Contextualizing Problems to Student Interests at Scale in Intelligent Tutoring System Using Large Language Models

Gautam Yadav[1], Ying-Jui Tseng[1] and Xiaolin Ni[1]

[1]*Carnegie Mellon University, 5000 Forbes Ave Pittsburgh PA 15213, United States*

Abstract

Contextualizing problems to align with student interests can significantly improve learning outcomes. However, this task often presents scalability challenges due to resource and time constraints. Recent advancements in Large Language Models (LLMs) like GPT-4 [1]offer potential solutions to these issues. This study explores the ability of GPT-4 in the contextualization of problems within CTAT [2], an authoring tool for the intelligent tutoring system, aiming to increase student engagement and enhance learning outcomes. Through iterative prompt engineering, we achieved meaningful contextualization that preserved the difficulty and original intent of the problem, thereby not altering values or overcomplicating the questions. To evaluate the effectiveness of these newly generated questions, we conducted focus groups and interviews with instructional designers. The positive assessment from the instructional designers signifies that these questions are suitable for implementation, potentially saving significant time spent on manual contextualization. Despite these promising findings, we acknowledge current limitations of our research, particularly with geometry problems, and emphasize the need for ongoing evaluation and research. Future work includes systematic studies to measure the impact of this tool on students' learning outcomes and enhancements to handle a broader range of problems.

Keywords

Large Language Models, Mass Production, Student Interests, Intelligent Tutoring System

1. Introduction

Research has demonstrated that integrating problem contextualization with student interests can significantly enhance learning outcomes in algebra, resulting in increased proficiency in problem-solving, improved accuracy, and the ability to transfer to future learning [3].

Teachers, who intimately comprehend their students' interests, often find the task of contextualizing problems according to these interests challenging, since the scalability of such task is often met with resource and time constraints. However, recent developments in Large Language Models (LLMs) may provide an opportunity to lessen the strains associated with the personalization of learning context for students.

This research aims to explore the capability of LLMs in contextualizing problems to align with student interests at a large scale within CTAT [2], an authoring tool for the intelligent tutoring system. In this study, we perform experiments using one of the most advanced LLMs

currently accessible, the GPT-4 [1], obtained via the OpenAI API. Our hypothesis suggests that the application of LLMs for problem contextualization, based on student interests, could result in increased student engagement and enhanced learning outcomes.

2. Prior Work

2.1. Context Personalization

The works of Walkington [3] introduced the concept of contextualizing algebraic questions based on students' interests. This innovative methodology, featuring student-created "algebra stories," aimed to boost engagement, cultivate ownership, and enhance understanding of algebraic principles. Interest has been identified as a pivotal factor in learning, impacting attention, persistence, and motivation. Personalized learning that mirrors individual interests has demonstrated a capacity to elicit positive emotional responses, enhance appreciation for instructional content, and leverage existing knowledge. [4, 5]

The efficacy of context personalization was investigated using both qualitative and quantitative research methods, demonstrating a positive association between these 'algebra stories' and improved student engagement and performance. Despite possible implementation obstacles due to the diversity of learners' interests, the use of digital tools has been proposed as a facilitative means for this personalization process. In its totality, contextual personalization has the potential to enhance learning effectiveness and accuracy, decrease the practice required for mastery, and foster transferable skills applicable to various scenarios.

2.2. Mass Production in Intelligent Tutoring Systems

Mass Production in Intelligent Tutoring Systems (ITS) is a technique that enables authors to parameterize previously authored problem-specific content, which can then be instantiated to suit a multitude of different problems. This technique essentially permits authors to manually generalize Example-Tracing expert models (known as behavior graphs) to accommodate all problems that share isomorphic solution structures.

The Mass Production technique in Intelligent Tutoring Systems (ITS) is leveraged to generate a multitude of diverse problems from a single, parameterized structure, enhancing mastery learning by allowing repeated practice in various contexts[6] . In our work, we utilized this principle within the CTAT platform, adjusting only the contextual 'cover stories' to individualize problems in alignment with student interests. This personalized approach, derived from mass production, could potentially foster increased student engagement, improved problem-solving skills, and enhanced subject matter understanding.

