Journal of General Education and Humanities
Vol. 5, No. 1, February 2026, pp. 2043 – 2058, https://doi.org/10.58421/gehu.v5i1.1131
ISSN 2963-7147

The Effect of Using ChatGPT (Generative Pre-Trained Transformer) and Gemini AI (Artificial Intelligence) on the Laziness of Thinking of Islamic Religious Education Students at the State Islamic University of Sulthan Thaha Saifuddin Jambi in the Process of Working on Assignments

Lala Rahayu Winara[1], Ahmad Ridwan[2], SettingsMuhamad Taridi[3]
[1,2,3]Program Studi Manajemen Pendidikan Islam Universitas Islam Negeri Sulthan Thaha Saifuddin Jambi, Indonesia

Article Info

Article history:

Received 2026-02-04
Revised 2026-02-09
Accepted 2026-02-27

Keywords:

Artificial Intelligence
Chatgpt
Gemini AI
Mental Laziness
Thinking Sloth

ABSTRACT

The use of generative AI such as ChatGPT and Gemini AI among Islamic Religious Education (PAI) students at UIN STS Jambi offers task efficiency, but has the potential to trigger laziness in thinking through cognitive offloading and disrupt the internalization of thinking, contemplation, and tazkiyah. This study aims to analyze the partial and simultaneous effects of both AIs on laziness in thinking and completing thesis assignments. A quantitative quasi-experimental posttest-only control group approach was used with a population of 5th-semester PAI students (N≈100), a sample of 56 respondents (experimental n=25, control n=31) via purposive sampling. The Likert questionnaire instrument was validated by Pearson and Cronbach's reliability $\alpha>0.70$; SPSS analysis: descriptive, t-test, multiple linear regression (F/t test, R^2). Results: Gemini AI had a significant effect ($p<0.01$, $\beta=0.919$), ChatGPT did not ($p>0.05$); Simultaneous $R^2=0.877$ ($p<0.01$), t-test $p=0.01$ ($d=0.72$). Conclusion: The differential impact of multimodal AI strongly triggers laziness; an ethical strategy is needed for PAI to integrate AI as a reflection aid.

Corresponding Author:

Lala Rahayu Winara
Program Studi Manajemen Pendidikan Islam Universitas Islam Negeri Sulthan Thaha Saifuddin Jambi
Email: lala01rahayu@gmail.com

1. INTRODUCTION

The phenomenon of using generative artificial intelligence such as ChatGPT and Gemini AI is increasingly widespread among Islamic Religious Education (PAI) students, offering high efficiency in completing academic tasks with quick and easy access [1]. Research at STIT Pemalang shows that 65-75% of students use it regularly, making it an integral part of daily learning activities, even though it triggers passive dependence that reduces substantial self-understanding [2]. At UIN Sulthan Thaha Saifuddin Jambi, an initial

survey of 150 PAI students revealed that 75% used AI without in-depth verification, with 65% having difficulty explaining concepts during presentations. Another study found a positive correlation between the intensity of ChatGPT and a decrease in motivation for independent research at ITS [3].

This phenomenon is increasingly complex because AI triggers mental laziness through cognitive offloading, where students shift their cognitive load to technology, at the expense of the process of critical and reflective analysis. For PAI students, this dependence interferes with the internalization of Islamic values such as fikr (deep thinking), tafakur (spiritual reflection), and tazkiyah an-nafs (cleansing of the soul), as affirmed by Q.S. Ali Imran 190-191, which demands contemplation of Allah's creation [4]. Superficial practices such as SKS (Overnight Speed System) are increasingly rampant, prioritizing the result rather than internalizing knowledge, and undermining naqd al-marwiyyat as the basis of Islamic science [5].

The problem is even more urgent because previous research has emphasized the efficiency of AI, such as the speed of tasks and access to information, while there has been a lack of study on its impact on critical thinking and spiritual values in PAI [6]. The erosion of Higher-Order Thinking Skills (HOTS) and metacognition due to AI instant solutions has not been widely measured empirically, especially the differences between ChatGPT (text-based) and Gemini AI (multimodal) in the context of PAI tasks [7]. At UIN STS Jambi, the absence of AI ethical policies weakens the process of fikr and tazkiyah, so that quantitative mapping is needed to prevent the cognitive-spiritual degradation of students [8].

This study aims to analyze the influence of ChatGPT and Gemini AI partially and simultaneously on the laziness of thinking of PAI UIN STS Jambi students in their assignments, with a high urgency to fill the interdisciplinary literature gap between educational technology and Islamic values in the digital era. Its novelty lies in a quantitative approach that differentiates the types of AI and integrates the spiritual indicators of PAI, providing a theoretical-practical basis for an ethical pedagogical strategy that makes AI an auxiliary tool, rather than a substitute for profound cognitive processes [9].

