



Generative AI-Based Tutoring for Enhancing Learning Engagement and Achievement

TIAN BELAWATI 

DIMAS PRASETYO 

**Author affiliations can be found in the back matter of this article*

RESEARCH ARTICLE



ABSTRACT

This paper presents the findings of a pilot study on the use of generative AI (GAI) in tutorial sessions within a large-scale distance education institution in Indonesia. The primary aim of the experiment was to assess the impact of GAI-based tutoring on student engagement and academic achievement. A secondary objective was to explore how GAI could reduce the workload of human tutors by automating routine tasks, such as responding to frequently asked questions and providing initial feedback, thereby allowing human tutors to concentrate more on moderation and higher-order teaching tasks. The development team adopted a modular design approach, utilizing OpenAI's ChatGPT models (versions 3.5 and 4.0) and integrating them with the Moodle learning management system. The resulting system consisted of three core modules: Management, LMS Integration, and Backend. These modules enabled efficient administration, seamless integration with the existing learning platform, and effective generation of AI responses for student interactions. The study involved data collection from four online courses with a total of 37,743 students. The results indicated that students in GAI-assisted classes participated more actively in discussion forums and achieved slightly higher scores in their assignments compared to those in non-GAI-assisted settings. These findings suggest that GAI-based tutor assistants can provide considerable enhancements in online learning environments, particularly in fostering engagement and improving student outcomes. The study suggests further improvement to make it more effective and have a greater positive impact on student engagement and achievement.

CORRESPONDING AUTHOR:

Tian Belawati

Universitas Terbuka, Indonesia

tbelawati@ecampus.ut.ac.id

KEYWORDS:

generative AI tutoring; AI instructions; learning engagement; learning achievement; AI-based learning; online tutorials; distance learning; distance education; AI for tutorials; generative AI in education

TO CITE THIS ARTICLE:

Belawati, T., & Prasetyo, D. (2025). Generative AI-Based Tutoring for Enhancing Learning Engagement and Achievement. *Open Praxis*, 17(2), pp. 211–226. DOI: <https://doi.org/10.55982/openpraxis.17.2.902>

The past several years have seen tremendous changes in the educational scene, mostly as a result of technological improvements. Artificial intelligence (AI) is a fascinating subject that is attracting the interest of stakeholders and educators alike. Of all the AI types, generative AI, or GAI, is particularly noteworthy because of its capacity to create new content based on certain prompt. The new and unique content can be either text, images, videos, music, or speech. This process involves three key components: (1) the model: this is the AI system itself, (2) user input (prompt): the questions or commands provided by users, and (3) iterative refinement: the AI continuously improves its output until it meets the user's expectations (Wood, 2023). These advanced neural network models utilize vast amounts of data and billions of learning parameters, enabling them to produce credible and high-quality content. The widespread availability, low costs, and user-friendly interfaces of GAI tools have made them increasingly popular in educational settings.

The journey of AI in education began as early as the 1960s, with the development of intelligent tutoring systems designed to personalize learning for individual students (Ross, 2008). This evolution continued into the early 2000s when computer scientists began creating adaptive learning systems and chatbots that focused on natural language processing and machine translation (Guo, Wang, Gu et al., 2021; Seehan, 2023; Teachflow, 2023). Fast forward to the 2020s, and we are witnessing rapid advancements in GAI, particularly with initiatives from organizations like OpenAI. For instance, ChatGPT—a popular GAI tool for content creation—gained over one million users within just a week of its launch in November 2022 (Mahajan, 2024). This success has sparked global interest in educational applications of GAI.

Today, GAI tools like ChatGPT are commonplace among students, teachers, and administrative staff. Students use these tools for research and assignments, while educators leverage them to streamline lesson planning and material development. Administrative staff find GAI helpful for tasks like report writing and data analysis. It's essential for educational institutions to embrace GAI not just as a passing trend but as a valuable resource for enhancing teaching and learning.

Many researches have investigated the impact of GAI in education. Results of previous studies indicate that GAI tools can facilitate the provision of immediate and tailored feedback thus streamlining the feedback process, which not only promotes student engagement but also enhances learning outcomes (Nguyen et al., 2024; Weeks et al., 2024). Similarly, Coursera's demonstrated that AI-powered learning assistants can effectively guide students through complex concepts, thereby reducing feelings of isolation and enhancing motivation (Coursera, 2023).

Recognizing the potential of GAI, Indonesia Open University or Universitas Terbuka (UT) has initiated an experiment to evaluate the effectiveness of GAI-guided tutoring compared to traditional human tutoring. As mentioned earlier, this experiment is particularly relevant for distance education institutions, which face challenges in providing meaningful learning experiences.

Founded in 1984, UT was established as an open university to enhance Indonesia's higher education capacity through distance learning. With over 550,000 students enrolled in around 1,000 courses across various levels, UT primarily relies on text-based learning materials, supplemented by audio, video, and computer-based resources. The university has adapted its tutorial services from face-to-face meetings to predominantly online interactions. This version aims to make the information more relatable and engaging while maintaining the core ideas and details from the original text.

One of the learning supports provided to students is online tutorials, which are delivered using Moodle-based Learning Management System. These tutorials are designed to be asynchronous, allowing students to access them at their convenience. Each semester, UT sets up online tutorials for all its courses, creating over 50,000 virtual tutorial classes. For each course, UT develops a master tutorial class that is then duplicated into multiple parallel classes, depending on the number of registered students. Some courses may have more than 1,000 parallel classes, resulting in a total of over 2.3 million student-course interactions. With this large number of activities, UT can definitely benefit from the use of GAI as assistant tutors.

Based on the above background, this study aims to examine the effects of Generative AI (GAI)-based tutoring on student engagement and achievement in a large-scale distance education setting. The main research question guiding this study is: How does the use of GAI-based tutor assistants impact student engagement and academic achievement compared to traditional human tutors in online tutorial classes at Indonesia Open University?