2.3. Instruction Generation with Large Language Models

Previous research involving large language models has explored their application in educational settings, such as the use of models like GPT for generating questions or providing hints/explanations to students [7]. Empirical evaluations of these applications and their impact on student

outcomes suggested that students perform better on human-generated content compared to these models [8, 9].

It is evident that large language models hold great potential in enhancing learning experiences, making them a promising tool for future educational endeavors. However, as our work proposes, a step further in personalized learning can be taken by leveraging these models in a more context-aware manner, which could further improve student engagement and outcomes.

3. System Design

To test the potential of GPT-4 in problem contextualization, we developed a prototype tool integrated with CTAT, illustrated by a few problem examples. This paper primarily focuses on outlining our approach and tool design. However, we recognize the necessity for rigorous evaluation. Future work includes systematic studies such as randomized controlled trials to comprehensively assess the tool's efficacy and its impact on students' learning outcomes.

3.1. Iterative Prompt Engineering in GPT-4

Our iterative prompt engineering was executed over four different problem sets of algebra questions in TutorShop [10], with the objective of achieving meaningful contextualization aligned with diverse interests. A few-shot learning approach was used, drawing on examples from a range of contexts and interests to foster coverage and generalization. Throughout this process, we gradually refined and added rules based on testing until our output could aptly accommodate novel interests such as TikTok and NBA, as demonstrated in Tables 1 and 2.

3.1.1. Prompt

We used the following prompt:

- Your task is to change context based on interest for a problem, for example:
  - Input Problem 1:
    Chaz and Nikki are standing in a long line to buy rock concert tickets. Nikki is 8 feet ahead of Chaz in the line. Let's compare Chaz's distance to Nikki's distance from the front of the line.
    When Nikki is 20 feet from the front of the line, how far away is Chaz?
    When Nikki is 16 feet from the front of the line, how far away is Chaz?
    In the row labeled "Expression", define a variable for Nikki's distance and use that variable to write an expression that will allow you to calculate Chaz's distance.

    Output Problem 1 based on interest "Video Games":
    In a video game, two players, Mario and Luigi, are standing at different points in a level. Luigi is 8 units ahead of Mario in the game. Let's compare Mario's distance to Luigi's distance from the level's end.
    When Luigi is 20 units from the end of the level, how far away is Mario?
    When Luigi is 16 units from the end of the level, how far away is Mario?

In the row labeled "Expression", define a variable for Mario's distance and use that variable to write an expression that will allow you to calculate Luigi's distance.

Output Problem 1 based on interest "basketball":
During a basketball game, two players, Jordan and Kobe, are standing at different positions on the court. Jordan is 12 feet ahead of Kobe on the court. Let's compare Jordan's distance to Kobe's distance from the basket.
When Kobe is 20 feet away from the basket, how far away is Jordan from the basket?
When Kobe is 16 feet away from the basket, how far away is Jordan from the basket?
In the row labeled "Expression", define a variable for Kobe's distance and use that variable to write an expression that will allow you to calculate Jordan's distance.

– Input Problem 2:
You are a product inspector for a company that produces light bulbs. You find that two out of every 300 bulbs are defective: they don't work properly.

Output Problem 2 based on interest "World of Warcraft":
You enjoy playing World of Warcraft on your computer. You notice that two out of every 300 times you defeat a monster, the monster has an epic item: a treasure that you want to collect.

– Input Problem 3:
y = 80 - 6x
If x = 10, what is y?
If x = 7, what is y?
If y = 8, what is x?
Write a story that could go along with the equation y = 80 - 6x.

Output Problem 3 based on interest "Video Games":
You are playing your favorite war game on the Xbox 360. When you started playing today, there were 80 enemies left in the locust horde. You kill an average of 6 enemies every minute.
(a) How many enemies are left after 10 minutes?
(b) How many enemies are left after 7 minutes?
(c) Write an algebra rule that represents this situation using symbols.
(d) If there are only 8 enemies left, how long have you been playing today?