2. METHOD

Types and Research Methods

This study applies a quantitative approach with a quasi-experimental method using a posttest-only control group design to measure the effect of the use of ChatGPT (X1) and Gemini AI (X2) on the laziness of thinking (Y) of PAI UIN Sulthan Thaha Saifuddin Jambi students in working on assignments [10]. The quantitative approach was chosen because it allows the collection of statistically analyzed numerical data to produce objective conclusions and strong generalizations of the population. At the same time, the quasi-experimental design is appropriate for higher education contexts where individual randomization is difficult to do ethically and practically, thus using intact classes as units of analysis. In this design, the experimental group received the treatment of using AI for the task, while the control group did not, followed by a posttest on both to compare the thinking laziness scores, thus controlling for pretest bias and improving internal validity compared to the one-group pretest-posttest design [11].

Data Collection Instruments and Techniques

The main instrument was a questionnaire based on a 4-point Likert scale (Strongly Agree=4 to Strongly Disagree=1) which measured the variables X1 (use of ChatGPT: frequency, critical verification, dependence), X2 (use of Gemini AI: multimodal frequency, validation, thinking influence), and Y (laziness of thinking: instant dependence, decrease in HOTS, lack of spiritual reflection of fikr-tafakur-tazkiyah), with a total of 43 items after validation [12]. Primary data collection techniques were through post-treatment self-administered questionnaires and documentation (attendance, assignments), while secondary data from journal literature and institutional reports for triangulation [13]. The instrument grid was constructed based on conceptual operational definitions, tested for construct validity with Product Moment ($r_{hitung} > r_{tabel}=0.355$, $df=29$, $\alpha=0.05$) and Cronbach's Alpha reliability (>0.7: X1=0.788 high, X2=0.739 high, Y=0.934 very high) using SPSS 25, ensuring accurate and consistent instruments for interval data from the Likert scale.

Population and Sample

The research population consists of 336 PAI students in the 5th semester of UIN STS Jambi who are active and experienced in academic tasks, chosen because they are relevant to the adaptation of AI technology such as ChatGPT and Gemini AI. Samples were taken through cluster random sampling of 12 classes, with lots yielding 2 intact classes: PAI A (experimental, n=25) and PAI C (control, n=31), a total of 56 respondents, according to quasi-experimental practices in which preexisting classes are used to maintain ecological equivalence and avoid contamination [14]. The difference in group size is accepted methodologically because of the intact group, by checking the similarity of respondent characteristics via descriptive (frequency, mean), to minimize selection bias.

Research Procedure

The procedure begins with the preparation of the instrument (grid, n=31 test for validity-reliability), class sample draw, assignment of similar tasks (e.g. PAI paper), treatment (experimental group is required to use ChatGPT & Gemini AI; manual control), followed by posttest questionnaires for O1 (experiment) and O2 (control), as well as process documentation. Data analysis included descriptive (mean, percentage), classical assumption test (Kolmogorov-Smirnov normality $p>0.05$, VIF<10 multicollinearity, Glejser heteroscedasticity $p>0.05$), Independent t-test (or Mann-Whitney if non-normal), multiple linear regression for partial (t-test) and simultaneous (F, R^2 test) influences, with SPSS 25 and criterion $p<0.05$ minus H0 [15]. The research was conducted at the Faculty of Tarbiyah UIN STS Jambi, in the even semester 2025/2026, during the intensive final project, for temporal validity [16].

3. RESULTS

Deskripsi Data Penelitian

a. Descriptive Statistics of ChatGPT Usage Variable (X1)

Table 1. Descriptive Statistics on the Use of ChatGPT (Experimental Class)

		X1
N	Valid	25
	Missing	0
Mean		35.16
Std. Error of Mean		1.325
Mode		33
Std. Deviation		6.625
Variance		43.890
Range		28
Minimum		17
Maximum		45
Sum		879

The use of ChatGPT in the experimental group showed an average of 35.16 with a standard deviation of 6.625, indicating a very high intensity of use. The middle score is 78, the minimum score is 39, and the maximum score is 80.

