Proceedings of Machine Learning Research 273:174–190, 2025 iRAISE Workshop at AAAI 2025

MultiTutor: Collaborative LLM Agents for Multimodal Student Support

Edward Sun EDWARDSUN12895@UCLA.EDU
University of California, Los Angeles
LeAnn Tai LEANNTAI@MIT.EDU
Massachusetts Institute of Technology

Abstract

The advent of Large Language Models (LLMs) has revolutionized education, introducing AI tools that enhance teaching and learning. Once purely natural language processors, LLMs have evolved into autonomous agents capable of complex tasks, from software development to high-level trading decisions. However, most educational applications only focus on classroom simulations or single-agent automation, leaving the potential of multi-agent systems for personalized support underexplored. To address this, we propose MultiTutor, a multi-agent tutoring framework tailored to individual student needs. MultiTutor uses internet searches and code generation to produce multimodal outputs like images and animations while expert agents synthesize information to deliver explanatory text, create visualizations, suggest resources, design practice problems, and develop interactive simulations. By identifying knowledge gaps and scaffolding learning, MultiTutor offers a transformative, accessible approach to education. Evaluation against baseline models across metrics like cognitive complexity, readability, depth, and diversity shows MultiTutor consistently outperforms in quality and relevance. Case studies further highlight its potential as an innovative solution for automated tutoring and student support.

Keywords: Large Language Model, Multi-agent LLMs, Personalized Tutor, Education

1. Introduction

One-on-one tutoring is highly effective in improving student learning outcomes Ballestar (2024); Kraft and Falken (2021); Nickow et al. (2020), relying on diverse skills to provide multimodal support, including explanations, diagrams, and tailored examples. While human tutors integrate these seamlessly, computational systems struggle to replicate such complexity, requiring advanced reasoning and multimodal capabilities.

Current approaches using single LLM agents employ prompt engineering, pre-trained knowledge, domain-specific retrieval, and multimodal integration. For instance, LLM tutors have been designed for Deaf and Hard-of-Hearing learners Cheng (2024), while ChatTutor handles tasks like memory recall and quizzing Chen (2024). Similar systems address grading, code explanation, and language practice Chen (2024); Sophia and Jacob (2021); Liu et al. (2022); TOPSAKAL and TOPSAKAL (2022); Tyen (2022). However, they fall short in handling domain-specific content, dynamic resources, and multimodal outputs, limiting their ability to match human tutors' adaptability and depth.

Multi-agent systems present a powerful alternative by dividing tasks among specialized agents to enhance efficiency and collaboration. These systems have demonstrated success in domains like software engineering, where frameworks such as Meta-GPT and AutoGPT

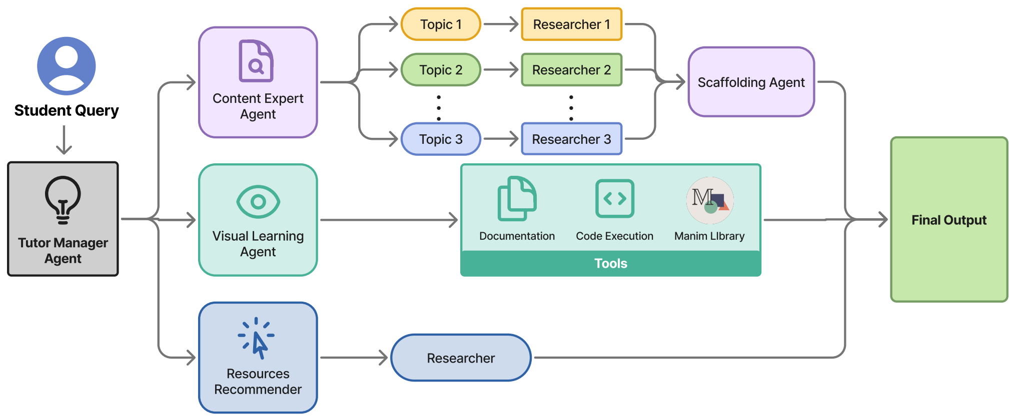


Figure 1: MultiTutor framework: harnessing the collaboration of various expert agents to craft a multi-modal in-depth response to student queries

assign agent roles to streamline task execution Yang et al. (2023); Hong (2024), and finance, where agents analyze large datasets for informed trading decisions Yu (2024); Zhang (2024). Through structured collaboration, these systems outperform single-agent models in reasoning and significantly reduce hallucinations Guo (2024). Building on these strengths, we propose **MultiTutor**, a multi-agent tutoring framework that integrates agents for content retrieval, research, visualization, and resource provision. MultiTutor delivers comprehensive and collaborative learning support, closely mimicking the multimodal and in-depth behavior of human tutors, who often draw diagrams, interact dynamically with students, and provide thorough explanations. Our contributions include the following:

- **Novel Architecture** MultiTutor introduces one of the first multi-agent AI tutoring systems, integrating expert agents for tasks like content retrieval, research, visualization, and code generation to produce detailed, multimodal outputs.