Numerous studies have explored the impact of GAI in education, focusing on areas like student engagement and academic performance. For example, Chan & Lee (2023) emphasized how AI can customize and personalize curricula, ultimately leading to improved educational quality. Their research indicates that when students are provided with tailored learning experiences, they are more likely to engage deeply with the material. Similarly, Nguyen et al. (2024) emphasized GAI's ability to enrich student engagement and learning experiences, suggesting that AI tools can create interactive environments that foster active participation. White (2024) further supports this notion by arguing that AI can facilitate personalized, research-based learning experiences that align with each student's unique strengths and interests. This tailored approach not only enhances engagement but also allows students to take ownership of their learning journey.

The potential of GAI to promote student-centered initiatives is further explored by Sevnarayan and Potter (2024), who examined its role in distance learning. Their findings indicate that GAI can effectively engage students by providing them with resources and support that cater to their individual needs. This is particularly important in online education, where maintaining student engagement can be challenging. Additionally, a study by Xu (2024) confirms that the integration of AI tools has resulted in positive outcomes in terms of academic performance, motivation, and overall engagement. These studies collectively underscore the importance of leveraging GAI to create engaging and supportive learning environments that can significantly enhance student experiences.

However, the benefits of GAI are not without their challenges. A review by Ogunleye et al. (2023) examined GAI's role in higher education and highlighted a critical concern: while AI tools can personalize learning experiences and improve student outcomes, there is evidence suggesting that students who rely heavily on these tools may score lower academically compared to their peers who do not use them. This paradox raises questions about the effectiveness of GAI in promoting genuine learning versus mere convenience. The study emphasizes the need for clear guidelines and best practices for the effective use of GAI in educational settings, advocating for a balanced approach that combines academic rigor with student engagement and empowerment.

Further research by Weeks et al. (2024) reveals that the negative impact of GAI usage is particularly pronounced among high-potential learners, indicating that excessive reliance on these tools might hinder their academic growth. This finding suggests that while GAI can enhance learning, it may also inadvertently stifle critical thinking and problem-solving skills if not used judiciously. The implications of these findings are significant for educators, institutions, and policymakers, highlighting the necessity for thoughtful integration of AI tools. By developing comprehensive guidelines and policies for GAI use in higher education, stakeholders can ensure that the benefits of these technologies are maximized while minimizing potential drawbacks, ultimately fostering a more effective and engaging learning environment for all students.

The potential of using GAI in distance educational settings is still very intriguing. With appropriate usage, GAI can provide immediate feedback to distance learners, which is crucial for maintaining engagement and motivation. Sodiq and Pearson (2024) argue that prompt feedback helps learners stay on track and address any misconceptions quickly, enhancing their learning experience. GAI can also generate personalized feedback tailored to each learner's needs. This feedback is not only prompt but also relevant to the individual's learning journey, making it more effective in addressing specific areas of improvement. Additionally, GAI tools can provide consistent feedback that aligns with established rubrics and assessment criteria. This consistency ensures that feedback is fair and reliable, which is essential for maintaining academic integrity and promoting equitable learning outcomes.

By providing prompt and personalized feedback, GAI can enhance the overall learning experience for distance learners. This includes better understanding of concepts, improved writing skills, and enhanced research capabilities, all of which are critical for academic success (Chan & Hu, 2023). Most importantly, the use of GAI for feedback can significantly reduce the time spent on marking and feedback (Sodiq & Pearson, 2024). This efficiency is particularly valuable in distance education settings where the number of students is high and resources are often limited. Overall, the promptness and personalization of GAI feedback make it an invaluable tool for enhancing the learning experience of distance learners, ensuring that they receive timely and relevant support throughout their educational journey.

Research supports these claims, highlighting the transformative impact of GAI in distance education. For instance, studies have shown that GAI tools can provide immediate and tailored feedback, which not only promotes student engagement but also enhances learning outcomes (Nguyen et al., 2024). Furthermore, the integration of GAI in assessment practices has been found to streamline the feedback process, allowing educators to focus more on instructional design and student support rather than administrative tasks (Wecks et al., 2024). This aligns with findings from Coursera's initiative, which demonstrated that AI-powered learning assistants can effectively guide students through complex concepts, thereby reducing feelings of isolation and enhancing motivation (Coursera, 2023).

METHODS

RESEARCH DESIGN

This study adopts a quasi-experimental design comparing two groups: students in tutorial classes with GAI-based tutors (experimental group) and those with human tutors (control group).

DATA COLLECTION

Data for this study were obtained from four general education courses offered during the second semester of the 2023/2024 academic year at UT. Among the participating classes, 500 tutorial classes used GAI-based tutor assistants, while another 500 tutorial classes served as the control group with human tutors. In total, 37,743 student records were analyzed, focusing on two main indicators: engagement and academic achievement. Student engagement was measured by the number of discussion forum posts, as recorded through the Learning Management System's (LMS) analytics. Academic achievement was assessed based on tutorial assignment scores retrieved from the university's Student Record System. Before analysis, the dataset was cleaned by removing incomplete records or entries with missing data.

DATA ANALYSIS

To determine whether there were statistically significant differences between the experimental and control groups, a series of normality tests (Kolmogorov-Smirnov) were conducted. As the data were not normally distributed, non-parametric statistical analysis using the Mann-Whitney U test was performed to compare engagement and achievement levels between the two groups. In addition, Cliff's Delta was calculated to estimate the effect size and assess the practical significance of any observed differences.

ABOUT THE EXPERIMENT

Recognizing the potential of GAI, UT has initiated an experiment to evaluate the effectiveness of GAI-guided tutoring compared to traditional human tutoring. As mentioned earlier, this experiment is particularly relevant for distance education institutions, which face challenges in providing learning supports and meaningful learning experiences.

Founded in 1984, UT was established as an open university to enhance Indonesia's higher education capacity through distance learning. With over 550,000 students enrolled in around 1,000 courses across various levels, UT primarily relies on text-based learning materials, supplemented by audio, video, and computer-based resources. The university has adapted its tutorial services from face-to-face meetings to predominantly online interactions. This version aims to make the information more relatable and engaging while maintaining the core ideas and details from the original text.