Now give output for

– input problem: 2x+3=15
– Interest: [The interest that the problem needs to be contextualized for.]
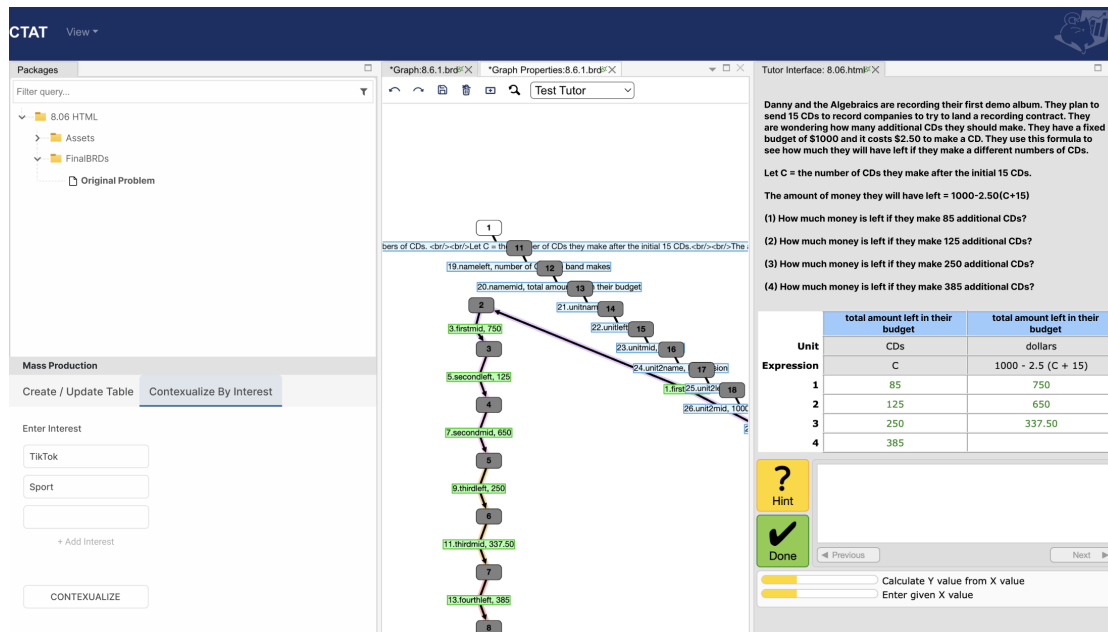
Some rules to follow:

1. don't change values

Figure 1: Mock-up depiction of the proposed extension to the CTAT system. Users can input their interests via the 'Contextualize By Interest' tab, which informs the system to generate personalized problem instances. **Note: The feature is currently under development and is not yet operational in the live system.

2. we want to have deeper contextualization not surface details based on Using Adaptive Learning Technologies to Personalize Instruction to Student Interests: The Impact of Relevant Contexts on Performance and Learning Outcomes
3. output question should ask same thing as input question, don't ask any additional question or complicate the info by adding unnecessary details

This strict adherence to rules ensures that we maintain consistency in problem difficulty and preserve the problem's original intent. This methodology respects the principle of not altering values or over-complicating the question by adding unnecessary details as observed in our earlier iterations.