- **Multi-Modal Outputs** Leveraging LLM visualization frameworks, MultiTutor creates diagrams and visualizations for complex STEM topics, addressing the limitations of text-only tutoring systems.

- **Structured Communication Protocol** A novel protocol eliminates the "telephone effect," ensuring coherent and structured outputs through agent collaboration.

- **Extensive Evaluation** We benchmark MultiTutor against single-agent systems and baseline LLMs, demonstrating superior performance through quantitative experiments and qualitative case studies showcasing its real-world applicability.

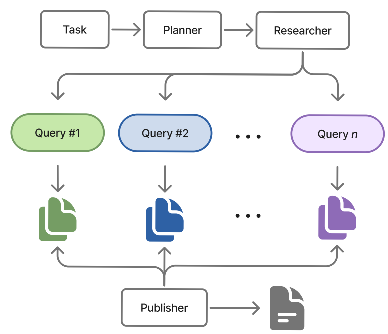


Figure 2: Researcher Agent sub-graph that parallelizes search

2. Methods

2.1. Agent Specialization

Assigning clear roles to LLM agents within a multi-agent workflow enables the efficient division of complex tasks into manageable subtasks Guo (2024); Hong (2024). In MultiTutor, six specialized agents collaborate: Manager, Content Expert, Scaffolding, Visual Learning, Resource Recommender, and Researcher. Each agent operates with distinct roles, tools, and constraints. For instance, the Researcher Agent uses internet search, while the Visualization Agent employs tools like Manim for creating visualizations. The specific roles and capabilities of each agent are as follows:

- **Manager Agent**: Handles student queries by selecting and coordinating downstream agents to ensure effective responses. It utilizes the Content Expert, Visual Learning, and Resource Recommender Agents to structure outputs.

- **Content Expert Agent**: Conducts in-depth research to generate detailed responses with explanations, examples, and practice problems. It coordinates Researcher Agents for parallel query execution.

- **Scaffolding Agent**: Structures the Content Expert Agent's output into a scaffolded, clear format that progresses in difficulty, enhancing clarity and coherence. The scaffolded information is then presented to the student over multiple rounds of interaction along with visualization aids.

- **Visual Learning Agent**: This agent creates high-quality images and animations using Manim, refining outputs through iterative code execution and evaluation The Manim Community Developers (2024). It utilizes a retrieval database containing scraped Manim documentation and basic visualization snippets to craft its visualization code.

- **Resource Recommender Agent**: Searches for relevant resources, such as practice problems and readings, leveraging Researcher Agents to provide supplementary materials.

A hidden **Organizer Agent** compiles the gathered information into a structured response for the student. The **Researcher Agent** functions as a subgraph of agents, distributing web-search tasks among multiple Tavily Search Agents (Figure 2) to optimize search throughput. Each agent's LLM backbone is selected based on task requirements: fast models like `gpt-4o-mini` and `gemini-flash` handle speed-critical tasks, while advanced models such as `o1-preview` support deeper reasoning and multi-round generation et al. (2024b); OpenAI (2024); Zhong (2024); et al. (2024a). For the Visual Learning agent, advanced code-writing models like `claude-3-5-sonnet` are employed to iteratively generate high-quality animations and visualizations Anthropic (2024).

2.2. Communication Protocol

Most multi-agent frameworks rely on natural language message histories for communication, but this approach often leads to issues like exceeding the context window of language models, losing critical details, and introducing a "telephone effect" where repeated messaging distorts information. While strategies like message summarization have been proposed, they often fail to address these limitations, resulting in poor performance for generating long-form, structured, and logically sound outputs.

To overcome these challenges, we developed a structured communication protocol inspired by Meta-GPT, using a global state dictionary that agents query for information and update with their execution results Hong (2024). This ensures that only relevant information is exchanged, avoids unnecessary messages, and maintains task organization. The protocol enables agents to contribute discrete, structured updates to the global state, collectively forming the final output. For instance, the Content Expert Agent queries the Manager Agent for assigned topics, conducts research, and updates the global state with findings, which the Scaffolding Agent later uses to organize the final response. The evaluation is conducted over 25 simulated student questions from 5 diverse subject areas (Math, Biology, Economics, Literature, and Computer Science). We present the specific questions asked in each specific area in Appendix B.

2.3. Experiment Set Up

We evaluate the outputs of our framework against two baselines: vanilla LLMs (`claude-3.5-sonnet` and `gpt-4o`) and their research-augmented versions (single agent), using our researcher framework (Figure 2). Evaluations combine LLM-based assessments for query-specific content quality with rule-based metrics analyzing cognitive distribution, readability, grade levels, and text complexity.

