THE EFFECTIVENESS OF USING ARTIFICIAL INTELLIGENCE (AI) TECHNOLOGY IN ENHANCING JUNIOR HIGH SCHOOL STUDENTS' MATHEMATICAL PROBLEM SOLVING SKILLS

Febry Azmiana Siregar[1], Niken Adelina Silalahi[2], Emya Lavigana Barus[3], Feby Armelia Rosi[4], Cornelia Valent Christin Sitanggang[5], Hanna Novita Sihombing[6], Daniel Sitohang[7]
Postgraduate Program in Mathematics Education, State University of Medan
E-mail: febryazmiana2802@gmail.com

Abstract
This study aims to evaluate the effectiveness of using Artificial Intelligence (AI) technology through ChatGPT in enhancing junior high school students' mathematical problem-solving skills. The research employed a quantitative method with a quasi-experimental design, specifically the pretest-posttest control group design, involving two groups: an experimental class using ChatGPT and a control class using conventional learning methods. The study involved 52 eighth-grade students at SMP Adhyaksa Medan, selected through purposive sampling. The research instruments included pretest and posttest assessments to measure problem-solving skills. Data analysis showed a significant improvement in both groups, with the experimental group showing a higher increase. The independent t-test revealed a significant difference between the posttest results of the two groups, and the N-Gain calculation showed an average of 69.85 in the experimental group (medium category) and 29.85 in the control group (low category). These findings indicate that using ChatGPT as an AI-based interactive learning tool is effective in improving students' mathematical problem-solving skills.
Keywords: Artificial Intelligence, Mathematical Problem-Solving Skills

1. Introduction

In this era of globalization and rapid technological advancement, the field of education is required to continuously adapt in order to produce a generation capable of facing the challenges of the 21st century. Sophisticated technologies such as Artificial Intelligence (AI), big data, and the Internet of Things (IoT) have transformed the way people work, think, and learn (Pratiwi & Mahmuddin, 2025). In this context, conventional learning methods are no longer sufficient to meet the demands of the times. Therefore, education must shift toward a more dynamic, innovative, and technology-based approach.

Problem-solving ability is an essential skill in mathematics learning and constitutes a crucial component of higher-order thinking skills that are highly needed in the 21st century (Irawadi, 2025; Suryandari et al., 2021). Through this ability, students are expected to analyze problems, formulate solution strategies, and make logical and systematic decisions. The National Council of Teachers of Mathematics (2000) states that problem-solving ability is a primary goal in mathematics education, where learning is focused, among other things, on the proficiency to apply mathematical concepts and skills to solve problems.

However, real conditions in the field show that students' problem-solving abilities—especially in mathematics—are still relatively low. Many students struggle to solve non-routine problems that require logical reasoning, conceptual understanding, and structured thinking strategies. This is supported by research conducted by Anggraeni and Kadarisma (2020), which found that students' mathematical problem-solving skills are still not optimal. Similarly, Asih and Ramdhani (2019) discovered that low levels of higher-order thinking are one of the main causes of weak problem-solving skills in mathematics learning. These conditions highlight an urgent need to implement more innovative, adaptive, and student-centered learning approaches that align with the characteristics of today's learners. Learning that remains teacher-centered and lacks exploration prevents students from fully engaging in the thinking and problem-solving processes (Mariani et al., 2025). Therefore, it is necessary to provide instructional interventions that foster independent learning, encourage critical thinking, and offer meaningful opportunities for students to experience the problem- solving process firsthand.

One of the opportunities that can be leveraged to address these challenges is the integration of digital technology into learning. The development of digital technology has opened up significant opportunities in education, particularly through the implementation of Artificial Intelligence (AI). In the field of education, AI is used in various forms, such as intelligent tutoring systems, teaching robots, adaptive learning, and human-

computer interaction (Ouyang & Pengcheng, 2021). One of the most exciting advancements in AI application for education is the emergence of generative AI tools like ChatGPT. ChatGPT is an AI-based language model developed by OpenAI, capable of explaining material, answering questions, and providing feedback interactively and in real time. In the context of mathematics learning, ChatGPT can engage students in interactive dialogue, offer personalized support, and explain concepts in a more natural and easily understood manner (Farhi et al., 2023; Govender, 2023). This technology offers a more engaging learning experience tailored to each student's needs (Rifky, 2024). With its ability to provide instant feedback, AI enables students to learn at their own pace and according to their learning styles, making the learning process more personalized, adaptive, and efficient (Govender, 2023; Korkmaz Guler et al., 2024).

