Thetawise: Redefining AI-Powered Math Tutoring

In the evolving landscape of AI-driven education, not all AI tutors are created equal. While models like ChatGPT, Claude, and Gemini demonstrate impressive problem-solving capabilities, they often fall short in delivering structured, effective tutoring experiences. Thetawise is changing that! 🥳

Through rigorous evaluations across 27 pedagogical dimensions[3], we have developed a superior AI tutoring system that significantly outperforms standard large language models (LLMs). Our approach enhances engagement, fosters critical thinking, and ensures students receive clear, structured guidance tailored to their learning needs.

I. The Importance of Evaluating AI Tutors

An effective AI tutor does more than provide correct answers: it must facilitate learning. Inspired by Google's LearnLM research[1], we established a comprehensive framework to assess AI tutoring effectiveness across key teaching principles, including:

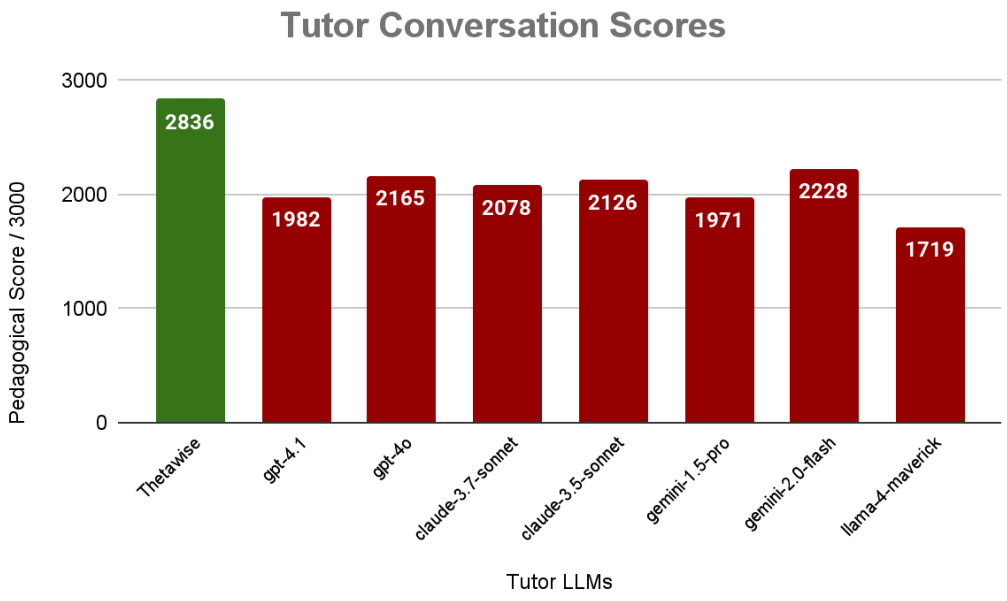
- Clarity and logical progression of explanations
- Encouragement of active student engagement
- Adaptability to individual learning styles
- Accuracy and precision in mathematical reasoning
- Structured, concise feedback delivery

Rather than evaluating AI performance solely on problem-solving accuracy, our methodology involves **simulated tutoring sessions**, where an AI tutor interacts with an AI student. A separate AI critic then assesses the session, scoring it against pedagogical best practices. This process provides a robust, real-world measure of an AI tutor's effectiveness.

II. Thetawise vs. Standard LLMs: Quantifiable Superiority

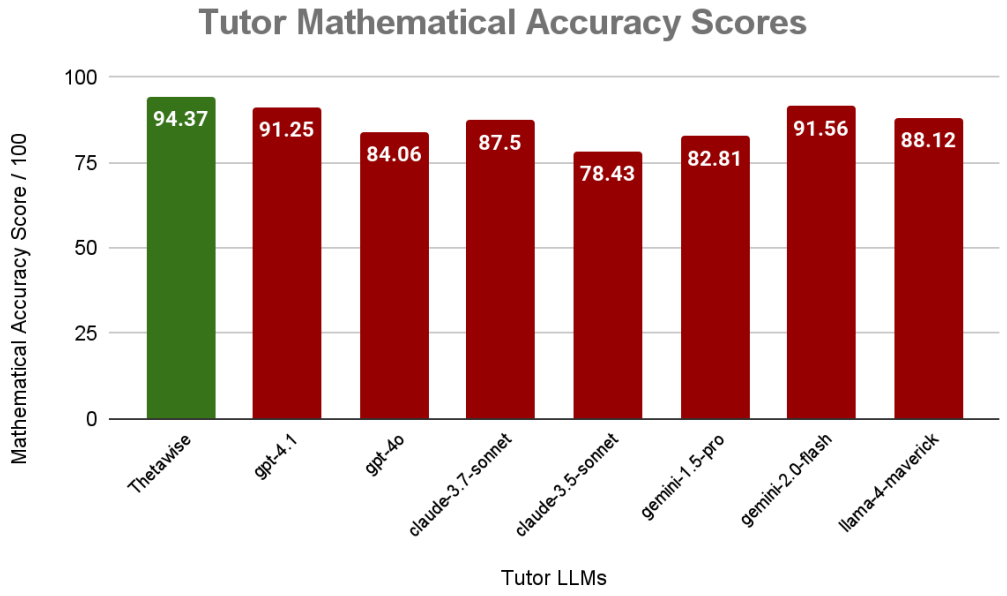
We tested Thetawise's "Tutor Mode" against baseline versions of GPT-4.1, GPT-4o, Claude 3.7 Sonnet, Claude 3.5 Sonnet, Gemini 1.5 Pro, Gemini 2.0 Flash, and Llama 4 Maverick. The results were definitive.