Table 2. Descriptive Statistics of ChatGPT Use (Control Class)

		X1
N	Valid	31
	Missing	0
Mean		28.77
Std. Error of Mean		1.137
Mode		24[a]
Std. Deviation		6.328
Variance		40.047
Range		26
Minimum		19
Maximum		45
Sum		892

The results of the calculation using SPSS 25 on the data after the treatment of the control class obtained a valid sample of 31, with a slightly higher average score. The use of ChatGPT in the experimental group showed a lower average of 28.77 with a smaller standard deviation of 6.328, a minimum value of 19, and a maximum value of 45.

b. Frequency Distribution and Histogram of ChatGPT Usage (X1)

Table 3. Frequency Distribution of ChatGPT Usage (Experimental Class)

No	Interval	F	Presentase
1	42 – 46	5	20%
2	37 – 41	5	20%
3	32 – 36	9	36%
4	27 – 31	4	16%
5	22 – 26	1	4%
6	17 – 21	1	4%
Total		25	100%

The frequency distribution showed that most of the students in the experimental class were in the high-medium category, particularly at intervals of 32-36 (36%) and 37-41 (20%). This indicates that the ability or tendency measured by the X1 variable in the experimental class tends to be in the medium to high category, with few respondents in the low category.

Table 4. Frequency Distribution of ChatGPT Usage (Control Class)

No	Interval	F	Presentase
1	44 – 48	1	3.2%
2	39 – 43	2	6.5%
3	34 – 38	2	6.5%
4	29 – 33	9	29.0%
5	24 – 28	12	38.7%
6	19 – 23	5	16.1%
Total		31	100%

The results of the distribution showed that the majority of students in the control class were in the low-medium category, especially in the interval of 24-28 (38.7%). Although there are scores in the high category, the number is relatively small. Thus, the variable X1 in the control class tends to be in the medium to lower category.

c. Descriptive Statistics of Gemini AI Usage Variable (X2)

Table 5. Descriptive Statistics of Gemini AI Usage (Experimental Class)

		X2
N	Valid	25
	Missing	0
Mean		26.00
Std. Error of Mean		1.066
Mode		25[a]
Std. Deviation		5.331
Variance		28.417
Range		22
Minimum		12
Maximum		34
Sum		650

The use of Gemini AI in the experimental group showed considerable variation among respondents, with an average of 26.00 with a standard deviation of 5,331, a minimum value of 39, and a maximum value of 80.

Table 6. Descriptive Statistics of ChatGPT Usage (Control Class)

		X2
N	Valid	31
	Missing	0
Mean		22.19
Std. Error of Mean		.668
Mode		20
Std. Deviation		3.719
Variance		13.828
Range		13
Minimum		17
Maximum		30
Sum		688

The results of calculations using SPSS 25 on the data after the control class treatment obtained 31 valid samples, Gemini AI Usage scores with a lower average of 22.19, a smaller standard deviation of 3.719, a minimum value of 17, and a maximum value of 30.

d. Frequency Distribution and Histogram of Gemini AI (X2) Usage

Table 7. Frequency Distribution of Gemini AI Usage (Experimental Class)

No	Interval	F	Presentage
1	32 – 35	3	12%
2	28 – 31	8	32%
3	24 – 27	8	32%
4	20 – 23	3	12%
5	16 – 19	2	8%
6	12 – 15	1	4%
Total		25	100%

The X2 value in the experimental class showed dominance in the medium-high category, especially at intervals 24-27 and 28-31, of 32% each. This condition is in line with the average X2 value of 26, which shows a tendency to be in the medium-high category.

Table 8. Frequency Distribution of Gemini AI Usage (Control Class)

No	Interval	F	Presentage
1	26 – 28	7	22.6%
2	23 – 25	6	19.4%
3	20 – 22	12	38.7%
4	17 – 19	6	19.4%
	Total	31	100%

Based on the table and histogram above, the frequency of using Gemini AI in the control class shows that most students are in the medium category, especially in the 20-22 interval, with the highest frequency of 38.7%. Intervals 26-28 that fall into the high category have a frequency of 22.6%, while the other two intervals, namely 17-19 (low) and 23-25 (medium-high), each recorded a frequency of 19.4%. These results show that the X2 value in the control class is more concentrated in the medium category, in line with the average X2 value of 22.19. Thus, the X2 variable in the control class tends to be at a moderate level of competence or tendency.

e. Descriptive Statistics of the Laziness of Thinking variable (Y)

Table 9. Descriptive Statistics of Laziness of Thinking (experimental class)

N	Valid	25
	Missing	0
Mean		62.80
Std. Error of Mean		2.411
Mode		78
Std. Deviation		12.055
Variance		145.333
Range		41
Minimum		39
Maximum		80
Sum		1570

The results of the calculation using SPSS 25 on the data showed that the number of valid samples was 25, the average score of the level of laziness thinking was relatively high, 62.80, with a standard deviation of 12.055, a minimum score of 39, and a maximum score of 80.

