPT-4, where it has been suggested that it has such an understanding that reaches past correlation between words. The study by Yousefzadeh and Cao (2023) indicates that GPT-4 only understands proofs for mathematical formulas that are more widespread on the internet, while when asked about theorems that are not as widespread, it fails to answer them correctly. That suggests that it does not have a higher understanding of mathematics but instead relies on the training data that it has.

LLMs have one problem with mathematics: the precision needed (Satpute et al., 2024). Mathematics also uses specialised language and symbols. Unlike natural languages, mathematics relies on unmentioned rules and assumptions, so even the most significant LLM has problems with mathematical reasoning. GPT-4 performs better regarding mathematics than some of the other LLMs, but it still fails to consistently answer all mathematical questions accurately.

Another area where LLMs' mathematical capabilities fall is regarding math word problems (Srivatsa & Kochmar, 2024). Some areas of math word problems make them challenging for LLMs. For example, a high diversity of mathematical operations that use infrequent numerical tokens makes a math word problem challenging for LLMs. Furthermore, if the problem is long, has low readability, or relies on real-world knowledge, it also makes it hard for LLMs to solve correctly.

### 2.2.3 LLMs and Education

There are several studies regarding using LLMs in education (Bonner et al., 2023; Calonge et al., 2023; Huber et al., 2024). They have been aimed for use in different parts of education both inside and outside of the classroom, for example, self-studying languages and being a supplement when teaching reading comprehension. There are also LLMs created more to support education, such as Google’s Bard, now known as Gemini (Calonge et al., 2023). Bard was tailored towards education and, as such, used pedagogical approaches such as giving step-by-step instructions. On the other hand, Bard specialised in its curriculums and depended on them.

# 3. Method

The method of this study was divided into three parts. The first part covers data gathering, the second covers data analysis, and the third covers testing initial prompts. The method was discussed in section 5.4, Methodology Reflection.

## 3.1 Research Context

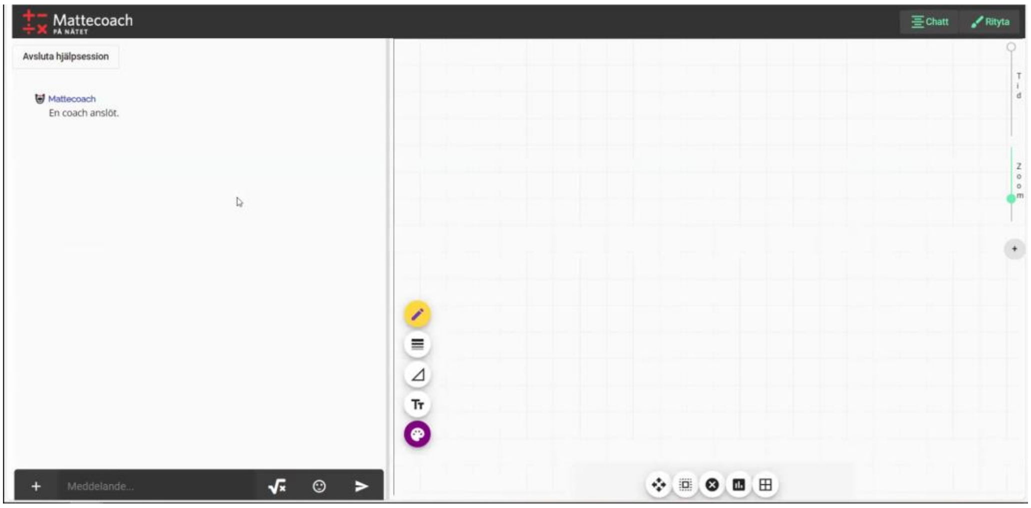
This study concerns an LLM meant to work with math coaches at Math Coach. As such, the research context is both Math Coach and the Math Coach LLM being worked on.

### 3.1.1 Math Coach

Mattecoach.se or Math Coach helps students in primary school up to upper secondary school with difficulties regarding mathematics in school. This is done using an online text-based chat and a whiteboard, as shown in Figure 1. The coaches are university students studying to become teachers who have done at least one course in mathematics didactics. The math coaches have also taken courses in online coaching before starting their work as math coaches. (Stenbom et al., 2016)

**Figure 1**

Screenshot of Math Coach’s chat room. To the left is a chat, and to the right is a whiteboard.

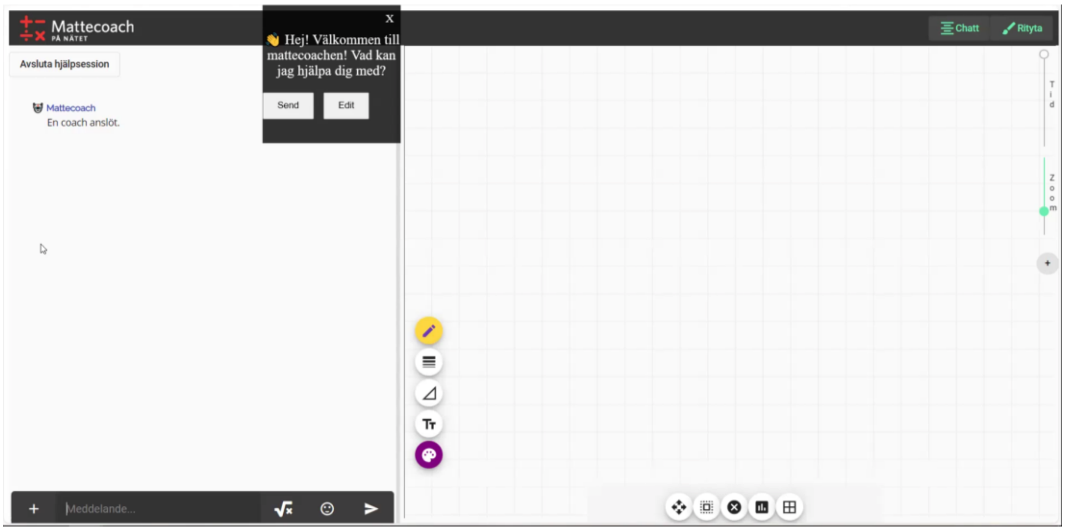


Mattecoach
En coach anslöt.