Several previous studies have shown that the use of AI in education can enhance learning efficiency, provide a more personalized learning experience, support student engagement, and improve learning outcomes (Diantama, 2023; Riska et al., 2025; Yassir & Saharuna, 2024). In line with this, research conducted by Egara et al. (2025) indicated that the use of ChatGPT in mathematics learning can support conceptual understanding and problem-solving skills, create more personalized learning experiences, accommodate diverse learning styles, and expand problem-solving strategies in the classroom.

However, most existing studies still focus on the context of higher education, while research on the effectiveness of artificial intelligence technology particularly ChatGPT in mathematics learning at the junior high school level remains very limited. The lack of empirical studies evaluating the impact of AI on junior high school students' problem-solving abilities reveals a gap in the literature. Therefore, this study aims to examine the effectiveness of using artificial intelligence (AI) technology in improving junior high school students' mathematical problem-solving skills. This research is expected to fill the gap in previous studies and provide a meaningful contribution to the development of adaptive and innovative technology-based learning models that align with the demands of 21st-century student competencies.

2. Research Method

This study aims to analyze the effectiveness of using Artificial Intelligence (AI) technology, particularly through ChatGPT, in improving junior high school students' mathematical problem-solving skills. A quantitative approach was employed using a quasi-experimental design with a pretest-posttest control group format. This design involved two groups: an experimental group and a control group. Both groups were given a pretest before the intervention and a posttest afterward to determine the effectiveness of AI technology in enhancing students' mathematical problem-solving abilities. The research subjects consisted of 52 eighth- grade students at SMP Adhyaksa Medan, selected through purposive sampling. The experimental group included 26 students who received mathematics instruction assisted by AI technology (such as ChatGPT), while the control group consisted of 26 students who received conventional instruction without AI technology.The research instruments included pretest and posttest assessments in the form of open-ended questions targeting mathematical problem-solving skills, structured according to Polya's four problem- solving steps: understanding the problem, devising a plan, carrying out the plan, and evaluating the solution. Content validity was reviewed by three experts, and the reliability of the instrument was tested using internal consistency analysis. Based on the analysis, all test items met the criteria for validity.

The study began with the administration of a pretest to both the experimental and control groups. The pretest aimed to measure the students' initial mathematical problem-solving abilities before the intervention. The pretest scores served as the basis for identifying the average initial abilities of the students. Subsequently, the learning process was carried out according to the treatment assigned to each group. The experimental group received instruction using Artificial Intelligence (AI) technology in the form of ChatGPT as an interactive learning tool. In this activity, students used their digital devices to interact with ChatGPT, either individually or in groups. They were able to ask questions about concepts they did not understand, request explanations of problem-solving steps, and receive feedback on their responses. Over several sessions, ChatGPT was used in a structured manner under the guidance of the teacher to ensure that the learning process remained focused and aligned with the learning objectives. Meanwhile, the control group received conventional instruction without the use of AI. After the entire learning process was completed, both groups were given a posttest to measure any changes in their mathematical problem-solving skills.

The hypothesis used in this study is as follows:

H_0: There is no significant improvement in students' mathematical problem-solving skills after using artificial intelligence (AI) technology.

H_1: There is a significant improvement in students' mathematical problem-solving skills after using artificial intelligence (AI) technology

The data analysis techniques in this study included normality and homogeneity tests as prerequisites for conducting parametric tests. An independent t-test was used to compare the posttest results between the experimental and control groups, while a paired t-test was used to examine the improvement within each group. If the data met the assumptions of normality and homogeneity, an N-gain score and N-gain percentage analysis were conducted to measure the effectiveness of using ChatGPT in enhancing students' mathematical problem-solving skills. The N-gain score was calculated using Hake's (1998) formula as follows:

$$\text{N Gain score} = \frac{(Posttest\ Score - Pretest\ Score)}{(Maximum\ Score - Pretest\ Score}$$

Furthermore, the categorization of N-gain scores used to measure the effectiveness level of AI implementation is presented in Table 1 below.

Table 1. N-Gain Score Categories

N-gain score Range	**Effectiveness Category**
g > **0,7**	High
0,3 ≤ g ≤ **0,7**	Medium
g < **0,3**	Low

Source: (Hake, 1998)

For the categorization of N-gain percentage (%) according to Arikunto (1999), the interpretation of effectiveness can be seen in Table 2 below.