3.2. CTAT Implementation

In this section, we propose a novel interaction design for contextualizing problems in Intelligent Tutoring Systems using CTAT and GPT-4 that emphasizes problem-authoring control. Teachers or instructional designers could contextualize existing problems simply by adding interest in the "Contextualized by Interest" tab in the Mass Production feature (Figure 1). After the user click the contextualize problem, the system will use GPT-4 and the prompt we mentioned in the prompt engineering section to generate variations of the problem for each interest. They can also preview and edit the contextualized result in the student-facing interface on the right panel to make sure whether they are satisfied with the generation result (Figure 2).
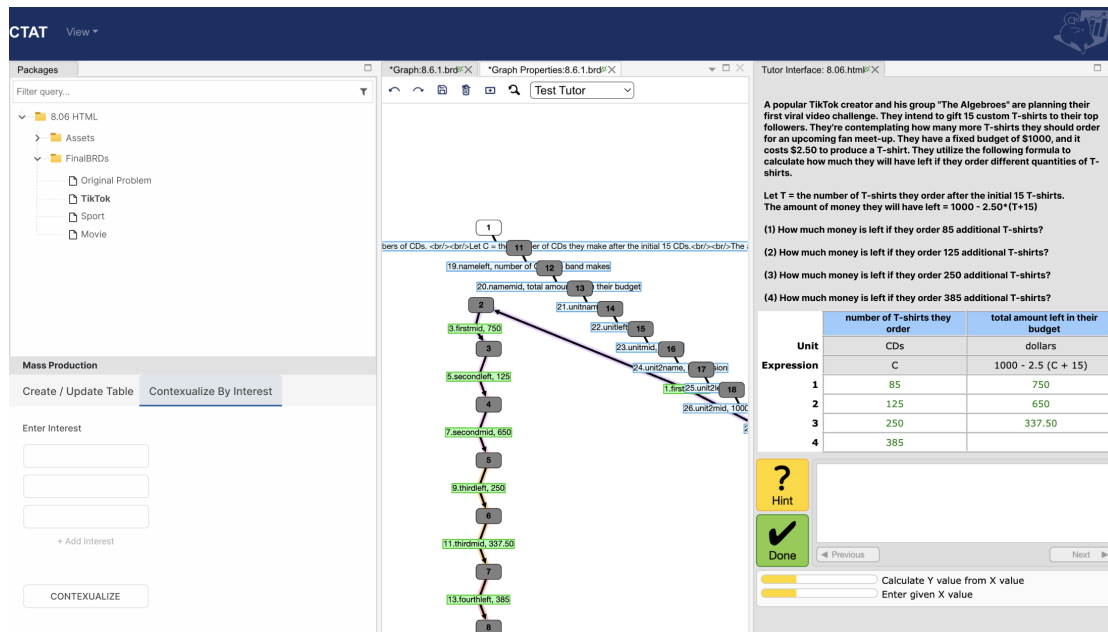
Figure 2: Depiction of the proposed extension to the CTAT system. After users click the 'CONTEXTU-ALIZE' button, they can preview and edit the contextualized problems. This figure demonstrates how the system adapts an original algebra problem based on 'TikTok' interest. **Note: this is a conceptual design and not yet a functional feature within the live system.

The application of mass production in CTAT offers significant value, as it facilitates the creation of a vast array of distinct problems using the same underlying structure. This contributes to mastery learning, allowing learners to practice similar problems in various contexts, ultimately strengthening their grasp of the subject [6]. In our approach, we utilized this principle, where only the contextual 'cover stories' were varied for the problem within the CTAT platform. This delivers similar problem-solving opportunities to students, yet personalizes these scenarios to align with their individual interests. The implications of this mass production approach based on interests are manifold; it can potentially increase student engagement, improve problem-solving abilities, and promote a better understanding of the subject matter.

4. Future Work and Limitations

While our approach benefits significantly from established pedagogical and technological foundations, it notably leverages the Mass Production facility to enhance problem diversity within existing problem sets. By utilizing mass production, we can incorporate existing cover stories and contextualized hints as examples in our few-shot learning approach with GPT-4, providing a coherent and contextual learning experience.

Systematic studies to measure the impact of this tool on students' learning outcomes remain a crucial part of our future work. This includes investigating improvements in initial response accuracy, learning efficiency enhancements, and the rate of progress towards proficiency.

However, we acknowledge certain limitations in our current model, particularly in handling geometry problems involving graphs, tables, or when diagrams would be beneficial. Despite its proficiency in algebra, GPT-4's present capabilities fall short in generating images that align with the problem text, hindering its effectiveness in scenarios where visual aids could significantly contribute to a student's comprehension of concepts. Future work will focus on overcoming these limitations to better facilitate student understanding across diverse problem types.

Acknowledgments

References

[1] OpenAI, Gpt-4 technical report, 2023. `arXiv:2303.08774`.

[2] V. Aleven, B. M. McLaren, J. Sewall, M. Van Velsen, O. Popescu, S. Demi, M. Ringenberg, K. R. Koedinger, Example-tracing tutors: Intelligent tutor development for non-programmers, International Journal of Artificial Intelligence in Education 26 (2016) 224–269.

[3] C. A. Walkington, Using adaptive learning technologies to personalize instruction to student interests: The impact of relevant contexts on performance and learning outcomes., Journal of educational psychology 105 (2013) 932.

[4] S. Hidi, K. A. Renninger, The four-phase model of interest development, Educational psychologist 41 (2006) 111–127.

[5] R. L. Goldstone, J. Y. Son, The transfer of scientific principles using concrete and idealized simulations, The Journal of the learning sciences 14 (2005) 69–110.