### 3.1.2 The Math Coach LLM

The LLM developed for Math Coach uses Llama2, an LLM developed by Meta, and is trained on Swedish text to communicate and further fine-tuned on Math Coach conversations. It works by suggesting a message for the math coach to send to the student based on the conversation. It also suggests a message at the beginning of the conversation, as shown in Figure 2. The Math Coach can edit or send the message suggested by pressing the respective button. The math coaches can close the window with the suggested response by pressing the x in the upper right corner of the black box in Figure 2. When editing a response, the math coaches can save a response, and the message shown in the black box in Figure 2 would change to what the math coach had written. The math coach can send or edit the response further or close the window. Although the LLM uses Llama2, there is talk regarding changing it to GPT-4.

**Figure 2**
Illustration of the Math Coach program with the LLM active.



Mattecoach
En coach anslöt.

👋 Hej! Välkommen till mattecoachen! Vad kan jag hjälpa dig med?

## 3.2 Data Gathering

The data gathering consisted of two tests in the form of a roleplay, both with group interviews afterwards. The tests occurred on two different occasions two days apart. Six math coaches in total took part, three during each test. All math coaches had at least one year of experience as math coaches. The tests were done by having three math coaches take turns acting as students with a math problem and asking Math Coach for help. One of the math coaches was to help the students with their problems. The tests differed here; for the first test, the math coach used ChatGPT 3.5 to help the students, while the second test used the Math Coach LLM. ChatGPT was chosen to approximate what it would be like to use GPT-4. They were encouraged to use the tool as much as they felt comfortable and to note things they noticed. With the math coaches ' consent, the tests and group interviews were recorded using Zoom on a computer with screenshare. The screenshare was of the math coach's computer. During the tests, the observer, sitting beside the math coach, asked the acting math coach clarifying questions and pointed things out to encourage them to discuss and comment on using the tool. The recordings were transcribed after the tests.

### 3.2.1 The Test Using ChatGPT

During the test, the acting math coach used ChatGPT as a tool to help two acting students simultaneously. The acting math coach and the acting student sat in two rooms to avoid affecting each other. There were no restrictions on how the math coach used ChatGPT, and they were allowed to use it as they saw fit. As such, they asked ChatGPT for the solution to the student's problem, and after that, they asked clarifying questions. The acting math coach changed after the acting students had been helped with their issues until all three math coaches had the opportunity to act as math coaches. After the test, the math coaches were interviewed in a qualitative semi-structured group interview. As such, it was a group interview with more open questions decided beforehand where the interviewees could be more thorough with their answers than they could in a structured interview. (Holme & Solvang, 1997). This was to encourage discussions between the interviewees so that they could build on each other's answers. They were asked whether ChatGPT helped them and how it was used. They were also asked if they noticed anything in particular when acting as students.

### 3.2.2 The Test Using the Math Coach LLM

During this test, the acting math coach helped one acting student at the time as the LLM was only usable on one computer and not between two computers. As such, the acting math coach and the acting student had to share one computer. The acting math coach was allowed to use the Math Coach LLM as they saw fit but were encouraged to use it as much as possible. They could send the suggested response directly to the student, edit the response before they send it, or ignore the suggested response and write their own. The math coach was to change when the student's problem was complete, but due to time constrictions, this did not happen, and it was prioritised that all math coaches could act as math coaches. After the test, the math coaches were interviewed in a qualitative semi-structured group interview similar to the test with ChatGPT. In this case, there were additional questions regarding how the interface was in Math Coach with the LLM, if there was anything they would like to change about the interface, how they liked the LLM and its responses, and if there was anything they noted about the suggested responses such as if the content was wrong at any point. They were also asked if they noted anything when acting as a student.

### 3.2.3 Possible Biases

The testing may have been biased because all the math coaches knew that ChatGPT or the LLM was used. As such, the answers on the acting student side may have depended on that knowledge. One possible bias may be the assumption that using the Math Coach LLM will decrease the response time. As such, the acting student could say that they noticed that the math coach answered faster, while the math coach may have never used the Math Coach LLM during the test. Possible bias was considered when analysing the results, and the recording was used to control what was said during the interview.

## 3.3 Data Analysis

The interviews were analysed using deductive analysis (Bingham & Witkowsky, 2022). This was done by creating codes and later sorting the data into categories based on the codes. The codes for the deductive analysis were based on the research questions: how the math coaches used the LLM as well as how they felt it could be used, what difficulties the LLM had according to the math coaches, and what the math coaches would want to change so that they could use it better. From the results of the data gathering and the analysis, some points of improvement were identified.

### 3.4 Prompt Testing

An LLM can be adjusted by changing the initial prompt. As such, adjusting the initial prompt could solve some difficulties with the Math Coach LLM. This was done by testing new initial prompts using ChatGPT. The testing was done by creating a base message to send to ChatGPT based on the initial prompts of the Math Coach LLM and then creating a new initial prompt by adding a sentence to the base message. The sentence added was a new initial prompt to solve the difficulty with the Math Coach LLM. The base message was:

*Hello. Can you act as a math teacher and help me with a math exercise? Do not give me the answer directly but ask leading questions to help me solve the exercise.*

The testing was done by recreating the problem from the Math Coach LLM test; if that did not work, the plan was to use a similar exercise where ChatGPT had difficulties. Once the situation where the Math Coach LLM had a problem was recreated, a new initial prompt that consisted of the base message with a new sentence added was used. The initial prompt was adjusted for every test, and any differences were documented. After the tests, a conclusion was made.

### 3.5 Ethical Considerations

The math coaches were asked for their consent regarding recording both tests and the group interviews and were informed of what the recording would be used for and what the study was regarding. Furthermore, they are not described or named in this study and, as such, are anonymous. The data gathered may not accurately reflect the results that could have been if the math coaches were to use the Math Coach LLM freely in their ordinary work at Math Coach as it may have been affected by both knowing that there was no real student on the other side and by having an observer sitting beside them. As the observer took part in the test by asking questions and talking to the acting math coach, it may have affected what the acting math coach did as the observer pointed out things of note that the math coach possibly may not have taken notice of on their own.

## 4. Results

The results are divided into two parts. The first part, the data gathering, showed that the math coaches were willing to use the LLMs as tools during the tests. There were difficulties regarding the LLMs' ability to solve math word problems and regarding their mathematical capabilities, notably in that the Math Coach LLM regarded an incorrect answer as correct when interacting with the student. The math coaches had several ideas on how to improve regarding the Math Coach LLM such as making sure that the suggested messages were short and not too long. The second part, the initial prompt testing, showed that changing the initial prompt did not affect ChatGPT’s response when answering incorrectly.