LLM-as-Judge We leverage `gpt-4o` as an expert judge, contextualized with the student query, to assess outputs across four dimensions: accuracy, depth, resourcefulness, and specificity. These metrics align with student needs by ensuring reliable information, fostering critical thinking, providing diverse references, and delivering precise, detailed responses.

Rule-based Metrics Recognizing the limitations of LLMs as judges, such as variability and hallucinations Bai (2023), we average results across 25 runs per query and incorporate rule-based NLP metrics for additional insight. These metrics evaluate multi-factor depth, length, and pedagogical clarity, detailed in Appendix A.

Generated Diagrams As no existing tutoring systems generate diagrams, quantitative evaluation is challenging. Instead, we provide qualitative results in Appendix C, showcasing the complexity and relevance of MultiTutor's diagrams and animations.

2.4. Results

2.4.1. LLM-as-Judge Metrics

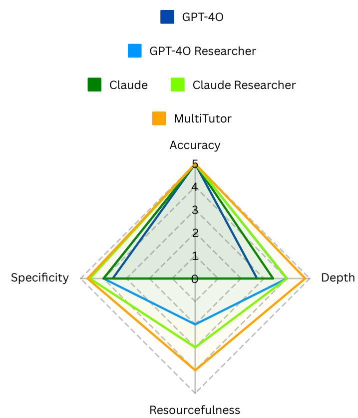


Figure 3: **MultiTutor's** performance on **LLM-based metrics** against baseline models

We present evaluation scores using `gpt-4o` as an impartial judge to assess tutoring responses from vanilla LLMs, research-enhanced LLMs, and MultiTutor across four dimensions: accuracy, specificity, depth, and resourcefulness. Results, shown in Figure 3, are based on strict 1-to-5 ratings normalized against the context of the student question and other outputs. MultiTutor consistently outperforms all baselines, particularly in specificity, depth, and resourcefulness. While accuracy remains similar across models—reflecting modern LLMs' strong baseline intelligence from pre-trained knowledge Srivastava (2023)—MultiTutor 's strengths lie in other dimensions. For specificity, we observe clear improvements from vanilla to research-enhanced models and a further leap with MultiTutor, demonstrating the impact of multi-agent systems in refining precision and detail. Similarly, depth improves as MultiTutor leverages additional researchers and scaffolding mechanisms, enabling richer, more comprehensive responses. The largest gain is in resourcefulness, driven by MultiTutor 's architecture of specialized agents. Unlike single-agent systems, which struggle to balance answering questions and providing references, MultiTutor assigns dedicated agents to each task. This specialization ensures detailed references and robust answers, illustrating the effectiveness of multi-agent systems in enhancing tutoring capabilities.

2.4.2. Rule-Based Metrics

We present metrics derived from traditional NLP rule-based methods, where MultiTutor consistently outperforms baselines across three dimensions: depth, length, and pedagogical clarity. These metrics analyze the structure and content of generated outputs using weighted compound measures.

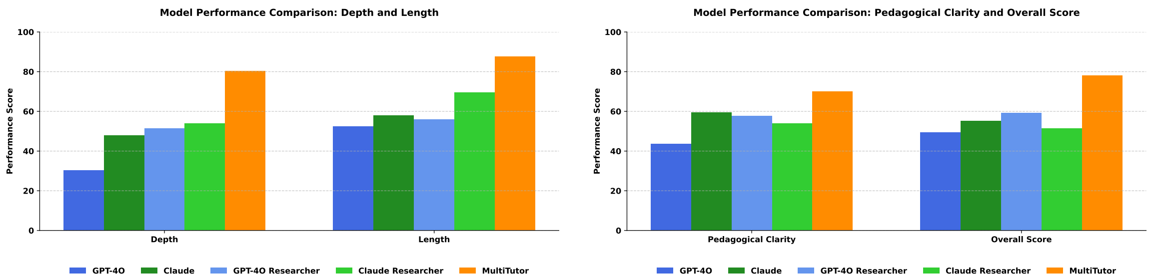


Figure 4: MultiTutor performance comparison against baseline models using **rule-based metrics**. Left: depth and length metrics. Right: pedagogical clarity and overall scores

As shown in Figure 4, depth metrics confirm the LLM-based evaluation, demonstrating that adding more researchers enhances output depth. Single-agent systems with one researcher outperform vanilla LLMs, while MultiTutor's multi-agent framework with specialized researchers achieves the highest depth scores. For output length, MultiTutor excels due to its structured communication protocol, which integrates contributions from individual agents to produce content that exceeds the typical limits of traditional LLMs. In contrast, systems like Claude plateau in length even with research augmentation, whereas MultiTutor continues to scale effectively.