Table 2. Interpretation Categories of N-Gain Percentage Effectiveness

Percentage (%)	**Category**
< 40	Not Effective
40 – 55	Less Effective
56 – 75	Fairly Effective
>75	Effective

Source: (Arikunto, 1999)

3. Result and Discussion

The research involved 52 eighth-grade students at SMP Adhyaksa Medan, consisting of 26 students in the control group and 26 students in the experimental group. Pretest data were collected to assess students' mathematical problem-solving skills before the intervention, while posttest data were collected to evaluate their skills after the intervention. The data from both groups were analyzed to determine whether the students who learned using ChatGPT technology showed greater improvement in mathematical problem-solving abilities compared to those who did not use the ChatGPT application during the learning process.

Table 3. Results of Students' Mathematical Problem-Solving Skills

Group	Pretest	Posttest	Minimum Score		Maximum Score	
	Mean	Mean	Pretest	Postest	Pretest	Posttest
Experimental (AI)	46,69	84,65	35	73	70	95
Control (Conventional)	44,34	61,61	30	45	68	75

Based on the data in Table 3, the average initial problem-solving ability of students in the experimental class was 46.69, while in the control class it was 44.34. This indicates that before the intervention, the initial abilities of both groups were relatively balanced, with a difference of only 2.35 points. From this description, it can be concluded that there was no significant difference between the pre-test scores of students in the two classes.

The average posttest score of students' problem-solving abilities in the experimental class was 84.65, while in the control class it was 61.61. This difference indicates that students in the experimental class achieved higher

scores compared to those in the control class, suggesting that the problem-solving skills of students in the experimental group were better. Thus, there is a significant difference in the posttest results of problem-solving abilities between the two classes.

Normality Test

The normality test was conducted to determine whether the data were normally distributed. In this study, the normality test was applied to both pretest and posttest scores using the Kolmogorov-Smirnov test with the help of SPSS software. The results of the normality test are presented in Table 4 below.

Table 4. Normality Test Results of Students' Mathematical Problem-Solving Skills Data

Data Type	Group	Kolmogorov-smirnov		
		Statistic	Df	Sig.
Pretest	Control	0,162	26	0,077
	Experimental	0,141	26	0,080
Postest	Control	0,142	26	0,196
	Experimental	0,129	26	0,20

Based on Table 4, the results of the normality test for students' pretest problem-solving data in both classes show significance values of 0.077 and 0.080, respectively—both of which are greater than or equal to 0.05. Therefore, the pretest data are normally distributed. Furthermore, the normality test results for the posttest data indicate significance values of 0.196 and 0.20 for the control and experimental classes, respectively—both exceeding 0.05. Thus, the posttest data for both classes are also normally distributed. Given that both the pretest and posttest data meet the assumption of normal distribution, the analysis can proceed with the homogeneity test.

Homogeneity Test

The homogeneity test is a statistical procedure used to determine whether the variances of the data are homogeneous across groups. In this study, the Levene's Test for Equality of Variances was used, assisted by SPSS software. The results of the homogeneity test are presented below.

Table 5. Homogeneity Test Results of Students' Mathematical Problem-Solving Skills Data

Data Type	Levene Statistic	df1	df2	Sig.
Pretest	0,003	1	50	0,956
Posttest	2,711	1	50	0,106

From Table 5, it can be seen that the significance value in Levene's Test for the pretest data is 0.956 > 0,05. This indicates that the pretest data on students' mathematical problem-solving skills in both classes have homogeneous variances. Similarly, for the posttest data, the Levene's significance value is 0.106 > 0,05. This shows that the posttest data for both groups also have homogeneous variances. Therefore, the data meet the homogeneity assumption, and the analysis can proceed with the Independent Samples T-test to compare the posttest results between the experimental and control groups.

Independent Samples T-test

The Independent Samples T-test was used to compare the posttest results between the experimental and control groups to determine whether there was a significant difference as a result of the AI-based (ChatGPT) learning intervention.

Table 6. Independent Samples T-test Results (Posttest)

Nilai t	df	Sig. (2-tailed)
9,925	50	0,000 (< 0,05)

The results of the independent t-test showed that the t-value was 9.925 with a significance value (Sig. 2-tailed) of 0.000. Since the significance value is less than 0.05, it can be concluded that there is a significant difference between the posttest results of students in the control group and the experimental group.