### 4.1 Data Gathering

The data-gathering process consists of two parts. The first part is the ChatGPT testing, and the second is the LLM testing.

### 4.1.1 Testing using ChatGPT

In the first test, the math coaches used ChatGPT to aid them when helping students. The result of this test is divided into two parts: content and usage. The content part focuses on the answers from ChatGPT, for example, if ChatGPT answered incorrectly. The usage part focuses on how the math coaches used the answers from ChatGPT, for instance, if they copied the answers directly, formulated a response using the answer, or asked clarifying questions to ChatGPT and used what they got from that.

#### *4.1.1.1 Content*

At the start of the conversation with the student, the math coaches put in the question they were asked in ChatGPT. ChatGPT answered with the solution to the exercise and did not answer as a math coach. In most cases, the math coaches could use the answers they got to help the student. But there were some errors with the answers. In some cases, the math was correct, but in others, the math was wrong. One example was when the question involved writing numbers using scientific notation. When asking ChatGPT for the solution, the math coach noted that Chat GPT had, at one point, answered that the base for 0.0208 was 208, which was incorrect. However, the coach asked ChatGPT several different questions regarding this exercise, and sometimes, the answer was accurate: The base for 0.0208 was 2.08. Similarly, during another exercise, the math coach noted that ChatGPT did all the steps of a solution correctly, but the final calculation was off every time. The calculation it failed with was $29{,}000 \times 1.06^3$, which is approximately 34,539, which ChatGPT answered was 34,554 and when refreshed 34,568.

#### *4.1.1.2 Usage*

The coaches mostly used ChatGPT to get an explanation of how to solve the exercise. Then, they formulated responses by taking them into account. One math coach asked ChatGPT to explain how to solve the question asked and sent the answer from ChatGPT to the student directly, but it was commented that it was a lot of text. In general, the coaches felt that it was too much text to be able to send it directly to the student. ChatGPT also, in most cases, gave the entire answer, and as the math coaches wanted the student to solve the exercise themselves, it was not ideal to send the solution and how to solve it to the student. One math coach used ChatGPT when a student got stuck by asking how one can explain why the value used in an exponential function where the percentual increase each year is 6% is 1.06 and not 0.06. The math coach used this to help explain to the student how to solve the problem.

### 4.1.2 Testing using the Math Coach LLM

In the second test, the math coaches used the Math Coach LLM. The result of this test was divided into two parts: content and usage. The content part focuses on the responses suggested by the Math Coach LLM, for example, if it suggested a response with incorrect information. The usage part focuses on how the math coaches used the suggested responses, such as whether they preferred to send the responses as is or send completely different responses.

#### *4.1.2.1 Content*

The responses that the Math Coach LLM suggested were mostly responses without any specific information regarding the exercise. There were no significant errors in the responses.

However, the Math Coach LLM took the student's answer as correct when they got an answer. It did this even when the answer was wrong. Furthermore, it was repetitive. After a while, the responses it suggested were almost the same. The math coaches had to go in and change the wording a little whenever it came to this as they felt that it would not be appropriate to say, for example, “Good! What’s next?” all the time. Similarly, when the student answered, it did not follow up in any way, for example, asking the student to summarise how they solved the problem or anything similar.

#### *4.1.2.2 Usage*

The math coaches used the responses suggested by the Math Coach LLM at the beginning of the help session. The math coaches often edited the responses that the Math Coach LLM suggested and sometimes ignored them completely. Mostly, the math coach edited when, for example, the response gave part of the answer instead of leading the student to the answer or asking the student what they knew. For instance, if the student was unsure about a problem with integrals, the Math Coach LLM suggested a response that explained how to solve the exercise. The math coach ignored the suggested response and asked the students what they knew about integrals. The coaches most used the suggested responses at the beginning of the help session when welcoming the students and asking what they needed help with.

### 4.1.3 Group Interviews

The two group interviews were analysed using deductive analysis and divided into three parts: usage, difficulties, and possible improvements. The possible improvements part consists only of responses from the second interview regarding possible enhancements to the Math Coch LLM based on using it. As such, there was no discussion regarding possible improvements during the first interview.

#### *4.1.3.1 Usage*

For the first test, the math coaches commented that they did not use ChatGPT much directly but instead based what they wrote on what ChatGPT had answered. That was because when they asked ChatGPT how to solve an exercise, ChatGPT gave the whole solution. The math coaches did not want to send that to the students as they were to guide them and not answer immediately. However, one math coach noted that ChatGPT could be used as a supplement when the coaches get asked about something they have not worked with for a while. They can then ask ChatGPT how to solve such an exercise without searching online. What one math coach said more precisely was:

> *What I think that it can help us the most with, or if it is such a question, is just this start. To start with the exercise. That one gets started with it quicker. Because now if one gets a type of exercise that one has not thought about in long, then one needs to first Google it, read up on it. But if you get a complete strategy from the beginning then the process of starting will go much faster which makes it so that we can surely be more effective in our work.*

Another math coach commented that ChatGPT helped a bit by giving steps to solve an exercise and that it was there where it could help them gain the most time.

For the second test, the math coaches felt that most of the time, the messages suggested by the Math Coach LLM were not good enough to send on their own. As such, they edited the messages or sent entirely new messages. On the other hand, they noted that the Math Coach LLM did better at the beginning of a conversation during the introduction and getting started with the exercise but was not as good for later parts when working with it. However, they commented that there may be parts of their conversation with students where the Math Coach LLM may be helpful. That part is, in particular, the end, where math coaches send a link to a form. The form is unique to the chat ID, and as such, it would be easier if the Math Coach LLM could handle the message with the link as it has all the information needed. Besides that, one math coach felt that the Math Coach LLM could be good, as it hinted at where they were when they last sent a message in the chat.