One of the learning supports provided to students is online tutorials, which are delivered using Moodle-based Learning Management System. These tutorials are designed to be asynchronous, allowing students to access them at their convenience. Each semester, UT sets up online tutorials for all its courses, creating over 50,000 virtual tutorial classes. For each course, UT develops a master tutorial class that is then duplicated into multiple parallel classes, depending on the number of registered students. Some courses may have more than 1,000 parallel classes, resulting in a total of over 2.3 million student-course interactions. With this large number of activities, UT can definitely benefit from the use of GAI as assistant tutors.

The development of GAI-based tutor assistants at UT is based on the need to provide better teaching services to students by improving learning experiences through increased interactive discussions facilitated by the tutor's faster response to student activities during tutoring sessions. This is supposed to enhance students' learning satisfaction. This is consistent with the findings of previous research, which show that increasing engagement in various online activities might help preserve student retention and motivation (Chan & Hu, 2023; Nguyen et al., 2024; Tierney, Peasey, & Gould, 2024; Zhou, Zhang, & Chan, 2024). Furthermore, this initiative aims to reduce the workload of tutors who have manage online classes through automation. Meanwhile, the human tutors' roles will be redefined as moderators.

In the early stages, the development of GAI-based tutor assistants at UT was designed for general courses that have a very large number of students (>1000) every semester. Therefore, five general courses were selected, namely: Civic Education, Indonesian Language, Basic Cultural Sciences, Islamic Education, and Management of Information Technology Service (Pendidikan Kewarganegaraan, Bahasa Indonesia, Ilmu Budaya Dasar, Pendidikan Agama Islam, dan Manajemen Layanan Teknologi Informasi). In the first semester or first stage, the number of classes used for the trial implementation of GAI-based tutor assistants is one class for each course with a total of 500 students; and in the second semester (second stage) it was increased to 100 classes/courses with a total of 500 classroom participated by approximately of 100.000 students (per class is set for a maximum of 200 students). Each stage has a different main goal. The focus of the first stage of development is to find out whether the system architecture, the integration of GAI-based tutor assistants can be done and how the most effective and efficient integration model is seen from the level of complexity of the development process, development and operation costs, and ease of maintenance as well as future development. Meanwhile, in the second stage, the development focuses on determining the level of system stability when facing an increasingly large utilization burden, including improvements in management system by adding features for fine tuning in the prompting process of the LLM model adjusted to instructional objectives, criteria and assessment rubrics on each topic discussed in the courses offered.

From the system perspective, the development team uses a modular and independent approach with the aim that this GAI-based tutor assistant system can use a variety of LLM models and can be integrated into any *Learning Management System*. Especially for current development needs, the team agreed to use an LLM model based on OpenAI ChatGPT 3.5 and 4.0 and LMS Moodle (which is an LMS that has been used at thUT). In these initial stages, this tutor assistant model would use global knowledge from ChatGPT training results. There are at least three major modules developed to support the GAI-based tutor assistant system, namely:

1. **Management Module:** This module was developed as the core of the tutor assistant system. This module is used as a classroom management administration system to activate GAI-based tutor assistants in certain course classes and make settings based on predetermined parameters. In this module, material developers can compile criteria and rubrics that will later be used as input prompting. Based on this prompting input, ChatGpt will respond to every response to discussions and assignments submitted by students. The response given will be in the form of a response or feedback to student answers and a score/rate that shows the quality of student answers based on the rubric that has been developed.
2. **LMS Integration Module.** This module is developed using the API that has been provided in the LMS so that the integration process can be carried out seamlessly and non-intrusively. This is done so that during each LMS update process according to the Moodle version, Developers do not need to re-engineer or redevelop the tutor assistant module. This module searches through every discussion and student assignment, stores all of the data in the management module, and then executes an interactive process by transferring the data to the backend module. The Backend module will provide feedback response data based on the answer input and prompting provided. This feedback data will be in the form of answers or feedback from ChatGPT and values that are then sent back to the integration module and displayed as a response to a discussion or feedback from assignment answers.
3. **Backend Module:** This module is developed as a service provider (web service) that connects the previous module with generative AI models. This module is in charge of receiving data in the form of student responses from the integration module combined with prompting models from the management module and then sent to the ChatGPT API to get a feedback response. This feedback response is then sent to the previous two modules.

This GAI-based tutor assistants model was tasked in assisting the human tutors in learning activities in UT tutor, and for the time being was only in the discussions and assignments, with the following criteria:

1. Classroom developers provide rubrics and criteria to guide the LLM model's reaction, feedback, and grading process. The more detailed and comprehensive the defined rubrics, the better and more accurate the feedback response will be. The LLM model will employ the global knowledgebase, which is updated on a regular basis by the model provider, as its foundation.
2. The discussion activity is still in the form of a question and answer model where the initiation of the topic of discussion starts from the tutor's question in each discussion. Students will answer the discussion and the LLM model will carry out an evaluation based on the discussion questions and rubrics developed. The LLM model will then provide a response based on these two parameters. The process of further discussion interactivity is limited to ensure that the response from the LLM mode does not go beyond the topic of discussion (out of context).
3. The interactivity and response speed of the tutor assistant are limited to simulate the discussion situation as naturally as possible. This is done by slowing down the answer response from the tutor assistant (by 10 minutes) in every answer or response sent by the student.
4. In assignments, the format is still limited to text-based student assignment responses, whether it is answers given online or in the form of files (docx and pdf formats). This is done because developers are still focusing on improving the prompting technique for a feedback response that is more focused on the topic of discussion. Further development such as handwriting recognition, mathematical character recognition and image recognition will be carried out at the stage of further development according to the needs of

The following serves as an example of an introduction to discussion (trigger questions) and the evaluation criteria/rubric for the Civic Education responses.

Trigger Question from the Human Tutor

As we all know, civic education is a compulsory subject for every Indonesian citizen through formal education, ranging from primary, secondary, and higher education. However, in some cases, citizens' awareness of their rights and obligations is still not in line with expectations. For example: (1) there are still illegal behaviors committed by citizens, (2) violations of traffic regulations are still frequent, (3) corrupt behavior is still ongoing, (4) terrorism crimes are also still present in our country, etc.