To ensure our metrics do not simply reward longer outputs, we also evaluate pedagogical clarity, which emphasizes clarity, structure, relevance, and detail rather than length alone. Illustrated in Figure 4, pedagogical clarity is assessed using weighted factors such as engaging vocabulary, clear sentence structure, effective examples, and educational strategies like practice problems. MultiTutor significantly outperforms all baselines in this dimension, thanks to its scaffolding agent, which ensures content is well-organized, coherent, and easy to understand, even for extended formats.

3. Conclusion

Our work introduces MultiTutor, a significant step forward in developing multi-agent AI tutoring systems that leverage specialized expert agents for planning, research, visual diagram creation, resource gathering, and pedagogical scaffolding. By mimicking the multimodal and in-depth behavior of human tutors, MultiTutor utilizes multi-agent frameworks to break down complex tasks and produce comprehensive, multimodal outputs through the collaboration of multiple LLM-based agents. This framework provides detailed, tailored support that meets diverse student needs. Comprehensive evaluations demonstrate MultiTutor's superiority over traditional methods, including vanilla LLMs and single-agent augmented systems. Case studies further highlight MultiTutor's ability to automatically generate and refine animations and diagrams through its dedicated visualization agent, significantly enhancing the learning experience. By adapting the multi-agent capabilities to the educational domain, MultiTutor showcases the transformative potential of this approach, representing a major advancement in the emerging field of multi-agent tutoring systems.

Moreover, our work identifies a critical gap in AI tutoring system benchmarks, particularly in assessing their effectiveness in assisting students. We hope this research not only inspires the development of more robust, multimodal, and potentially embodied tutoring agents but also drives the creation of evaluation benchmarks for automated tutoring systems.

References

Anthropic. Claude 3.5 sonnet. https://www.anthropic.com, 2024. Large language model, released June 20, 2024, updated October 22, 2024.

Yushi et al. Bai. Benchmarking foundation models with language-model-as-an-examiner, 2023. URL https://arxiv.org/abs/2306.04181.

María Teresa et al. Ballestar. Effectiveness of tutoring at school: A machine learning evaluation. *Technological Forecasting and Social Change*, 199:123043, February 2024. ISSN 0040-1625. doi: 10.1016/j.techfore.2023.123043. URL http://dx.doi.org/10.1016/j.techfore.2023.123043.

Yulin et al. Chen. Empowering private tutoring by chaining large language models, 2024. URL https://arxiv.org/abs/2309.08112.

Haocong et al. Cheng. Llm-powered ai tutors with personas for d/deaf and hard-of-hearing online learners, 2024. URL https://arxiv.org/abs/2411.09873.

Gemini Team et al. Gemini: A family of highly capable multimodal models, 2024a. URL https://arxiv.org/abs/2312.11805.

OpenAI et al. Gpt-4 technical report, 2024b. URL https://arxiv.org/abs/2303.08774.

Rudolf Flesch and J. Peter Kincaid. Flesch–kincaid readability tests, 1975. Readability formula developed for the U.S. Navy.

Taicheng et al. Guo. Large language model based multi-agents: A survey of progress and challenges, 2024. URL https://arxiv.org/abs/2402.01680.

Sirui et al. Hong. Metagpt: Meta programming for a multi-agent collaborative framework, 2024. URL https://arxiv.org/abs/2308.00352.

Matthew A. Kraft and Grace T. Falken. A blueprint for scaling tutoring and mentoring across public schools. *AERA Open*, 7, January 2021. ISSN 2332-8584. doi: 10.1177/23328584211042858. URL http://dx.doi.org/10.1177/23328584211042858.

Shuqi Liu, SiuYing Man, and Linqi Song. An nlp-empowered virtual course assistant for online teaching and learning. In *2022 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)*, pages 373–380, 2022. doi: 10.1109/TALE54877.2022.00068.

Andre Nickow, Philip Oreopoulos, and Vincent Quan. *The Impressive Effects of Tutoring on PreK-12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence*. July 2020. doi: 10.3386/w27476. URL http://dx.doi.org/10.3386/w27476.

OpenAI. Learning to reason with llms - openai o1 model. https://openai.com/index/learning-to-reason-with-llms/, 2024. Accessed: 2024-11-21.

J.Jinu Sophia and T.Prem Jacob. Edubot-a chatbot for education in covid-19 pandemic and vqabot comparison. In *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)*, pages 1707–1714, 2021. doi: 10.1109/ICESC51422.2021.9532611.

Aarohi et al. Srivastava. Beyond the imitation game: Quantifying and extrapolating the capabilities of language models, 2023. URL https://arxiv.org/abs/2206.04615.

The Manim Community Developers. Manim – mathematical animation framework, April 2024. URL https://www.manim.community/.

Oguzhan TOPSAKAL and Elif TOPSAKAL. Framework for a foreign language teaching software for children utilizing ar, voicebots and chatgpt (large language models). *The Journal of Cognitive Systems*, 7(2):33–38, December 2022. ISSN 2548-0650. doi: 10.52876/jcs.1227392. URL http://dx.doi.org/10.52876/jcs.1227392.