#### *4.1.3.2 Difficulties*

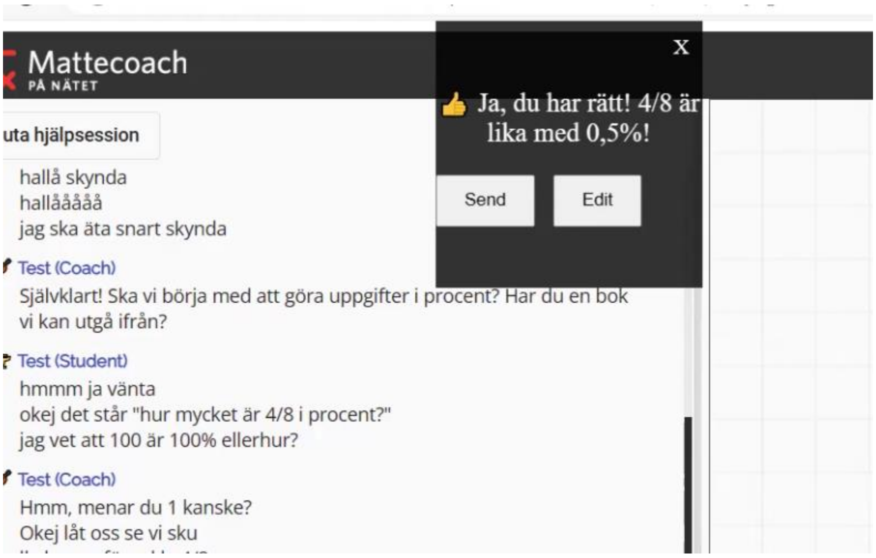
There were several difficulties that ChatGPT had that the math coaches commented on:

- It was unreliable regarding calculations and gave a wrong answer. One math coach said that ChatGPT had all the right steps until the final calculation, but it never got right. Another case was regarding an exercise with a scientific notation where ChatGPT got several calculations wrong. However, the coaches noted that the steps to the answer were correct, so they did not feel it was a big deal.
- In one case, they had an exercise where ChatGPT did not register part of the question. When asking ChatGPT to show how to solve this question, it ignored the first part of the exercise and focused solely on the latter part. When asking further questions about the first part, it reiterated what it had answered regarding the latter part. The math coach commented that, in the end, they had to reformulate the request they had sent to ChatGPT not to include the parts of the question. They formulated it not as asking for clarification regarding the exercise but instead asking outside of the exercise.
- In one case, it gave both the wrong solution and the wrong answer. This differed from an earlier problem where ChatGPT gave an incorrect answer but had a correct solution. The problem came when the math coach got confused by what ChatGPT had answered and was unsure if they understood the question themselves.

There were also some difficulties during the second test:

- Sometimes the messages suggested were a bit long and could, in those cases, be shorter.
- One problem that could have caused a more significant problem was that, at one point, the student gave a wrong answer that the Math Coach LLM accepted as correct, which can be seen in Figure 3. The math coach noted during the interview that they almost sent that message and that if something like that had happened in a conversation with an actual student, it would have been bad. They noted that it would negatively affect students' trust in their math coach if the math coach made such mistakes that they had to apologise for often.

**Figure 3**
An example of when the Math Coach LLM accepted a wrong answer as correct. The suggested response translates to *Yes, you are right! 4/8 is equal to 0.5%!*



👍 Ja, du har rätt! 4/8 är lika med 0,5%!

hallå skynda
hallåååååå
jag ska äta snart skynda

Test (Coach)
Självklart! Ska vi börja med att göra uppgifter i procent? Har du en bok vi kan utgå ifrån?

Test (Student)
hmmm ja vänta
okej det står "hur mycket är 4/8 i procent?"
jag vet att 100 är 100% ellerhur?

Test (Coach)
Hmm, menar du 1 kanske?
Okej låt oss se vi sku

### *4.1.3.3 Possible Improvements*

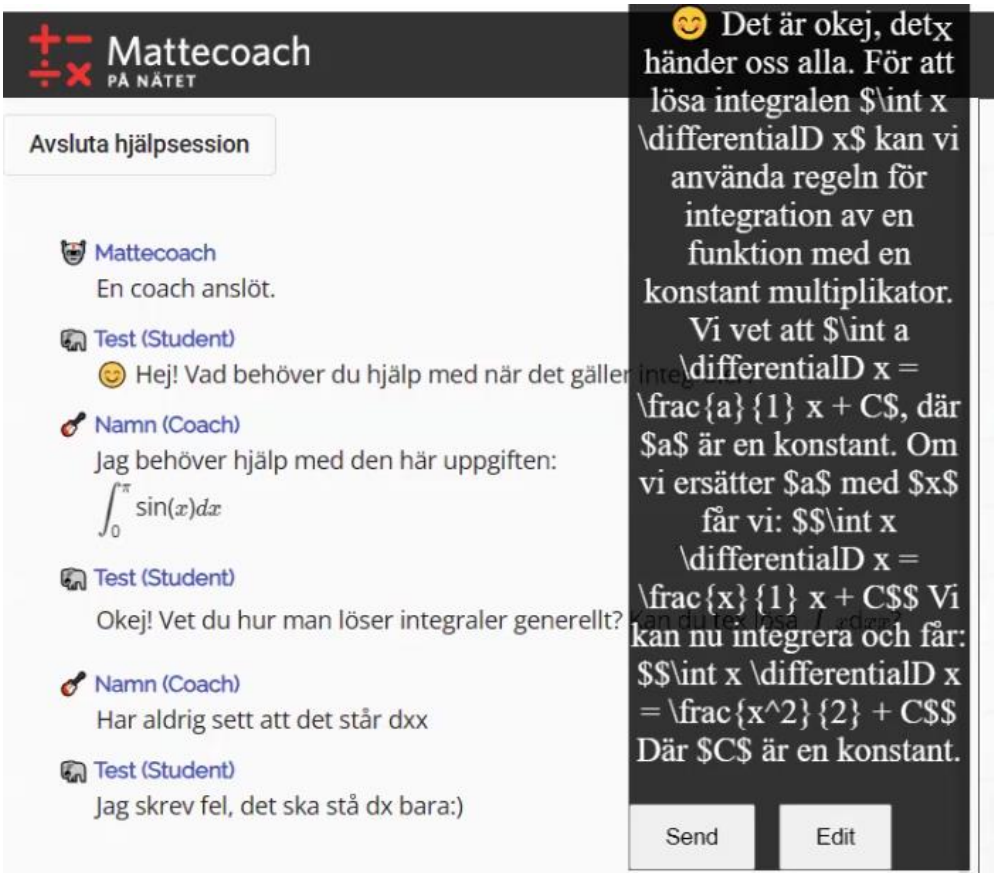
The math coaches had two ideas for improvements regarding the Math Coach LLM and several regarding the interface for the Math Coach LLM. The two improvements for the Math Coach LLM were having the Math Coach LLM create the survey links and ensuring it keeps the messages short. They said that it would be good if the Math Coach LLM could be used to generate the survey links they send at the end of the conversation, as it would be nice if that were more automatic. The Math Coach LLM was generally good at keeping the messages short, but on some occasions during the test, it suggested a lengthy response. As such, it was noted that one thing that could be improved was ensuring that it does not suggest long responses.