1. *In your opinion, why does this happen?*
2. *As a college student, what can you do? Associate it with the major or study program you are taking now!*

Have a good discussion!

Criteria/Rubric for GAI-based Tutor Assistant

The following is an assessment rubric to evaluate answers in discussion forums regarding the importance of Civic Education for Indonesia citizens, focusing on questions about the reasons for unlawful behavior and the role of students from their respective departments/ study programs in overcoming these problems:

1. Problem Introduction (10 points)
 - The answer begins with a brief introduction to the importance of Civic Education. (5 points)
 - Stating some examples of unlawful behavior that often occurs, such as traffic violations, corruption, and terrorism. (5 points)

2. Cause Analysis (40 points):

- Explaining the reason why there are still many citizens who are not aware of their rights and obligations. (and/or)
- Associating low legal awareness with the poor quality of Civic Education delivery method. (and/or)
- Providing concrete examples that are relevant to social or cultural conditions in Indonesia.

3. Role of Students and Departments/Study Programs (40 points)

- Explaining the role of students in general in raising legal and civic awareness. (and/or)
- Connecting the role with the major/study program taken, providing examples of activities or initiatives that can be done. (and/or)
- Conveying innovative ideas or creative solutions that can be applied to overcome these problems from the perspective of the department/study program.

4. Conclusion and Advice (10 points)

- Presenting conclusions that reflect a deep understanding of the importance of Citizenship Education. (and/or)
- Providing constructive suggestions for policy makers, educational institutions, or the wider community in improving the effectiveness of Civic Education.

The correct answer is the one that is able to comprehensively answer the question by including an analysis of the causes, the role of students from various departments/study programs in overcoming problems, and providing constructive suggestions to increase citizens' awareness of their rights and obligations.

Score: 0 to 94

Instructions:

- Evaluate each criterion step by step, give a cumulative score, the maximum total must be 94.
- If the answer does not answer the question directly, give it a score of 10.

The examples below show how the GAI-based tutor assistant responded to two distinct student comments, using the discussion trigger and rubric mentioned above as a guide.

Response of Student 1

The presence of law in society is an important means to shape social welfare. However, why did the violation occur? Examples of violations of this law are diverse, for example violations of traffic regulations, corrupt behavior, and terrorism crimes. The economy is predicted to be the main cause of rampant violations of the law. However, poverty is not the main reason, because the perpetrators of violations of the law are not only from the poor, the upper economic class is also rampant in violations. Generally, perpetrators of law violations commit violations according to their social reach. In other words, those in the upper class are capable of committing major violations of the law, such as corruption, money laundering, bribery, and others.

There are several factors that cause unlawful behavior, including:

1. Weak law enforcement.

One of the causes of weak law enforcement in Indonesia is the quality of law enforcers. The low morality results in a lack of professionalism and unwillingness in law enforcement. This morality is also related to corruption committed by law enforcement personnel (judicial corruption). Law enforcers who are supposed to enforce the law are actually involved in corrupt practices.

2. Violations are considered commonplace.

The second reason why violations of the law occur is because the public considers the violation as normal or not something wrong. For example, taking the law into one's own hand. However, people often consider this method as a step to defend the "victim". The vigilante itself is also considered the right action in fulfilling the required justice.

3. Low legal compliance in society.

The reason for the lack of awareness in the community is that community protection is still not reflected in the rules and regulations that have been enacted by the authorised institutions. The general public also believes that law enforcement officers, who are the ones who create and administer laws, are still unable to effectively carry out the regulations that have been established. In actuality, law enforcement officers who are meant to be the executors frequently break the law themselves.

4. Lack of moral and character education.

Moral and character education is very related to shaping an ethical, intelligent, and decent person. However, in Indonesia there is still a crisis of moral and character education. This is because of the very rapid development, causing moral and character education that has been developed since a long time ago, simply eroded by globalization.

5. Environmental factors.

A less conducive environment, such as full of violence, poverty, and injustice, can affect the behavior of a citizen.

6. Media factors.

Social media and mass media can affect the perspective and behavior of citizens.

7. Educational factors.

Inadequate and unbalanced education can affect citizens' awareness of their rights and obligations.

8. Cultural factors.

A culture that adheres to values that are less supportive of citizenship and obedience to the law can also influence the behavior of citizens, for example, the values of violence, nepotism, or corruption that still exist in the culture of a particular society.

9. Political factors.

Unlawful behavior can also be influenced by political factors, such as the existence of corrupt or irresponsible rulers.

Feedback of GAI-based Tutor Assistant

Your response provides good introduction of problems by mentioning the importance of the law and examples of violations of the law. You have also successfully analyzed the causes of unlawful behavior in sufficient detail, covering various factors such as weak law enforcement, public views on violations, low legal compliance, lack of moral and character education, as well as environmental, media, educational, cultural, and political factors. However, your answer is lacking in explaining the role of students from their respective departments/study programs in overcoming these problems. In addition, conclusions and suggestions for policy makers, educational institutions, or the wider community in improving the effectiveness of Civic Education have also not been conveyed. Therefore, you get a score that reflects the strength of your analysis but also considers aspects that have not been covered.

Score: 70 (Max)

Response of Student 2

1. Weak Law Enforcement. How the law is enforced in a region or country has a big impact on the violations that occur.

2. Lack of Legal Compliance. Another factor that contributes to the violation of the law is the lack of legal compliance in society.

3. Efforts that can be made by students to uphold human rights include writing articles and posters about how human rights should be enforced, condemning the wounded victims of human rights violations, conducting defense actions or demonstrations against the injustice received by victims of human rights violations.

Feedback of GAI-based Tutor Assistant

Your answer includes several important points, such as weak law enforcement and lack of legal compliance as the cause of unlawful behavior. You also mentioned several efforts that students can make to uphold human rights. However, your answer is still lacking in explaining the analysis of the causes of unlawful behavior and the role of students from their respective departments/study programs. In addition, your answers have also not provided conclusions and suggestions that reflect a deep understanding of the importance of Citizenship Education. Therefore, the score you get is 40.