Gladys et al. Tyen. Towards an open-domain chatbot for language practice. In Ekaterina Kochmar, Jill Burstein, Andrea Horbach, Ronja Laarmann-Quante, Nitin Madnani, Anaïs Tack, Victoria Yaneva, Zheng Yuan, and Torsten Zesch, editors, *Proceedings of the 17th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2022)*, pages 234–249, Seattle, Washington, July 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.bea-1.28. URL https://aclanthology.org/2022.bea-1.28.

Hui Yang, Sifu Yue, and Yunzhong He. Auto-gpt for online decision making: Benchmarks and additional opinions, 2023. URL https://arxiv.org/abs/2306.02224.

Yangyang et al. Yu. Fincon: A synthesized llm multi-agent system with conceptual verbal reinforcement for enhanced financial decision making, 2024. URL https://arxiv.org/abs/2407.06567.

Wentao et al. Zhang. A multimodal foundation agent for financial trading: Tool-augmented, diversified, and generalist, 2024. URL https://arxiv.org/abs/2402.18485.

Tianyang et al. Zhong. Evaluation of openai o1: Opportunities and challenges of agi, 2024. URL https://arxiv.org/abs/2409.18486.

Appendix A. Evaluation Details

A.1. Depth Score Calculation

The depth metric (D) is calculated as a weighted combination of five components, each normalized by its respective threshold:

$$
\begin{aligned}
D = w_1 \cdot \min\left(\frac{\text{example\_count}}{T_{\text{examples}}}, 1\right) + \\
w_2 \cdot \min\left(\frac{\text{explanation\_markers}}{T_{\text{markers}}}, 1\right) + \\
w_3 \cdot \min\left(\frac{\text{nested\_explanations}}{T_{\text{nested}}}, 1\right) + \\
w_4 \cdot \min\left(\frac{\text{technical\_terms}}{T_{\text{technical}}}, 1\right) + \\
w_5 \cdot \min\left(\frac{\text{average\_example\_length}}{T_{\text{length}}}, 1\right)
\end{aligned}
\tag{1}
$$

Where:

- T_{examples}: Threshold for example count
- T_{markers}: Threshold for explanation markers
- T_{nested}: Threshold for nested explanations
- $T_{\text{technical}}$: Threshold for technical terms
- T_{length}: Threshold for average example length
- w_1, w_2, w_3, w_4, w_5: Weights for each component, summing to 1

Each Threshold is set dynamically based on the length of the document automatically. To avoid bias for longer documents with more potential trigger words. The final value of D is scaled to be presented as a percentage.

A.2. Length Score Calculation

$$
\begin{aligned}
\text{Length Score} = & \, w_1 \cdot \min\left(\frac{\text{Total Words}}{T_{length}}, 1\right) \\
& + w_2 \cdot \min\left(\text{Content Density}, 1\right) \\
& + w_3 \cdot \min\left(\frac{\text{Average Paragraph Length}}{100}, 1\right)
\end{aligned}
\tag{2}
$$

Where:

- w_1, w_2, w_3: Weights for each component's contribution, summing to 1

- Total Words: Total number of words in the text
- T_{length}: A length threshold that is dynamically adjusted based on the full set of generated responses and normalized relative to their distribution.
- Content Density: $\frac{\text{Number of meaningful words}}{\text{Total Words}}$
- Average Paragraph Length: $\frac{\text{Total words across paragraphs}}{\text{Number of paragraphs}}$

The final Length Score is presented as a percentage.

A.3. Pedagogical Clarity Score Calculation

The pedagogical clarity score is computed by combining three components: engagement score, clarity score, and structure score, each normalized between 0 and 1. The formula is as follows:

$$\begin{aligned}\text{Base Score} = &\frac{w_e \cdot \text{Engagement Score} +}{w_e + w_c + w_s} \\ &\frac{w_c \cdot \text{Clarity Score} +}{w_e + w_c + w_s} \\ &\frac{w_s \cdot \text{Structure Score}}{w_e + w_c + w_s}\end{aligned} \tag{3}$$

A.3.1. Engagement Score

$$\begin{aligned}\text{Engagement Score} = &\, w_{e_1} \cdot \text{Normalized Readability} \\ &+ w_{e_2} \cdot \text{Has Examples} \\ &+ w_{e_3} \cdot \min\left(\frac{\text{Question Count}}{q_{\max}}, 1\right) \\ &+ w_{e_4} \cdot \min\left(\frac{\text{Paragraph Breaks}}{p_{\max}}, 1\right)\end{aligned} \tag{4}$$

A.3.2. Clarity Score

$$\begin{aligned}\text{Clarity Score} = &\, w_{c_1} \cdot (1 - \min(\text{Complex Word Ratio}, 1)) \\ &+ w_{c_2} \cdot \min(\text{Vocab. Diversity}, 1) \\ &+ w_{c_3} \cdot \left(1 - \min\left(\frac{\text{Avg. Sent. Len.}}{s_{\max}}, 1\right)\right)\end{aligned} \tag{5}$$