[6] C. J. MacLellan, K. R. Koedinger, Domain-general tutor authoring with apprentice learner models, International Journal of Artificial Intelligence in Education (2022) 1–42.

[7] S. Elkins, E. Kochmar, J. C. Cheung, I. Serban, How useful are educational questions generated by large language models?, arXiv preprint arXiv:2304.06638 (2023).

[8] E. Prihar, M. Lee, M. Hopman, A. Kalai, S. Vempala, A. Wang, G. Wickline, N. Heffernan, Comparing different approaches to generating mathematics explanations using large language models, in: Proceedings of the AIED2023 Conference, 2023. To be published.

[9] Z. A. Pardos, S. Bhandari, Learning gain differences between chatgpt and human tutor generated algebra hints, arXiv preprint arXiv:2302.06871 (2023).

[10] V. Aleven, J. Sewall, O. Popescu, F. Xhakaj, D. Chand, R. Baker, Y. Wang, G. Siemens, C. Rosé, D. Gasevic, The beginning of a beautiful friendship? intelligent tutoring systems and moocs, in: Artificial Intelligence in Education: 17th International Conference, AIED 2015, Madrid, Spain, June 22-26, 2015. Proceedings 17, Springer, 2015, pp. 525–528.

Table 1

Contextualized Algebra Problem with the cover story based on interests Tiktok and NBA

Interest	Question
Original	Danny and the Algebraics are recording their first demo album. They plan to send 15 CDs to record companies to try to land a recording contract. They are wondering how many additional CDs they should make. They have a fixed budget of $1000 and it costs $2.50 to make a CD. They use this formula to see how much they will have left if they make a different number of CDs. Let C be the number of CDs they make after the initial 15 CDs. The amount of money they will have left = 1000 - 2.50$(C + 15)$ 1. How much money is left if they make 85 additional CDs? 2. How much money is left if they make 125 additional CDs? 3. How much money is left if they make 250 additional CDs? 4. How much money is left if they make 385 additional CDs?
TikTok	An upcoming TikTok creator, Danny and his creative team, are planning to promote their first viral video. They plan to run a paid advertising campaign to 15 prominent TikTok influencers to try to gain more followers. They are wondering how many more influencers they should reach out to. They have a fixed budget of $1000 and it costs $2.50 to run an ad per influencer. They use this formula to see how much they will have left if they promote to a different number of influencers. Let I = the number of influencers they target after the initial 15 influencers. The amount of money they will have left = 1000-2.50$(I + 15)$ 1. How much money is left if they reach out to 85 additional influencers? 2. How much money is left if they reach out to 125 additional influencers? 3. How much money is left if they reach out to 250 additional influencers? 4. How much money is left if they reach out to 385 additional influencers?
NBA	The Lakers are planning their season's training sessions. They have a budget of $1000 to buy basketballs for practice. Each basketball costs $2.50. They are wondering how many additional basketballs they should buy. They already plan to buy 15 basketballs to start off. They use this formula to see how much they will have left if they buy a different number of basketballs. Let B = the number of basketballs they buy after the initial 15 basketballs. The amount of money they will have left = 1000-2.50$(B + 15)$ 1. How much money is left if they buy 85 additional basketballs? 2. How much money is left if they buy 125 additional basketballs? 3. How much money is left if they buy 250 additional basketballs? 4. How much money is left if they buy 385 additional basketballs?

Table 2

Contextualized Algebra Problem without cover story based on interests Tiktok and NBA

Interest	Question
Original	$2x + 3 = 15$
TikTok	In the realm of TikTok, you're working on a new viral challenge. You have found a way to get 2 new followers every time you post a video. Additionally, you already have 3 followers to start with. Your aim is to have 15 followers. Create an equation to show how many TikTok videos 'x' you need to post to reach your follower goal.
NBA	During an NBA game, a player earns points for their team by scoring baskets. Each 2-point field goal adds 2 points and every free throw adds a single point to the team's total. Imagine a situation where a player, LeBron, makes a certain number of 2-point field goals and 3 successful free throws, resulting in 15 points for his team. Write an equation that would help determine the number of 2-point field goals LeBron made. Use 'x' to denote the number of 2-point field goals.