Score: 40

As shown by the discussion examples above, the feedback given by the GAI-based tutor assistant is based on the student's response and thus unique and personal in accordance with the respective response.

For comparison, consider the following example of comments from the human tutor to students' responses to the identical discussion trigger. For the human tutor, the marking guide was different from the rubric used for the GAI-based tutor assistant.

Guide for the Human Tutor

1. If students show opinions in accordance with theory, from other sources & UT course materials, and valid data, then give a score of 80–100.
2. If students show opinions accompanied by the following two out of the following elements: according to theory, UT course materials, valid data, then give a score of 60–80.
3. If students show opinions accompanied by one of the following elements: according to theory, sourced from UT course materials, valid data, then give a score of 40–60.
4. If students show opinions without being accompanied by reinforcing elements, such as: according to theory, sourced from BMP, valid data, then then give a score of 20–40.
5. If students only give opinions to strengthen the opinions of other friends, then give a score of 0–20.

Response Student 1

The occurrence of unlawful behavior in our country carried out by the community is caused by various aspects, one of which is the lack of a sense of nationalism, where there is a lack of pride in their national identity. Many Indonesian people, both young and old, are self-interested and do not care or do not want to be involved in the problems that exist around them, both in the scope of the home and the interests of the state. As students, we must increase the sense of nationalism and instill it in ourselves, for example by caring for the environment of the surrounding community, participating in activities organized by the government, and prioritizing common interests over personal interests. Sumer reference: Teaching Materials for Nationalism Education and Training Subjects, Euis Naya Sari, S.ST, M.Si.

Feedback from Human Tutor

Thank you for responding to the discussion in this session, try to first come up with theories related to the discussion sourced from UT course materials, then relate it to cases/data/facts/examples in the community, then analyze. Also write down the reference sources you use to avoid plagiarism.

Score: 50

Response Student 2

In my opinion: The occurrence of unlawful behavior and lack of awareness of rights and obligations as citizens. Although civics education is taught in schools, the curriculum is often adequate but the implementation is less effective. Thus, understanding and awareness of rights and obligations as citizens are not well formed and the lack of effective law enforcement or the existence of corruption in law enforcement institutions can also affect the level of compliance with the law. As a student to increase awareness of the rights and obligations of citizens, especially based on the majors or study programs taken, students can carry out educational and advocacy activities about the rights and obligations of citizens in the surrounding environment, either at school, community, or social media. This can be done through counseling, seminars, discussions, or campaigns that aim to increase public understanding.

Feedback from the Human Tutor

Thank you for responding to the discussion in this session, try to first come up with theories related to the discussion sourced from UT course materials, then relate them to cases/data/facts/examples in the community, then analyze. Also write down the reference sources you use to avoid plagiarism.

Score: 50

As seen from the example above, the feedback given by the human tutor were not as elaborate and personalized as the ones given by AI-based tutor assistant.

FINDINGS AND DISCUSSION

FINDINGS

As described earlier, to see the impact of the use of GAI-based tutor assistants in the online tutoring, data are taken from the second semester of the experiment (the second semester of 2023/2024) where there was 100 virtual classes for each of the 5-courses with a total of 500 classes (and approximately 100.000 students) using Gen-AI-based tutor assistants. Due to some technical problems, only data from four courses is included in the analysis namely MKDU4109 (Basic Cultural Sciences/*Ilmu Sosial dan Budaya Dasar*), MKWU4101 (Islamic Education/*Pendidikan Agama Islam*), MKWU4108 (Indonesian Language/*Bahasa Indonesia*), and MKWUU4109 (Civic Education/*Pendidikan Kewarganegaraan*). For the purposes of data comparison analysis, 100 virtual classes that are not using Gen-AI-based tutors were randomly selected as the control group. Table 1 presents the number of students data that are included in the analysis from both the AI and non-AI tutorial classes. Therefore as seen from the table, the analysis consists of data from 37,743 students.

CLASS	COURSE				TOTAL	%
	MKDU4109	MKWU4101	MKWU4108	MKWU4109		
AI	4621	4720	4673	4699	18713	49.58
Non-AI	4732	4776	4778	4744	19030	50.42
Total	9353	9496	9451	9443	37743	100.00

Table 1 Data/number of students included in the analysis.

In one semester, the tutorials of the bachelor level programs are broken down into 8 sessions. Table 2 shows the mean number of students' posting in each discussion session and in total. The table indicates that, on average, there are somewhat more student postings in the class utilising the GAI-based assistant tutor (7,42) as compared to the class using a human tutor (7.32).

CLASS	SESSION								TOTAL
	1	2	3	4	5	6	7	8	
AI	1.02	.96	1.00	.97	.90	.88	.87	.81	7.42
Non-AI	1.06	.97	.96	.93	.89	.86	.85	.81	7.33
Total	1.04	.97	.98	.95	.89	.87	.86	.81	7.37

Table 2 Mean of students' posting in discussion forum.

A number of normality tests were run for the data variances to determine the significance of those mean differences. The findings of the Kolmogorov-Smirnov test indicate that the significance level is less than 0.05, indicating that the data are not normally distributed (Table 3). The significance of the mean difference between the GAI-based tutor assistance class and the human tutor one must therefore be examined using a Mann-Whitney U test.

SESSION	CLASS	KOLMOGOROV-SMIRNOV ^a		
		STATISTIC	df	SIG.
Session 1	AI	.415	18713	.000
	Non-AI	.448	19030	.000
Session 2	AI	.396	18713	.000
	Non-AI	.419	19030	.000
Session 3	AI	.385	18713	.000
	Non-AI	.421	19030	.000
Session 4	AI	.376	18713	.000
	Non-AI	.419	19030	.000
Session 5	AI	.376	18713	.000
	Non-AI	.419	19030	.000
Session 6	AI	.412	18713	.000
	Non-AI	.465	19030	.000
Session 7	AI	.398	18713	.000
	Non-AI	.445	19030	.000
Session 8	AI	.413	18713	.000
	Non-AI	.433	19030	.000
Total	AI	.253	18713	.000
	Non-AI	.296	19030	.000

Table 3 Normality test for the discussion data variance.

a. Grouping Variable: class.