At several points, the math coaches felt the user interface of Math Coach could be improved.

- While the suggested responses could contain LaTeX code, there was no way to see how it would look when sent in the chat. As such, one suggestion for improvement was that the window in Figure 3 could show a preview of what will be sent and, as such, not show the LaTeX code as it currently does, as shown in Figure 4.

**Figure 4**

An example of a suggested response with LaTeX code.



Mattecoach
En coach anslöt.

Test (Student)
😊 Hej! Vad behöver du hjälp med när det gäller inte

Namn (Coach)
Jag behöver hjälp med den här uppgiften:
$\int_0^\pi \sin(x)dx$

Test (Student)
Okej! Vet du hur man löser integraler generellt? Kan du tex lösa

Namn (Coach)
Har aldrig sett att det står dxx

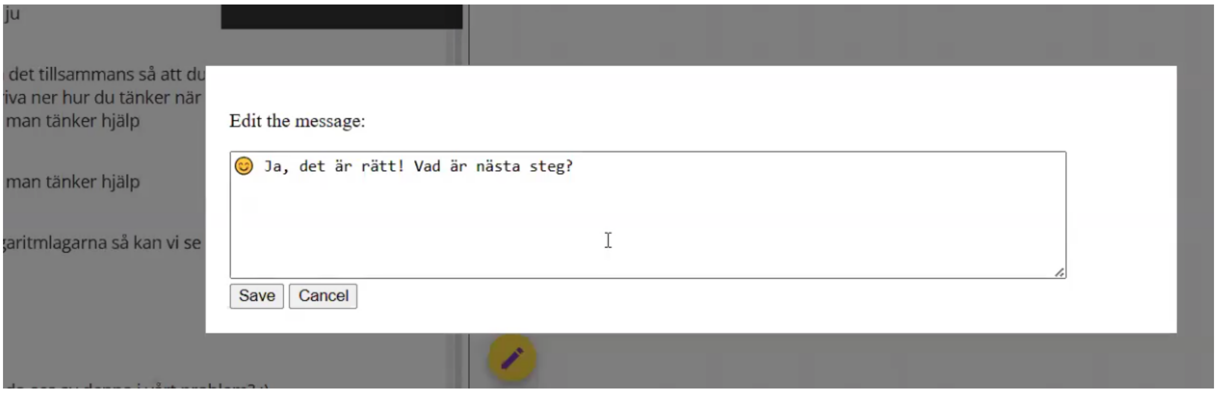
Test (Student)
Jag skrev fel, det ska stå dx bara:)

😊 Det är okej, det händer oss alla. För att lösa integralen \$\int x \differentialD x\$ kan vi använda regeln för integration av en funktion med en konstant multiplikator. Vi vet att \$\int a \differentialD x = \frac{a}{1} x + C\$, där \$a\$ är en konstant. Om vi ersätter \$a\$ med \$x\$ får vi: \$\$\int x \differentialD x = \frac{x}{1} x + C\$\$ Vi kan nu integrera och får: \$\$\int x \differentialD x = \frac{x^2}{2} + C\$\$ Där \$C\$ är en konstant.

- When testing, there was an occasion when the suggested response was so long that one could not see the send or edit button unless one zoomed out. This revealed two points that could be improved.
    - The ability to scroll in the window in exchange for keeping it at a fixed size. If the window is fixed size and one can scroll in it, there is less of a problem if the response is too long.
    - A smaller font size. According to one of the math coaches, the font size did not need to be so big, as the size of the messages in Math Coach was smaller.
- When editing, the math coaches would like the ability to scroll in the math coach chat and move the window where they edit messages. The math coaches said they sometimes like to look through the messages sent while formulating a response and cannot do that with the current window. As shown in Figure 5, the window may also

hide part of the messages and what is drawn on the whiteboard when editing a response. However, the math coaches generally liked the editing window as they could easily see what they had written.

**Figure 5**
The window that is shown when editing a suggested response from the LLM.



[redacted]

ju

det tillsammans så att du
iva ner hur du tänker när
man tänker hjälp

man tänker hjälp

garitmlagarna så kan vi se

Edit the message:

😊 Ja, det är rätt! Vad är nästa steg?

## 4.2 Prompt Testing

A priority regarding improvements was to solve the problem of the LLM accepting wrong answers as correct. This was because the math coach commented that they almost sent the response even though it was incorrect. They commented that it may cause problems if they must apologise for saying something wrong, which may lead to students losing trust in them. To improve this problem, tests were done with ChatGPT to determine whether the initial prompt could be adjusted to catch when an answer is wrong. To test this, a base message was first sent to ChatGPT, asking it to act as a mathematics teacher and ask leading questions to help solve a mathematics exercise. Then, the problem was recreated by asking ChatGPT to help solve the 4/8 in percentage, the exercise where the problem occurred with the Math Coach LLM during the tests. The answer of 4/8, being 0.5%, was sent to ChatGPT, to which it explained that the answer was wrong. As such, the testing had to be done with another exercise. The exercise chosen was an exercise from the first test, to which ChatGPT gave the wrong answer during the test. The question was:

*I have an exercise where I am supposed to write numbers in scientific notation. The numbers are a) 50000, b) 1910000, c) 0.00004, and d) 0.0208. How do I solve it?*

The goal was then to have ChatGPT accept a wrong answer. This was done by suggesting that 50000 in scientific notation was $5 \times 10^3$ when it is supposed to be $5 \times 10^4$. Then the base message was adjusted by adding the sentence: Keep in thought that I can be wrong so please check my answers. During the tests, different variations of such a sentence were used to try to force ChatGPT to either check the answer given or, as it turned to in later tests, ask to explain the answer. The last initial prompt was:

*Hello. Can you act as a math teacher and help me with a math exercise? Do not give me the answer directly but ask leading questions to help me solve the exercise. Go through my solution with me when I give an answer.*

This was a try to make it so that the student at least must repeat how they came to the answer they came to instead of giving the response, “Exactly! 50000 in scientific notation is $5 \times 10^3$“ or similar. In some cases, ChatGPT, in a way, ignored the response and acted as if the answer given was $5 \times 10^4$. If corrected with what was answered, it changed the response to that being correct. As such, none of the tested initial prompts worked to solve the problem.