A.3.3. Structure Score

$$\begin{aligned}\text{Structure Score} = & w_{s_1} \cdot \text{Has Examples} \\ & + w_{s_2} \cdot \text{Has Definitions} \\ & + w_{s_3} \cdot \text{Has Transitions} \\ & + w_{s_4} \cdot \mathbb{1}(\text{Question Count} > 0)\end{aligned}$$

A.3.4. Component Descriptions

- **Weights:**
  - w_e, w_c, w_s: Weights for engagement, clarity, and structure in the base score formula.
  - $w_{e_1}, w_{e_2}, w_{e_3}, w_{e_4}$: Weights for components in the engagement score.
  - $w_{c_1}, w_{c_2}, w_{c_3}$: Weights for components in the clarity score.
  - $w_{s_1}, w_{s_2}, w_{s_3}, w_{s_4}$: Weights for components in the structure score.
- **Parameters:**
  - $q_{\max}$: Maximum question count for normalization.
  - $p_{\max}$: Maximum paragraph breaks for normalization.
  - $s_{\max}$: Maximum average sentence length for normalization.
- **Normalized Readability:** The Flesch Reading Ease score normalized to a 0-1 range Flesch and Kincaid (1975).
- **Has Examples, Has Definitions, Has Transitions:** Binary values (0 or 1) indicating presence.
- **Complex Word Ratio:** Ratio of complex words to total words.
- **Vocabulary Diversity:** Unique non-stopword tokens divided by total tokens.
- **Average Sentence Length:** Total words divided by sentence count.

A.3.5. Flesch Reading Score Normalized to [0, 1]

The Flesch Reading Ease (FRE) formula[1] is computed as:

$$\begin{aligned}FRE = 206.835 - 1.015 & \left(\frac{\text{Total Words}}{\text{Total Sentences}}\right) \\ -84.6 & \left(\frac{\text{Total Syllables}}{\text{Total Words}}\right)\end{aligned} \tag{6}$$

where:

- Total Words is the number of words in the text.

1. https://en.wikipedia.org/wiki/Flesch-Kincaid_readability_tests

- Total Sentences is the number of sentences in the text.
- Total Syllables is the number of syllables in the text.

To normalize the FRE score to the range $[0, 1]$, we use the following formula:

$$FRE_{norm} = \frac{\text{FRE} - \text{Min FRE}}{\text{Max FRE} - \text{Min FRE}} \tag{7}$$

Appendix B. Evaluation Student Questions

MultiTutor is designed to adapt to a wide range of subjects and academic questions. For evaluation, we calculate the average score by querying MultiTutor on five subjects, each consisting of five questions, resulting in a total of 25 evaluations.

B.1. Math

- How can the Fundamental Theorem of Calculus be used to evaluate definite integrals, and how does it connect differentiation and integration?
- What is the significance of eigenvalues and eigenvectors in understanding the properties of a matrix, and how are they applied in real-world problems?
- Can you explain the structure of a group under addition modulo n and how it satisfies the group axioms?
- How is Bayes' Theorem used to update the probability of an event based on new evidence, and what are its practical applications?
- What is the principle of mathematical induction, and how can it be applied to prove statements about sequences or summations?

B.2. Biology

- How do differences in the structure and function of rough and smooth endoplasmic reticulum contribute to their roles in the cell?
- Can you explain how epigenetic modifications, such as DNA methylation and histone acetylation, regulate gene expression in eukaryotes?
- How do keystone species influence the stability and biodiversity of an ecosystem?
- What role do telomeres and telomerase play in cellular aging and cancer?
- How does the process of osmoregulation differ between freshwater and marine fish, and what adaptations support these differences?

B.3. Economics

- How do price ceilings and price floors impact market equilibrium, and what are the potential consequences for consumer and producer surplus?
- Can you explain how changes in interest rates influence aggregate demand and supply in the context of monetary policy?
- What are the advantages and disadvantages of free trade agreements, and how do they impact domestic industries and labor markets?
- How do cognitive biases like loss aversion and anchoring affect consumer decision-making in markets?
- What role does human capital investment play in economic growth, and how can governments encourage such investments in developing countries?

B.4. Literature

- How does the use of symbolism in The Great Gatsby reflect the themes of ambition and disillusionment in the American Dream?
- What are the similarities and differences in how identity is explored in James Baldwin's Giovanni's Room and Virginia Woolf's Orlando?
- How did the cultural and political atmosphere of the Harlem Renaissance shape the works of poets like Langston Hughes and Claude McKay?
- In what ways does stream-of-consciousness narration in Virginia Woolf's Mrs. Dalloway enhance the reader's understanding of the characters' inner lives?
- How do Gothic elements in Mary Shelley's Frankenstein challenge Enlightenment ideals and reflect anxieties about scientific progress?