The results of the Mann-Whitney U test, presented in Table 4, compare the average number of student discussion posts between classes using GAI-based tutor assistants and those with human tutors. The analysis shows that statistically significant differences were found in session 2 ($p = 0.010$), session 7 ($p = 0.038$), and the overall total across sessions ($p = 0.046$), based on a two-tailed significance threshold of 0.05. The significant result for the total sessions indicates that, on average, students in the GAI-assisted classes posted more frequently than those in the human-tutored classes. This suggests that students supported by GAI-based tutors were generally more engaged and active in discussions. These findings are consistent with previous studies, such as those by Chan and Lee (2023) and Nguyen et al. (2024), which also reported increased engagement when AI-supported learning tools were used.

Table 4 Significance test for the mean difference.

a. Grouping Variable: class.

TEST STATISTICS ^a	SESSION 1	SESSION 2	SESSION 3	SESSION 4	SESSION 5	SESSION 6	SESSION 7	SESSION 8	TOTAL
Mann-Whitney U	17585.00	16593.50	17245.50	18300.50	18340.00	18152.50	16863.00	17066.00	16494.50
Wilcoxon W	36113.00	35121.50	35773.50	36828.50	36868.00	36680.50	35391.00	35594.00	35022.50
Z	-1.28	-2.56	-1.70	-.38	-.32	-.56	-2.07	-1.74	-2.00
Asymp. Sig. (2-tailed)	.20	.01	.09	.71	.75	.56	.038	.081	.046

Given the relatively small differences in student discussion postings across sessions (as shown in Table 2), it was important to assess not only the statistical significance but also the practical significance of the results. To do this, Cliff's Delta was used as a non-parametric measure of effect size. This method evaluates the degree of overlap between two distributions, with values ranging from -1 to 1. A value close to 0 indicates a high degree of overlap (i.e., little difference), while values closer to -1 or 1 suggest a large difference between groups. The analysis of the overall data yielded a Cliff's Delta value of 0.01, indicating a very small and practically negligible

effect size (Meissel & Yao, 2024). Thus, while the Mann-Whitney U test showed a statistically significant difference between the GAI-assisted and human-tutored groups, the actual impact of GAI on student engagement was minimal in practical terms.

With regard to achievement, Table 5 presents the mean of students' assignment scores. By design, each student have to complete and submit three assignments during the tutorial sessions. Table 5 demonstrates that, on average, students in the GAI-based tutor assistant class received somewhat higher assignment total scores (206.155) than those in the human tutor class (202.885).

CLASS	ASSIGNMENT			TOTAL
	1	2	3	
AI	72.734	66.75	66.67	206.155
Non-AI	72.269	65.40	65.22	202.885
Total	72.499	66.07	65.94	204.506

Table 5 Mean of students' student assignment scores.

Once more, to ascertain the significance of those mean discrepancies, several normality tests were conducted for the data variances. The results show that the data are not normally distributed, with a significance level of less than 0.05 (Table 6). A Mann-Whitney U test must thus be used to determine the significance of the mean difference between the GAI-based tutor support class and the human tutor one.

ASSIGNMENT	CLASS	KOLMOGOROV-SMIRNOV ^a		
		STATISTIC	df	SIG.
1	AI	.291	18713	.000
	Non-AI	.319	19030	.000
2	AI	.291	18713	.000
	Non-AI	.301	19030	.000
3	AI	.294	18713	.000
	Non-AI	.305	19030	.000
Total	AI	.221	18713	.000
	Non-AI	.234	19030	.000

Table 6 Normality test for the assignment score data variance.

a. Grouping Variable: class.

Table 7 presents the results of the Mann-Whitney U test comparing the average assignment scores between students in GAI-assisted classes and those in human-tutored classes. The analysis shows that, with the exception of Assignment 3, all comparisons yielded statistically significant differences at the $p < 0.05$ level. This indicates that, overall, students in the GAI-based tutoring group performed better on their assignments than those in the control group with human tutors. These findings suggest that GAI-based tutoring may contribute to improved academic achievement and are consistent with the previous study by Chan and Lee (2023), which also found positive effects of AI-supported learning on student performance.

TEST STATISTICS ^a	ASSIGN 1	ASSIGN 2	ASSIGN 3	TOTAL
Mann-Whitney U	13934.000	16362.500	18521.500	15374.000
Wilcoxon W	32462.000	34890.500	37436.500	33902.000
Z	-4.322	-2.089	-.095	-2.974
Asymp. Sig. (2-tailed)	.000	.037	.924	.003

Table 7 Significance test for the mean difference (assignment).

a. Grouping Variable: class.

To further assess the significance and practical impact of the differences in assignment scores (as shown in Table 5), Cliff's Delta was again used to calculate the effect size based on the total data. The analysis produced a value of 0.015, indicating a small but noticeable effect (Meissel & Yao, 2024). Although the effect size is modest, it supports the Mann-Whitney U test findings, confirming that students receiving GAI-based tutoring performed better, on average, than those in classes with human tutors.

The results of this study offer important insights into the potential role of generative AI (GAI)-based tutoring systems in large-scale distance education. The data from four general education courses at Indonesia Open University show that students in GAI-assisted tutorial classes demonstrated a statistically significant increase in discussion participation and slightly higher assignment scores compared to their peers in traditional human-tutored classes. While the differences are relatively small in terms of effect size, they provide meaningful evidence that GAI-based tutoring can positively influence online student engagement and achievement.

ENGAGEMENT: MORE THAN JUST PARTICIPATION?

Student engagement, often considered a proxy for learning involvement and motivation, showed measurable improvement in GAI-based environments. The higher number of student discussion posts in the GAI group supports prior research that suggests GAI tools can promote active learning by providing faster and more personalized responses (Nguyen et al., 2024; Sevnarayan & Potter, 2024). The asynchronous, immediate feedback offered by GAI may encourage students to contribute more confidently, knowing that their inputs will receive attention and evaluation. However, the small Cliff's Delta effect size (0.01) cautions against overstating the practical impact. While students may post slightly more frequently, this does not necessarily indicate deeper engagement or higher-order thinking.