# 5. Discussion

This study had three research questions: how math coaches used the LLM, what difficulties there were with it, and what can be done to make it match what the math coaches want. This section will discuss these and the methodology, limitations, and future directions, both practical and theoretical.

## 5.1 The usage of the LLM

The tests showed that math coaches were willing to use AI as help when coaching. Generally, the responses were edited before being sent. During the tests where ChatGPT was used, it was utilised as a tool to get an idea of how to solve an exercise, and then the responses were based on the answers. The group interview for the test using ChatGPT showed that ChatGPT answered with too much text during the tests, as it was commented on when a response was copied directly from ChatGPT. During the test with the Math Coach LLM, the math coaches sent a few messages directly but often erased parts of the message or wrote a new one entirely. This was because the Math Coach LLM did better when suggesting a response at the beginning of a conversation rather than any later parts, as explained by the interview. As such, the LLM did better during more regular conversations rather than helping a student solve an exercise. This can be explained by LLMs being trained more to be chatbots than math coaches (Routray et al., 2023). Furthermore, the latter parts of the conversations concerned the scaffolding of the exercise, where the coach determined what support the students needed and gave it to them (Wood et al., 1976). To help the student, the coach must determine what the student knows and does not know, thus determining the student's zone of proximal development (Vygotsky, 1978). This is more specialised, and as LLMs generally are not trained for education, one may need an LLM aimed at education to use it when coaching students.

## 5.2 The difficulties of the LLM

One difficulty that both ChatGPT and the Math Coach LLM had problems with was mathematical capabilities similar to what the study by Satpute et al. (2024) showed. ChatGPT could explain all the steps of a solution but failed to solve the exercise, while the Math Coach LLM accepted the wrong answer as correct. This is similar to what Satpute et al. (2024) discussed regarding LLMs having trouble with the precision needed for mathematics. In general, the Math Coach LLM did not calculate a lot as it is trained to give leading questions and, as such, does not give the answer to the student. As such, the Math Coach LLM’s mathematic capabilities were not shown as much during the test except when it was accepted that 4/8 was equal to 0.5%. Furthermore, ChatGPT has problems with math word problems where it did not acknowledge a part of the problem in one case and, in another case, could not solve the problem correctly. While it was able to solve the question regarding the money exchange (Exercise 1 in Appendix), it was not able to even get the steps right regarding the population change (Exercise 2 in Appendix), as shown by the math coach getting confused about how to solve the exercise after reading the answer on how to solve it from ChatGPT. The problem regarding the math word problem with the population change could be that it required real-world knowledge and contained different values (Srivatsa & Kochmar, 2024). There was real-world knowledge needed for what 6000 people moving out each year meant regarding the question of change per day. Then, there were three different sorts of numbers: a base population of 500,000, a change per day of 20, and a change per year of 6000. This could have caused ChatGPT to have a problem solving the problem.

## 5.3 Possible improvements to the Math Coach LLM

Work is needed with the suggested messages and the user interface to make the Math Coach LLM better suit the math coaches' needs. Regarding the user interface, the group interview provided some suggestions for improvements to the message preview in the way both see the preview of the message they are to send and a sort of fail-safe so that it is always easy to press the edit and send buttons. There was also a suggestion to create a function to see what has been sent before in the chat when writing a response in the editing window.

Regarding the messages, the length and mathematical capabilities are primarily a problem. The tests showed that there was further work to do to make sure that the messages were shorter. The test with ChatGPT showed that it was noticeable to the recipient when the messages were directly from ChatGPT due to their length. During the tests with the Math Coach LLM, there were a few occasions during the actual tests where the suggested responses were too long, but it was noted during the group interview after the test to ensure that the messages were not too long. There was a problem regarding the Math Coach LLM accepting a wrong answer as correct, and from testing, it is uncertain if that can be solved by changing the initial prompt. The testing showed that while the Math Coach LLM had problems with the exercise, with 4/8 equal to 0.5%, ChatGPT did not. However, ChatGPT did have issues with another exercise in a similar way to the Math Coach LLM. As such, the problem with the Math Coach LLM accepting a wrong solution as correct may be solved if it has greater mathematical capabilities. This connects to LLMs' general problem: their inability to fact-check themselves (Routray et al., 2023). As such, it may have been trained with sources that say that 4/8 equals 0.5% or at least something similar. It will accept that as true. It may also be that it is trained with many examples of 4/8 being equal to 0.5 and not many of 4/8 not equal to 0.5%. As such, it may lean on the first to say it is equal. With more focus on mathematical capabilities, it may better identify if an answer is incorrect.

There were no significant concerns outside the Math Coach LLM's mathematical capabilities. As shown in Figures 2 and 3, it used emoticons like math coaches do as nonverbal cues to the students (Walther & D’Addario, 2001). Math coaches regularly do this. The Math Coach LLM could also write LaTeX code and send messages like math coaches. During the interview, it was noted that the Math Coach LLM sounded like a math coach.

## 5.4 Methodology Reflection

The usage of testing the Math Coach LLM and ChatGPT by having math coaches act as students and math coaches was good. The math coaches knew what questions students ask and how they behave, and as such, they could simulate a conversation that at least somewhat was like one. There were math coaches on both sides of the screen, so they were more at ease sending messages they might typically not send. If the math coaches had been chatting with real students, they would not send as many messages that were suggested as they did during the tests. An example was during the first test with ChatGPT when the coach copied text from ChatGPT and sent it. The math coach acting as a student commented on a lot of text, and considering that the math coaches wanted the Math Coach LLM to keep it short, they most likely avoided sending long messages to students. As such, if it were with a real student, the math coach most likely would not have sent it. However, for this study, sending messages they may have been somewhat unsure of was good as it gave data to work with.

One thing that could have been done better regarding the testing is using the base message used later in the study during the testing with ChatGPT. During the testing with ChatGPT, the math coaches never got to use it, similar to how the Math Coach LLM was used. They asked ChatGPT for a solution, and then they had to pick out what they could use from what ChatGPT sent. When working with initial prompts later in the study, it turned out that starting the conversation with ChatGPT with a similar initial prompt as the Math Coach LLM worked similarly. It could have helped with the testing if something like that had been used so that the two tests could have been better compared.