Paired Samples t-Test

The Paired Samples t-Test was conducted to determine the improvement from pretest to posttest scores within each group, both the experimental and control groups.

Table 7. Paired Samples t-Test Results

Group	***Sign 2 tailed***
Experimental	0.000
Control	0.032

In the experimental class, the results of the paired samples t-test show a Sig. (2-tailed) value of 0.000, indicating a significant increase between the pretest and posttest scores. Meanwhile, in the control class, the Sig. (2-tailed) value is 0.032, which also indicates an improvement, although not as substantial as in the experimental class. Since the significance values are less than 0.05, it can be concluded that H_0 is rejected. The rejection of H_0 leads to the acceptance of H_1, which states that "there is a significant improvement in students' mathematical problem-solving abilities after using artificial intelligence (AI) technology." This hypothesis is supported by the difference in average scores between the pretest and posttest, where the posttest mean is notably higher than the pretest mean.

N-Gain Score Test

Next, the N-Gain Score test was conducted to determine the level of effectiveness of using AI specifically ChatGPT in improving students' mathematical problem-solving skills. The results of the N-Gain test are shown in Table 7 below.

Table 8. N-Gain Score Results

Group	N	N-Gain score	N-Gain Percentage
Experimental	26	0,6985	69,85
Control	26	0,2985	29,85

Based on the data in Table 8, it is shown that the average N-Gain score in the experimental class is 69.85, while in the control class it is 29.85. This indicates that the improvement in students' mathematical problem- solving abilities after the learning intervention was higher in the class that used artificial intelligence (AI) technology, namely ChatGPT. According to Hake's (1998) classification, the N-Gain score for the experimental class falls into the medium category, while the control class falls into the low category. These findings indicate that the use of ChatGPT as an interactive learning tool is proven to be more effective in enhancing students' problem-solving abilities compared to conventional teaching methods

Overall, the results of the study indicate that the use of artificial intelligence (AI) technology specifically through the ChatGPT application significantly enhances junior high school students' mathematical problem-solving abilities compared to conventional teaching methods. The significant improvement in the experimental group is evident from the higher average posttest scores, statistically significant t-test results, and the N-Gain category that falls within the medium to high range.

The use of AI assists students in understanding problem-solving steps in an interactive and personalized manner, accelerates feedback, and encourages logical and systematic thinking. This aligns with the characteristics of 21st-century learning, which emphasize the integration of technology to enhance students' problem-solving abilities. These findings also support previous studies showing that technology-based learning can improve student engagement, conceptual understanding, and learning motivation. Therefore, the implementation of AI in mathematics instruction can serve as an effective strategy to address the complex challenges of learning in the digital era.

4. Conclusion

Based on the results of the study, it can be concluded that the use of artificial intelligence (AI) technology through ChatGPT in mathematics learning is effective in improving junior high school students' problem-solving abilities. This is evidenced by a significant increase in pretest and posttest scores in both the experimental and control classes, with the experimental group showing a much greater improvement. The results of the independent t-test revealed a significant difference between the two groups, and the N-Gain analysis further supported this finding, with the experimental class achieving an average N-Gain score of 69.85, which falls into the medium category. Thus, AI-based learning contributes positively to the development of students' critical thinking and mathematical problem-solving skills.

Bibliography

Anggraeni, R., & Kadarisma, G. (2020). Analisis kemampuan pemecahan masalah matematik siswa SMP kelas VII pada materi himpunan. Jurnal Cendekia : Jurnal Pendidikan Matematika. 4(2), 1072– 1082. doi: https://doi.org/10.31004/cendekia.v4i2.334

Arikunto, S. (1999). Dasar-Dasar Evaluasi Pendidikan Edisi 2. Bumi Aksara.

Asih, N., & Ramdhani, S. (2019). Peningkatan kemampuan pemecahan masalah matematis dan kemandirian belajar siswa menggunakan model pembelajaran means end analysis. Jurnal Pendidikan Matematika. 8(3), 435–446. doi: https://doi.org/10.31980/mosharafa.v8i3.534

Diantama, S. (2023). Pemanfaatan Artificial Intelegent (AI) Dalam Dunia Pendidikan. DEWANTECH: Jurnal Teknologi Pendidikan. 1(1): 8 – 14.