B.5. Computer Science

- What are the key differences between binary search trees, AVL trees, and red-black trees in terms of balance and efficiency?
- What is backpropagation and how does the detailed math work?
- How does breadth-first search work and how does it differ from depth-first search?
- What is paging in operating systems?
- How does the Transmission Control Protocol (TCP) ensure reliable data transfer over an unreliable network, and what are the limitations of this approach?

Appendix C. Additional Case Studies

We illustrate a typical output format of MultiTutor in Figure 5. As shown, MultiTutor leverages its research capabilities to produce detailed language outputs while also utilizing Manim to generate animations and images, simulating the multimodal diagram-drawing abilities of human tutors.

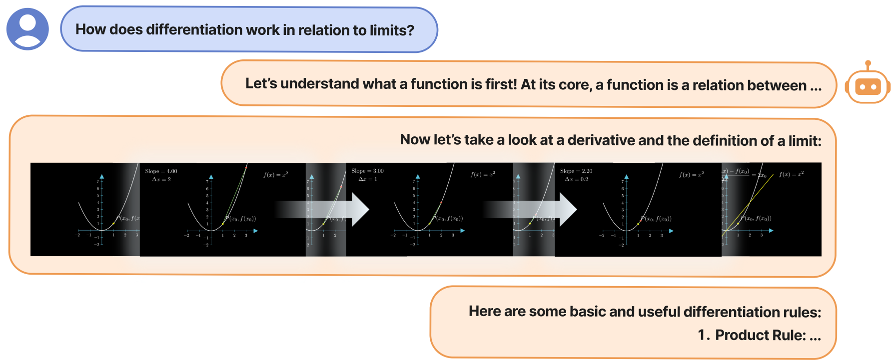


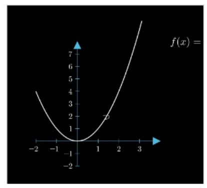
Figure 5: An example of a multi-modal output from MultiTutor showcases animations generated using Manim, combined with insights from our multi-agent framework to deliver comprehensive and informative results.

Here, we present additional case studies of MultiTutor's responses to student questions. For clarity and conciseness, the outputs have been truncated to fit a graphical format. These samples showcase MultiTutor's capabilities but do not represent the full extent of its outputs. Truncations are indicated with "...".

To understand differentiation and limits, you must first have a good grasp of how functions work:

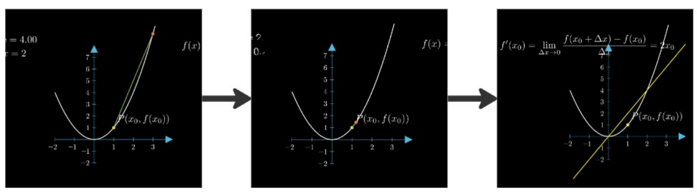
Understanding Functions

At its core, a function is a relation between a set of inputs and a set of permissible outputs, where each input is associated with exactly one output. Mathematically, a function \(f \) from a set \(X \) to a set \(Y \) is denoted as \(f: X \rightarrow Y \)...



Concept of a Derivative

The derivative of a function at a specific point measures how the function's output changes as its input changes. Formally, if we have a function f(x), the derivative at a point x = a, denoted as f'(a), is defined as the limit of the difference quotient as h approaches zero: f'(a) = lim(h→0) [(f(a + h) - f(a))/h]...



Limit Definition of a Derivative

Mathematically, the derivative of a function at a specific point is the slope of the tangent line to the function at that point. The derivative of a function \(f(x) \) at a point \(x = a \) is defined by the expression: \(f'(a) = \lim_{{h \to 0}} \frac{{f(a + h) - f(a)}}{h} \). ...

Mathematics: Basic Differentiation Rules

The process of differentiation is essentially about finding the derivative, denoted as \(f'(x) \) or \(\frac{dy}{dx} \), for a function \(y = f(x) \). The derivative represents the slope of the tangent line to the function's graph at any given point, offering a precise measure of how the function changes at that point. This understanding is crucial for analyzing and predicting the behavior of functions in real-world applications.