Another thing that could have been done was to interview math coaches about what they would like of a Math Coach LLM. This was done during the initial part of the development of the Math Coach LLM (Jansson et al., 2024), but it could be interesting to hear how the ideas of the math coaches changed by testing the Math Coach LLM. They would have a more nuanced view of which of their ideas are possible, and it would give more opinions to math coaches regarding what they want of the LLM.

The initial prompt's testing would have been better done with the actual Math Coach LLM, as it would have shown if there was any real change. The testing may not accurately reflect what it would be with the Math Coach LLM, but due to limitations, such as the cost of running the Math Coach LLM, the testing was done with ChatGPT.

## 5.5 Future Studies

This study has shown several ways to explore the area of LLMs in mathematics education. Some ways are connected directly to the work in this study, such as future work with the Math Coach LLM. However, there are also some possible directions for research outside of the scope of the Math Coach LLM.

### 5.5.1 Directions for the Math Coach LLM

For future work with the Math Coach LLM, it is good to continue to work closely with the math coaches. In this study, some suggestions from math coaches have been brought up, but they do not have to reflect the opinions of all math coaches. As such, for example, the improvements to the interface the math coaches suggested are examples of what three math coaches want.

One possible thing to work on regarding the Math Coach LLM is what LLM to use in the background. During this study, it was Llama2 due to limitations regarding the fact that the LLM needed to be run and store data locally, but a better LLM may exist to use instead. For example, GPT-4 is said to understand mathematics better, and Gemini may be better for explaining step-by-step how to solve exercises (Calonge et al., 2023). Some of the difficulties identified during this study may be solved by using a different LLM as a base, for example, as shown from the prompt testing where ChatGPT did not have the same problems with an exercise as the Math Coach LLM had.

Another thing to work on is further improving the possible usage of the LLM regarding the beginning and end of a conversation. The study showed that the Math Coach LLM did the best during the first part of a conversation, welcoming the students and asking them what they needed help with. There was potential there to automatise this. The study also showed that there might be some potential in automating the latter part of a conversation where the coach wraps everything up in the way of, for example, the automatic survey in the Math Coach's case. These parts of the conversation may not contain as much coaching but regular conversation, which LLMs are more suited for, considering their current use as chatbots (Routray et al., 2023).

### 5.5.2 Research Directions

There are several directions this could go in the future. One is working on an AI that can act as a mathematics teacher. A base message was used to test initial prompts with ChatGPT during this study. The message asked ChatGPT to act as a mathematics teacher and help with an exercise. Similarly, the Math Coach LLM has an initial prompt that it is a mathematics teacher. As such, it would be interesting to see if an AI can act as a mathematics teacher. Considering the results of this study, it would need to be able to guide students without needing to calculate, as ChatGPT showed that an LLM can give the steps to solve an exercise but is more uncertain regarding calculations. As such, it could guide the student through the steps to solve an exercise but does not tell if an answer is right or wrong. Maybe it could ask the students when they came to an answer about how they solved it to give them time to think through their solution and find if they made any miscalculations. It could also work in an entirely different way.

## 5.6 Limitations of the Study

There were some limitations regarding this study. One limitation was that the study started when the one who had developed the Math Coach LLM was leaving the project. As such, there was a lot of work to do with changing who had control over the work. That caused the work of this study to fall behind a bit. Furthermore, a decision was made to change which LLM was the base for the Math Coach LLM, possibly to GPT-4, which was part of why a test was done using ChatGPT. This caused some problems as this study worked with the assumption that the development would continue with what was there, and there was an uncertainty about what work in this study could be directly transferred to work in the future. As such, parts of this study were adjusted with this in mind. One last limitation came from the fact that the Math Coach LLM that was worked on during this study was still in the hands of the one who developed it. As such, there had to be communication with someone outside of Math Coach and the Department of Digital Learning on KTH to be able to use the Math Coach LLM. Due to this, the fact that there was talk of potentially changing the LLM from Llama2 to GPT-4, and the fact that running the Math Coach LLM cost money, ChatGPT was used for testing further than the tests with the math coaches.

# 6. Conclusions

In conclusion, this study evaluated the Math Coach LLM and found some key areas for future work to improve the Math Coach LLM. This was done by addressing three research questions: how did math coaches use the Math Coach LLM, what difficulties did the Math Coach LLM have, and how can the Math Coach LLM be adjusted to fit what the math coaches want? The results showed that the math coaches were willing to use the Math Coach LLM in a controlled situation where no actual students from their work were involved. They mostly edited the response before sending it and did not often send it as is. Still, they did note that the suggested responses at the beginning of a conversation were good, and that was the part of the conversation where the suggested responses were used the most without changing them. There were some difficulties with calculations and repetition. The Math Coach LLM did not do many calculations, as it mostly suggested leading questions, but it showed when it accepted a wrong answer. If a math coach sends such a suggested response, it could cause problems as they become more unreliable in the eyes of the students. The Math Coach LLM often repeated itself, and the suggested response had to be adjusted somewhat each time to make it less repetitive. The group interviews of this study showed some things to work on regarding the Math Coach LLM. Most were regarding the interface, such as being able to have the suggested responses in the preview window match what would be seen in the chat in cases of LaTeX code being used and either a fixed size and the possibility to scroll in the preview window or smaller font size for the preview window as a failsafe if the suggested response is too long. Regarding other areas, there was a suggested feature for automatically generating links to an end survey to send to the student at the end of a conversation and some way to ensure that the suggested responses are not too long. For future directions regarding the Math Coach LLM, one could work on implementing the suggested improvements or possibly compare different LLMs to evaluate which LLM would be the best fit for Math Coach.

# References

Beal, C., Walles, R., Arroyo, I., & Woolf, B. (2007). On-line Tutoring for Math Achievement Testing: A Controlled Evaluation. *Journal of Interactive Online Learning*, *6*.

Bingham, A. J., & Witkowsky, P. (2022). Deductive and inductive approaches to qualitative data analysis. In C. Vanover, P. Mihas, & J. Saldana (Eds.), *Analyzing and interpreting qualitative data: After the interview*. Sage Publications.

Bonner, E., Lege, R., & Frazier, E. (2023). Large Language Model-Based Artificial Intelligence in the Language Classroom: Practical Ideas for Teaching. *Teaching English with Technology*, *23*(1), 23–41.