Tutoring Effectiveness Scores (Higher is better, out of 3000)



This represents an improvement of up to **65% in tutoring quality** by leveraging Thetawise's proprietary system prompt and optimization techniques.

Mathematical Accuracy Scores (Higher is better, out of 100)



By refining how LLMs process mathematical reasoning and leveraging computational tools, Thetawise not only enhances the quality of tutoring but also **improves problem-solving accuracy**.

However, some math problems require an extra level of reasoning and verification beyond standard tutoring. That's where our **Advanced Solver** comes in: designed to handle especially complex mathematical challenges, it integrates multiple specialized models, including OpenAI's **o4-mini Chain-of-Thought model**, to provide deeper exploration and understanding.

III. What Sets Thetawise Apart?

Thetawise's superior performance is driven by several key innovations:

- **Advanced Tutor Mode Optimization** - Our meticulously crafted system prompt ensures that AI tutors adhere to fundamental pedagogical principles, promoting clarity, engagement, and logical structuring.
- **Integrated Mathematical Reasoning & Visualization Tools** - Thetawise utilizes coding and visualization tools to enhance conceptual understanding and provide accurate solutions.
- **Dynamic and Adaptive Teaching** - Unlike standard LLMs, Thetawise's AI adjusts its teaching style based on student responses, ensuring a more personalized learning experience.
- **Streamlined Explanations & Constructive Feedback** - Thetawise minimizes unnecessary repetition, delivering clear, structured feedback that keeps students engaged and progressing efficiently.

IV. Technical Implementation Details

Thetawise's AI tutoring evaluation framework is built upon a multi-model architecture that integrates various LLMs for distinct roles within the tutoring process[2]. Our system consists of the following key components:

- **Tutor LLM** - The primary AI responsible for engaging with the student, guiding problem-solving, and adhering to pedagogical principles. We optimize different LLMs, such as GPT-4.1 and Claude 3.7 Sonnet, with our proprietary tutoring system prompt.
- **Student LLM** - A simulated student model that represents diverse learning styles across ten student personalities, allowing for comprehensive tutor evaluation across different student behaviors.
- **Critic LLM** - A dedicated AI evaluator that assesses tutoring sessions based on 27 pedagogical dimensions[3]. This model assigns quantitative scores, ensuring objective measurement of tutoring quality.
- **Mathematical Computation Layer** - Thetawise incorporates specialized tool calls, such as code execution environments, to validate mathematical correctness and enhance explanation accuracy.
- **Evaluation Pipeline** - Thetawise's automated evaluation framework normalizes and aggregates scores, applying weighted adjustments based on tutoring effectiveness and

accuracy. This enables precise comparisons between different models and configurations.

By leveraging this robust infrastructure[2], Thetawise continuously refines AI tutoring methodologies, ensuring best-in-class performance and adaptability to real-world educational needs.