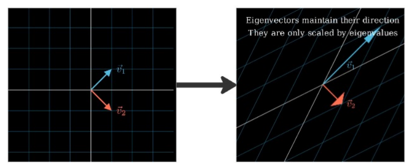
Here are some additional learning resources to better understand derivatives:

- Introductory text to derivatives: https://math.libretexts.org/Courses/Mount_Royal_University/Calculus_for_Scientists_I/3%3A_Derivatives/2.2%3A_Techniques_of_differentiation\
- Khan Academy on derivatives: https://www.khanacademy.org/math/ap-calculus-ab/ab-differentiation-2-new/ab-3-5a/a/review-categorizing-functions-for-taking-derivatives
- Calculator of derivatives: https://www.derivative-calculator.net/
- Interactive demonstration of limits and derivatives: https://www.intmath.com/differentiation/1-limits-and-differentiation.php

Figure 6: MultiTutor output on the question **"How does differentiation work in relation to limits?"**

Eigenvalues and Eigenvectors Definition

At the core, an eigenvector of a square matrix is a non-zero vector that, when transformed by the matrix, results in a vector that is a scalar multiple of the original. This scalar is known as the eigenvalue. Mathematically, this relationship is expressed as A*v = λ*v, where A is the matrix, v is the eigenvector, and λ is the eigenvalue. This equation indicates that the transformation by A does not change the direction of v, only its magnitude, which is scaled by λ.



To determine the eigenvalues of a matrix, one must solve the characteristic equation. This equation is derived from the determinant of the matrix A - λI, where I is the identity matrix of the same dimension as A. The characteristic equation is given by det(A - λI) = 0...

Eigenvalues and eigenvectors possess several important properties. For eigenvalues, the sum of all eigenvalues of a square matrix equals the trace of the matrix, and their product equals the determinant. If the matrix is real and symmetric...

Characteristic Equation

The characteristic equation is a pivotal concept in linear algebra and differential equations, serving as a key tool for understanding the properties of matrices and systems of linear equations. It is a polynomial equation derived from a square matrix, which reveals the matrix's eigenvalues...

Diagonalization

Diagonalization is a fundamental concept in mathematics, particularly in linear algebra, where it serves as a tool to simplify complex matrix operations...

Applications of Eigenvalues:

Dimensionality Reduction: Dimensionality reduction is a fundamental concept in mathematics and data science, essential for simplifying complex datasets while preserving their core characteristics...

Vibration Analysis: Vibration analysis is a crucial component in the field of engineering, focusing on the study of oscillatory systems...

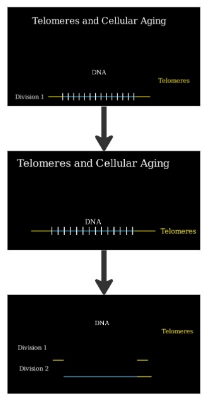
Google PageRank: The PageRank algorithm assigns a numerical weight to each page within a hyperlinked set of documents, such as the World Wide Web. The core idea is that more important websites are likely to receive more links from other sites...

Here are some additional resources to guide your learning:
- MATLAB Reading: https://www.mathworks.com/help/matlab/math/eigenvalues.html
- Georgia Tech Notes: https://textbooks.math.gatech.edu/ila/diagonalization.html
- GeeksforGeeks on Eigenvalues: https://www.geeksforgeeks.org/applications-of-eigenvalues-and-eigenvectors/

Figure 7: MultiTutor output on the question **"What is the significance of eigenvalues and eigenvectors in understanding the properties of a matrix, and how are they applied in real-world problems?"**

Telomeres: Structure and Function

Telomeres are essential structures located at the ends of linear chromosomes, composed of repetitive nucleotide sequences. These sequences, rich in guanine bases, are crucial for maintaining genomic stability and integrity. In humans, the telomeric sequence is TTAGGG, repeated thousands of times. This repetitive nature is vital for the protective role telomeres play in cellular processes. The telomeric DNA is associated with a specialized set of proteins known as the shelterin complex, which includes TRF1, TRF2, POT1, TIN2, TPP1, and RAP1. This complex is essential for protecting telomeres from being mistakenly identified as DNA damage and for regulating their length...



The End Replication Problem

The end replication problem is a critical issue in molecular biology, primarily affecting eukaryotic cells with linear chromosomes. This problem arises during DNA replication..

Telomerase: Structure and Function

The structure of telomerase is complex, comprising both protein and RNA components. The core components include the telomerase reverse transcriptase (TERT) and the telomerase RNA component (TERC)...

Cellular Aging and Telomeres

Telomeres are composed of repetitive DNA sequences, rich in guanine, and are associated with a specialized set of proteins known as the shelterin complex. In humans, the telomeric sequence is TTAGGG, repeated thousands of times. The primary role of telomeres is to protect chromosome ends from being mistaken for DNA damage..

Cancer and Telomerase Activation

Cancer is a complex and multifaceted disease characterized by the uncontrolled growth and division of cells. One of the key factors contributing to the immortality of cancer cells is the activation of telomerase...

References:
- NIH Article: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/telomerase
- Khan Academy on Telomerase: https://www.khanacademy.org/science/biology/dna-as-the-genetic-material/dna-replication/a/telomeres-telomerase
- Uniprot Entry: https://www.uniprot.org/uniprotkb/O14746/entry

Figure 8: MultiTutor output on the question **"What role do telomeres and telomerase play in cellular aging and cancer?"**