Egara, F. O., Mogege, M & Moeketsi, M. (2025). Secondary School Students' Perceptions of Their Usage of Artificial Intelligence-Based ChatGPT in Mathematics Learning. Journal of Education. 98: 124 – 146. doi: https://doi.org/10.17159/2520-9868/i98a07.

Farhi, F., Jeljeli, R., Aburezeq, I., Dweikat, F. F., Al-shami, S. A., & Slamene, R. (2023). Analyzing The Students' Views, Concerns, and Perceived Ethics About ChatGPT Usage. Computers and Education: Artificial Intelligence. 5. 1-8. doi: https://doi.org/10.1016/j.caeai.2023.100180.

Govender, R. (2023). The Impact of Artificial Intelligence and The Future of Chatgpt for Mathematics Teaching and Learning In Schools and Higher Education. Pythagoras. 44(1).1-2. doi: https://doi.org/10.4102/PYTHAGORAS.V44I1.787.

Hake, R. R. (1998). Interactive-Engagement Versus Traditional Methods: A Six-Thousand Student

Survey of Mechanics Test Data for Introductory Physics Courses. American Journal of Physics. 66(1), 64–74. doi: https://doi.org/10.1119/1.18809

Irawadi, A. A., Sripatmi., Junaidi & Baidowi. (2025). Kemampuan Pemecahan Masalah Matematis Siswa Ditinjau dari Kepercayaan Diri dan Kecemasan Matematika. Journal of Classroom Action Research. 7(1). 22 – 32. doi: https://doi.org/10.29303/jcar.v7i1.10081.

Kemdikbud. (2023). Transformasi Digital dalam Pendidikan di Indonesia. Diakses dari https://www.kemdikbud.go.id/transformasi-digital-pendidikan

Korkmaz Guler, N., Dertli, Z. G., Boran, E., & Yildiz, B. (2024). An Artificial Intelligence Application in Mathematics Education: Evaluating Chatgpt's Academic Achievement in A Mathematics Exam. Pedagogical Research. 9(2). doi: https://doi.org/10.29333/pr/14145.

Mariani, D., Mustaji & Utari, D. (2025). Pengaruh Model Problem Based Flipped Learning terhadap Kemampuan Berpikir Kreatif dan Kemampuan Pemecahan Masalah Matematis Siswa SMP Sekolah Indonesia Kuala Lumpur. JIIP (Jurnal Ilmiah Ilmu Pendidikan). 8 (2): 1492-1497.

Natsir, Nurasia., dkk. (2003). Belajar di Era Digital. Mutiara Intelektual Indonesia Press

NCTM. (2000). Principles and Standards for School Mathematics. Reston: NCTM.

Ouyang, Fan & Pengcheng Jiao. (2021). Artificial Intelligence in Education: The Three Paradigms. Computers and Education: Artificial Intelligence. 2. doi: https://doi.org/10.1016/j.caeai.2021.100020.

Pratiwi, R. T. L & Mahmuddin, Y. (2025). Manfaat dan Tantangan Penggunaan Artificial Intelligence (AI) Bagi Guru dan Peserta Didik di Era Society 5.0. Journal of Innovation and Teacher Professionalism. 3(2): 488 – 494.

Rifky, Sehan. (2024). Dampak Penggunaan Artificial Intelligence Bagi Pendidikan Tinggi. Dampak Penggunaan Artificial Intelligence Bagi Pendidikan Tinggi. Indonesian Journal of Multidisciplinary on Social and Technology. 2 (1): 37-42. doi: https://doi.org/10.31004/ijmst.v2i1.287.

Riska, Nina., Ila, R & Dase, E. J. (2025). Integrasi Teknologi AI Dalam Pembelajaran Adaptif Untuk Meningkatkan Keterampilan Abad 21 di Sekolah Dasar. Jurnal Inovasi dan Teknologi Pendidikan. 4(1): 180 – 198. doi: https://doi.org/10.46306/jurinotep.v3i1.

Suryandari, K. C., Rokhmaniyah, & Wahyudi. (2021). The effect of scientific reading based project model in empowering creative thinking skills of preservice teacher in elementary school. European Journal of Educational Research. 10(3), 1329–1340. doi: https://doi.org/10.12973/eu-jer.10.3.1329

Yasir, M & Saharuna. (2024). Pengaruh Artificial Intelligence (AI) Terhadap Hasil Belajar Mahasiswa yang Dimediasi oleh Motivasi Belajar dan Kreativitas. Jambura: Journal of Educational Management. 5(1): 45-54