There is also the issue of the *quality* of engagement. Increased quantity of posts could potentially be superficial — driven by the automated nature of feedback rather than critical reflection or dialogue. As other studies have pointed out (Ogunleye et al., 2023; Weeks et al., 2024), GAI tools may lead to a form of passive learning if students rely on AI responses rather than critically engaging with peers or tutors. This raises questions about the pedagogical depth of GAI-mediated discussions and signals the need for future research to explore the *cognitive complexity* of such interactions.

ACHIEVEMENT: SUPPORTIVE, NOT TRANSFORMATIONAL

Assignment scores were also marginally higher among students supported by GAI-based tutors. This suggests that GAI can assist students in producing better work, likely through consistent, rubric-aligned feedback and guidance during the learning process. These findings are in line with Chan & Hu (2023), who reported that students benefit from formative feedback that helps clarify learning goals and performance expectations. Moreover, GAI tools might have helped reduce ambiguity in instructions and provided motivational reinforcement, especially valuable in a distance learning setting where students often report feelings of isolation (Coursera, 2023).

Nevertheless, the modest difference in scores and the low effect size (Cliff's Delta = 0.015) highlight the limited transformative power of GAI when it comes to learning outcomes. These findings align with critiques suggesting that AI alone cannot drive substantial gains in learning unless it is integrated into a well-designed pedagogical framework (Yu & Guo, 2023). The results reinforce the importance of human oversight and instructional design — GAI is most effective when used to enhance, not replace, good teaching.

IMPLICATIONS FOR DISTANCE EDUCATION AT SCALE

From an institutional perspective, the results suggest that integrating GAI-based tutor assistants can be a promising strategy to manage instructional load in large-scale online education. In contexts like UT, where hundreds of thousands of students are enrolled in asynchronous courses, the automation of feedback and moderation through GAI can significantly reduce tutor workload. By shifting human tutors into more strategic roles as moderators or mentors, institutions can maintain quality at scale without overburdening staff.

However, successful implementation requires not only technological integration but also faculty development and clear policy frameworks. The quality of AI feedback heavily depends on the quality of rubrics and prompts provided, which calls for new roles and competencies among educators. It also raises important questions about ethics, fairness, and academic integrity — especially in how AI-generated feedback is used for grading and evaluation.

Limitations and Future Directions

Despite the contributions of this study, several limitations must be acknowledged. First, the analysis focused on basic engagement metrics (number of posts) and aggregate achievement (assignment scores). More nuanced indicators of learning quality, such as discourse analysis, critical thinking assessment, or longitudinal retention rates, were not examined. Second, the study did not account for possible variations in student digital literacy or prior experience with AI tools, which could influence outcomes. Third, the use of Cliff's Delta revealed only minor effects, which should be interpreted with caution in making policy decisions.

Future research should consider mixed-method approaches to capture student perceptions, learning depth, and the emotional dimensions of AI interaction. Additionally, further exploration is needed on how GAI can be adapted to support different disciplines, student profiles, and types of assessments beyond text-based assignments.

CONCLUSION, IMPLICATIONS AND SUGGESTIONS

This study examined the impact of generative AI (GAI)-based tutor assistants on student engagement and academic achievement in the context of distance higher education in Indonesia. Data were collected from four large-enrollment courses (MKDU4109, MKWU4101, MKWU4108, and MKWUU4109), involving approximately 37,743 students. The results indicate that students in GAI-assisted classes participated slightly more in discussion forums (mean of 7.42 posts) compared to those in human-tutored classes (mean of 7.33 posts). The Mann-Whitney U test revealed statistically significant differences in discussion participation, with AI-supported classes showing higher engagement levels. In terms of academic performance, students who received GAI-based tutoring also achieved higher average assignment scores (mean total score of 206.155) compared to their peers with human tutors (mean total score of 202.885).

In conclusion, the use of GAI-based tutor assistants in online tutorial settings appears to offer modest but meaningful improvements in both student engagement and academic achievement. While the effect sizes were small, the findings are consistent with prior research (e.g., [Chan & Lee, 2023](#); [Nguyen et al., 2024](#)) supporting the positive role of AI in education. Importantly, the use of GAI-based tutors did not negatively affect student engagement, suggesting that they can serve as a viable and scalable complement to human tutors. Given their potential to reduce tutor workload and enhance instructional efficiency, GAI-based tutors present a promising approach to supporting online learning and improving student outcomes in large-scale distance education. However, for ethical assurance, the formal deployment of GAI-based tutors at scale must be transparent to students. As a result, further improvement is required to make it more effective and ethical, as well as to have a greater positive impact on student engagement and achievement.

DATA ACCESSIBILITY STATEMENT

The data that support the findings of this study are available from Universitas Terbuka database but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Universitas Terbuka.

SUSTAINABLE DEVELOPMENT GOALS (SDGs)

This study is linked to the following SDG(s): Quality education (SDG 4).

ETHICS AND CONSENT

As data used were collected from the university system and anonymous, no consent and approval of use were requested to the students.

COMPETING INTERESTS

The authors have no competing interests to declare.

Tian Belawati: Conceptualization, methodology, formal analysis, investigation, visualization, writing—original draft preparation, writing—review and editing; Dimas Prasetyo: experimentation, data collection and curation, writing—original draft preparation. All authors have read and agreed to the published version of the manuscript.

AUTHOR NOTES

Based on *Academic Integrity and Transparency in AI-assisted Research and Specification Framework* (Bozkurt, 2024), the authors of this paper acknowledge that this paper was reviewed, edited, and refined with the assistance of Open ChatGPT (the GPT-4 architecture, specifically the GPT-4-turbo, released in November 2023) and Quibott (free web-based version), complementing the human editorial process. The human authors critically assessed and validated the content to maintain academic rigor. The authors also assessed and addressed potential biases inherent in the AI-generated content. The final version of the paper is the sole responsibility of the human authors.