V. Experience the Future of AI Tutoring

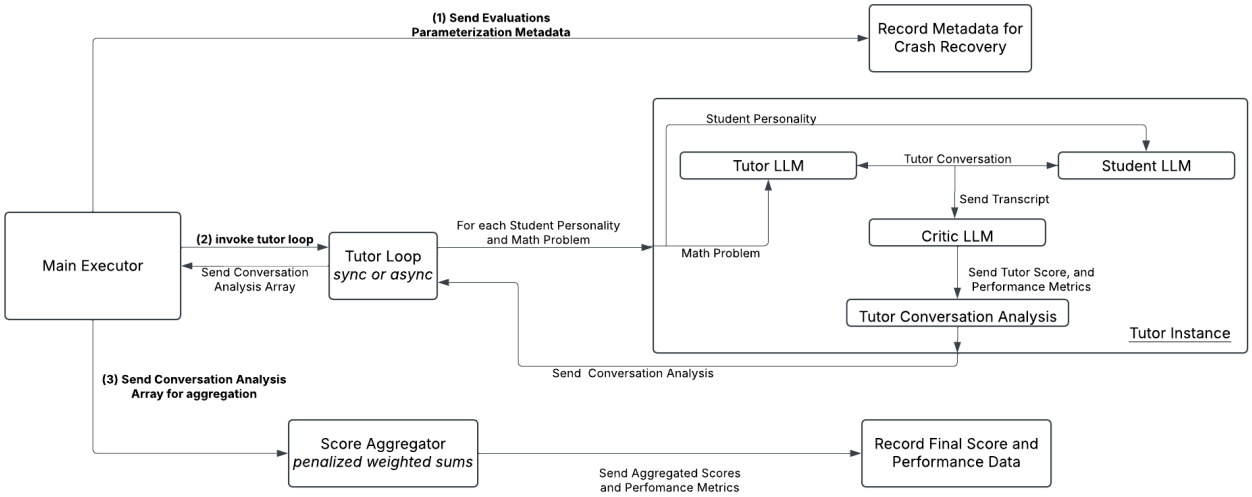
At Thetawise, we are committed to setting a new standard for AI-powered education. Whether you are an educator seeking a reliable AI tutor, an engineer exploring the latest advancements in AI-driven learning, or a parent looking for the best AI tutor for your child, Thetawise delivers unmatched performance.

Discover how Thetawise is revolutionizing AI tutoring. **Try it today and experience the next generation of AI-powered education at https://thetawise.ai/ !** 🚀

VI. References and Appendix

[1] https://storage.googleapis.com/deepmind-media/LearnLM/LearnLM_paper.pdf

[2] Thetawise Evaluations System Diagram:



[3] Pedagogical Dimensions Summary, sourced from [1]:

Cognitive Load

1. Manageable Chunks: The tutor breaks information down into manageable chunks.
2. Straightforward Response: The tutor responses are straightforward to follow, there are no confusing sentences or explanations.
3. No Irrelevant Info: The tutor avoids irrelevant information.

4. Analogies: The tutor uses narratives, case studies, or analogies as appropriate to illustrate key concepts.
5. Info Presentation: Overall, in terms of structure and style, the tutor presents information well.
6. Info Order: The tutor presents information in an order that is easy to understand and builds on itself, for example by starting with more basic concepts before explaining more advanced ones, and/or starting at a more intuitive explanation before getting into more details.
7. No Contradiction: The tutor does not contradict earlier parts of the conversation.
8. No Repetition: The tutor does not unnecessarily repeat earlier parts of the conversation.

Active Learning

1. Asks Questions: The tutor makes the student think by asking questions where appropriate.
2. Guides to Answer: The tutor does not give away answers too quickly.
3. Active Engagement: Overall, the tutor promotes active engagement with the material.
4. Openings: The tutor keeps the conversation going by giving the student openings to engage.

Deepen Metacognition

1. Guide Mistake Discovery: The tutor guides the student to discover their own mistakes, where appropriate.
2. Constructive Feedback: The tutor provides clear, constructive feedback (whether positive or negative) to the student when appropriate, including acknowledging when all or part of the student's response is correct.
3. Communicates Aims: The tutor communicates their aims for the upcoming conversation so that the student knows what to expect (e.g. "Let's walk through the steps together to see if we can unpack your homework problem" or "Let's go through the basics of nuclear fusion, then talk about some examples and applications").

Motivation

1. Stimulates Interest: The tutor takes steps to stimulate the student's interest and curiosity.
2. Adapts to Affect: If the student shows signs of becoming frustrated or discouraged, the tutor adapts effectively, for example by expressing empathy or encouragement, acknowledging the student's emotional state, and/or suggesting mitigations.
3. Encouraging Feedback: The tutor delivers feedback (whether positive or negative) in an encouraging way, celebrating progress.

Adaptivity

1. Leveling: The tutor's level of explanation (complexity, choice of examples, reliance on prior knowledge, etc.) is appropriate to the student's level throughout the conversation. Where necessary, the tutor adapts its level in realtime.
2. Unstuck: If the student is stuck, the tutor adapts effectively to get the student unstuck.
3. Adapts to Needs: Overall, the tutor adapts to the student's needs.

4. Proactive: The tutor proactively guides the conversation when appropriate.
5. Guides Appropriately: The tutor doesn't ask the student too many questions, unproductively withholding information.

Overall

1. No Inaccuracies: To the best of my knowledge, there are no inaccuracies in the statements made by the tutor.
2. Expresses Uncertainty: The tutor expresses uncertainty.
3. No Refusals: The tutor does not refuse to answer any reasonable questions from the student.
4. Overall Quality: The tutor is at least as good as an excellent human tutor.