Calonge, D. S., Smail, L., & Kamalov, F. (2023). Enough of the chit-chat: A comparative analysis of four AI chatbots for calculus and statistics. *Journal of Applied Learning and Teaching*, *6*(2), Article 2. https://doi.org/10.37074/jalt.2023.6.2.22

Denis, B., Watland, P., Pirotte, S., & Verday, N. (2004). Roles and Competencies of the e-Tutor. *Proceedings of the Networked Learning Conference (NLC 2004), Lancaster, UK, April*.

Holme, I. M., & Solvang, B. K. (1997). *Forskningsmetodik: Om Kvalitativa Och Kvantitativa Metoder* (2nd ed.). Studentlitteratur.

Hrastinski, S., & Stenbom, S. (2013). Student–student online coaching: Conceptualizing an emerging learning activity. *The Internet and Higher Education*, *16*, 66–69. https://doi.org/10.1016/j.iheduc.2012.02.003

Huber, S. E., Kiili, K., Nebel, S., Ryan, R. M., Sailer, M., & Ninaus, M. (2024). Leveraging the Potential of Large Language Models in Education Through Playful and Game-Based Learning. *Educational Psychology Review*, *36*(1), 25. https://doi.org/10.1007/s10648-024-09868-z

Jacovina, M. E., & McNamara, D. S. (2017). Intelligent Tutoring Systems for Literacy: Existing Technologies and Continuing Challenges. In *Grantee Submission*. https://eric.ed.gov/?id=ED577131

Jansson, M., Tian, K., Hrastinski, S., & Engwall, O. (2024). An initial exploration of semi-automated tutoring: How AI could be used as support for online human tutors. *Networked Learning Conference*, *14*. https://doi.org/10.54337/nlc.v14i1.8070

Kim, W.-H., & Kim, J.-H. (2020). Individualized AI Tutor Based on Developmental Learning Networks. *IEEE Access*, *8*, 1–1. https://doi.org/10.1109/ACCESS.2020.2972167

Kraft, M. A., & Lovison, V. S. (2024). The Effect of Student-Tutor Ratios: Experimental Evidence from a Pilot Online Math Tutoring Program. EdWorkingPaper No. 24-976. In *Annenberg Institute for School Reform at Brown University*. Annenberg Institute for School Reform at Brown UniversityBrown University Box 1985, Providence, RI

02912http://www.annenberginstitute.orgTel.: 401-863-7990, Fax: 401-863-1290. https://www.proquest.com/eric/docview/3075706868/98C1B3B24C674124PQ/4?sourcetype=Reports

Lee, V., & Wagner, H. (2002). The Effect of Social Presence on the Facial and Verbal Expression of Emotion and the Interrelationships Among Emotion Components. *Journal of Nonverbal Behavior*, *26*(1), 3–25. https://doi.org/10.1023/A:1014479919684

Maciejewski, A. A., & Leung, N. K. (1992). The Nihongo Tutorial System: An Intelligent Tutoring System for Technical Japanese Language Instruction. In *CALICO Journal* (Vol. 9, Issue 3, pp. 5–25). https://www.proquest.com/eric/docview/62852138/A8D5DD63B8DF49ECPQ/8?sourcetype=Reports

*Mattecoach på nätet*. (n.d.). Retrieved 20 March 2024, from https://www.mattecoach.se/

McKenna, K. Y. A., Green, A. S., & Gleason, M. E. J. (2002). Relationship Formation on the Internet: What's the Big Attraction? *Journal of Social Issues*, *58*(1), 9–31. https://doi.org/10.1111/1540-4560.00246

Nguyen, D. M., & Kulm, G. (2005). Using Web-Based Practice to Enhance Mathematics Learning and Achievement. *Journal of Interactive Online Learning*, *3*(3).

Routray, S. K., Javali, A., Sharmila, K. P., Jha, M. K., Pappa, M., & Singh, M. (2023). Large Language Models (LLMs): Hypes and Realities. *2023 International Conference on Computer Science and Emerging Technologies (CSET)*, 1–6. https://doi.org/10.1109/CSET58993.2023.10346621

Säljö, R. (2015). *Lärande – en introduktion till perspektiv och metaforer*. Gleerups.

Satpute, A., Giessing, N., Greiner-Petter, A., Schubotz, M., Teschke, O., Aizawa, A., & Gipp, B. (2024). *Can LLMs Master Math? Investigating Large Language Models on Math Stack Exchange* (arXiv:2404.00344). arXiv. https://doi.org/10.48550/arXiv.2404.00344

Srivatsa, K. A., & Kochmar, E. (2024). *What Makes Math Word Problems Challenging for LLMs?* (arXiv:2403.11369). arXiv. https://doi.org/10.48550/arXiv.2403.11369

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), Article 1. https://doi.org/10.24059/olj.v20i1.563

Stenbom, S., Hrastinski, S., & Cleveland-Innes, M. (2012). Student-Student Online Coaching as a Relationship of Inquiry: An Exploratory Study from the Coach Perspective. *Journal of Asynchronous Learning Networks*, *16*(5). https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-104457

Tamkin, A., Brundage, M., Clark, J., & Ganguli, D. (2021). *Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models* (arXiv:2102.02503). arXiv. https://doi.org/10.48550/arXiv.2102.02503

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press.

Walther, J. B., & D'Addario, K. P. (2001). The Impacts of Emoticons on Message Interpretation in Computer-Mediated Communication. *Social Science Computer Review, 19*(3), 324–347. https://doi.org/10.1177/089443930101900307

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

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, *55*(1), 90–112. https://doi.org/10.1111/bjet.13370

Yousefzadeh, R., & Cao, X. (2023). *Large Language Models' Understanding of Math: Source Criticism and Extrapolation* (arXiv:2311.07618). arXiv. https://doi.org/10.48550/arXiv.2311.07618

# Appendix

## Exercise 1

Therese is going to exchange money for a trip to London. She finds two exchange offices with different pricing models. Exchange office A: rate 11.74 SEK/£ and no fixed charge. Exchange office B: 11.64 SEK/£ and fixed charge of 40 SEK. In which intervals (expressed in £) is it beneficial to exchange in exchange office A, and when is it beneficial to exchange in exchange office B? Which exchange office should Therese choose if she has 5000 SEK to buy £?

## Exercise 2

In a city with 500,000 residents, around 20 people move to the city per day. Every year around 6000 choose to move from the city a) describe the change per day with a linear equation according to the straight-line equation, b) does the population increase or decrease each year?

TRITA – ITM-EX 2024:515
Stockholm, Sweden 2024

www.kth.se