The Introduction section of this paper was initially drafted using Open ChatGPT (the GPT-4 architecture, specifically the GPT-4-turbo, released in November 2023) and Perplexity. This section was subsequently reviewed, critically edited, and validated by the human authors to ensure academic rigor and adherence to ethical standards. The authors also assessed and addressed potential biases inherent in the AI-generated content. The final content, conclusions, and assertions in this paper are the sole responsibility of the human authors.

AUTHOR AFFILIATIONS

Tian Belawati  orcid.org/0000-0001-7354-1500

Universitas Terbuka, Indonesia

Dimas Prasetyo  orcid.org/0000-0002-3213-0171

Universitas Terbuka, Indonesia

REFERENCES

- Bozkurt, A. (2024). GenAI et al.: Cocreation, Authorship, Ownership, Academic Ethics and Integrity in a Time of Generative AI. *Open Praxis*, 16(1), 1–10. <https://doi.org/10.55982/openpraxis.16.1.654>
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20(43). <https://doi.org/10.1186/s41239-023-00411-8>
- Chan, C. K. Y., & Lee, K. K. W. (2023). The AI generation gap: Are Gen Z students more interested in adopting generative AI such as ChatGPT in teaching and learning than their Gen X and millennial generation teachers? *Smart Learn. Environ*, 10(60). <https://doi.org/10.1186/s40561-023-00269-3>
- Coursera. (2023). *Leveraging GAI to empower learners*. Coursera Blog. <https://blog.coursera.org/coursera-coach-leveraging-GAI-to-empower-learners/>
- Guo, L., Wang, D., Gu, F., et al. (2021). Evolution and trends in intelligent tutoring systems research: a multidisciplinary and scientometric view. *Asia Pacific Educ. Rev.*, 22, 441–461. <https://doi.org/10.1007/s12564-021-09697-7>
- Mahajan, V. (2024). *100+ Incredible ChatGPT statistics & facts in 2024*. Notta. <https://www.notta.ai/en/blog/chatgpt-statistics>
- Meissel, K., & Yao, E. S. (2024). Using Cliff's Delta as a non-parametric effect size measure: An accessible web app and r tutorial. *Practical Assessment, Research, and Evaluation*, 29(2), 1–12. <https://openpublishing.library.umass.edu/pare/article/1977/galley/1980/view/#:~:text=Cliff's%20%CE%B4%20can%20be%20interpreted,d%20for%20non%2Dnormal%20data>
- Nguyen, A., Kremantzis, M., Essien, A., Petrounias, I., & Hosseini, S. (2024). Enhancing student engagement through artificial intelligence (AI): Understanding the basics, opportunities, and challenges. *Journal of University Teaching and Learning Practice*, 21(6). <https://doi.org/10.53761/caraaq92>
- Ogunleye, B., Zakariyyah, K. I., Ajao, O., & Olayinka, O. (2023). A Systematic Review of Generative AI for Teaching and Learning Practice. *Journals Education Sciences*, 14(6), 636. <https://www.mdpi.com/2227-7102/14/6/636>
- Ross, P. (2008). Intelligent tutoring systems. *Journal of Computer Assisted Learning*, 3(4), 194–203. <https://doi.org/10.1111/j.1365-2729.1987.tb00331.x>

- Sevnarayan, K., & Potter, M. (2024). Generative Artificial Intelligence in distance education: Transformations, challenges, and impact on academic integrity and student voice. *Journal of Applied Learning and Teaching*, 7(1), 104–114. <https://doi.org/10.37074/jalt.2024.7.1.41>
- Sodiq, A., & Pearson, J. (2024). Can GAI tools help with marking and feedback? FutureLearn. <https://www.futurelearn.com/info/courses/generative-ai-in-higher-education/0/steps/388993>
- Teachflow. (2023). The evolution of AI in education: Past, present, and future. <https://teachflow.ai/the-evolution-of-ai-in-education-past-present-and-future/>
- Tierney, A., Peasey, P. V., & Gould, J. J. M. (2024, Accepted/In press). Student perceptions on the impact of AI on their teaching and learning experiences in higher. <https://doi.org/10.58459/rptel.2025.20005>
- Wecks, J. O., Voshaar, J., Plate, B. J., & Zimmermann, J. (2024). Generative AI usage and academic performance. *arXiv:2404.19699*. <https://doi.org/10.2139/ssrn.4812513>
- White, C. (March 15, 2024). How generative AI can engage and empower students. NewSchool. <https://www.newschools.org/blog/how-generative-ai-can-engage-and-empower-students/>
- Wood, R. (2023). AI + you: How to use ChatGPT for content creation. <https://create.microsoft.com/en-us/learn/articles/how-to-use-chatgpt-for-content-creation>
- Xu, Z. (March 2024). AI in education: Enhancing learning experiences and student outcomes. *Applied and Computational Engineering*, 51(1):104–111. <https://doi.org/10.54254/2755-2721/51/20241187>
- Yu, H., & Guo, Y. (01 June, 2023). Generative artificial intelligence empowers educational reform: Current status, issues, and prospects. *Frontiers in Education, Sec. Digital Learning Innovations*, 8. <https://www.frontiersin.org/journals/education/articles/10.3389/feduc.2023.1183162/full>
- Zhou, X., Zhang, J. J., & Chan, C. (April 2024). Unveiling students' experiences and perceptions of artificial intelligence usage in higher education. *Journal of University Teaching and Learning Practice*, 21(6). <https://doi.org/10.53761/xzjprb23>

TO CITE THIS ARTICLE:

Belawati, T., & Prasetyo, D. (2025). Generative AI-Based Tutoring for Enhancing Learning Engagement and Achievement. *Open Praxis*, 17(2), pp. 211–226. DOI: <https://doi.org/10.55982/openpraxis.17.2.902>

Submitted: 11 April 2025

Accepted: 10 June 2025

Published: 10 July 2025

COPYRIGHT:

© 2025 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Open Praxis is a peer-reviewed open access journal published by International Council for Open and Distance Education.