Tabel 10. Distribusi Frekuensi Kemalasan Berpikir (Kelas Kontrol)

		Y
N	Valid	31
	Missing	0
Mean		55.52
Std. Error of Mean		1.520
Mode		63
Std. Deviation		8.465
Variance		71.658
Range		38
Minimum		44
Maximum		82
Sum		1721

The results of the calculation using SPSS 25 on the data showed that the number of valid samples in the control class was 31, the average score of the level of laziness thinking was relatively low, 55.52, with a standard deviation of 8,465, a minimum value of 44, and a maximum value of 82.

f. Frequency Distribution and Histogram of Laziness of Thinking (Y)

Table 11. Frequency Distribution of Laziness of Thinking (Experimental Class)

No	Interval	F	Presentase
1	74 – 80	5	20%
2	67 – 73	6	24%
3	60 – 66	5	20%
4	53 – 59	5	20%
5	46 – 52	1	4%
6	39 – 45	3	12%
Total		25	100%

The frequency distribution showed that the majority of respondents in the experimental class were in the high category, particularly in the 67-73 interval (24%). Intervals 60-66, 53-59, and 74-80 also recorded high frequencies (20% each), thus illustrating a strong tendency that the level of the thinking laziness window (Y) is in the high category. This is consistent with a mean of 62.80.

Table 12. Frequency Distribution of Laziness of Thinking (Control Class)

No	Interval	F	Presentase
1	74 – 80+	1	3.2%
2	68 – 73	0	0%
3	62 – 67	9	29.0%
4	56 – 61	5	16.1%
5	50 – 55	6	19.4%
6	44 – 49	10	32.3%
	Total	31	100%

Hypothesis Testing

1. Group difference test

Table 13. Independent Samples t-test results

		t-test for Equality of Means					
		F	**t**	**df**	**Sig. (2-tailed)**	**Mean Difference**	**Std. Error Difference**
Laziness of Thinking	Equal variances assumed	3.829	-2.652	54	0.010	-7.284	2.747
	Equal variances not assumed		-2.555	41.617	0.014	-7.284	2.850

2. Partial test (t-test)

Table 14. Summary of Partial Test (T-test)

Class	Variabel	t_{count}	Sig.	Remarks
Eksperimen	X1→ Y	0.078	0.938	Insignificant
	X2→ Y	3.926	0.001	Signifikan
Control	X1→ Y	-0.448	0.658	Insignificant
	X2→ Y	1.834	0.077	Insignificant

Based on the results of the t-test in the table above, it can be seen that in the experimental class, the variable X1 (the use of ChatGPT) has a Sig value of 0.938 > 0.05 and a calculation of 0.078 < the table (2.074). Thus**, H1 is rejected,** which means that it has no significant effect on the variable Y (Laziness of Thinking). In contrast, the variable X2 (Use of Gemini AI) showed a sig value of 0.001 < 0.01 and a tcount of 3,926 > ttable (2,074). Thus, **H2 is accepted**, so it can be concluded that X2 has a significant effect on the laziness of students' thinking in the experimental class.

For the control class, the X1 variable (ChatGPT Usage) had a significance value of 0.658 > 0.05 and t count -0.448 < t table (2.048). Thus, **H4 was rejected**. This means that there is no significant influence of X1 on Y. Variable X2 (Use of Gemini AI) has a

significance value of 0.077 > 0.05 and t calculation of 1.834 < t table (2.048). Thus, , **H5 is rejected**. This means that there is no significant influence of X2 on Y.

3. Simultaneous Test (F Test)

Table 15. Simultaneous Test Results (F Test)

Experiment						
Model		**Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
1	Regression	3057.505	2	1528.752	78.125	.000[b]
	Residual	430.495	22	19.568		
	Total	3488.000	24			
Control						
Model		**Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
1	Regression	311.676	2	155.838	2.374	.112[b]
	Residual	1838.066	28	65.645		
	Total	2149.742	30			

Based on the results of the F test in the table above, it is known that in the experimental class, the significance value is 78.125 with a significance of 0.000 < 0.05. Count 78.125 > Ftable (3.44). Thus, **H3 is accepted**. This means that simultaneously the variables used by ChatGPT (X1) and Gemini AI (X2) have a significant effect on Y (Laziness of Thinking).

On the other hand, in the control class, the Fcal value was 2.374 with a significance of 0.112 > 0.05. Count 2,374 > Ftable (3.34). Thus, **H6 was rejected**. This means that simultaneously the variables used by ChatGPT (X1) and Gemini AI (X2) did not have a significant effect on Y (Laziness of Thinking).

4. Coefficient of Determination (R2)

Tabel 16. Hasil uji Determinasi (R2)

Experimental Classes							
Model	**R**	**R Square**	**Adjusted R-Square**	**Std. Error of the Estimate**	**Change Statistics**		
					R Square Change	**F Change**	**df1**
1	.936[a]	.877	.865	4.424	.877	78.125	2
Control Class							
Model	**R**	**R Square**	**Adjusted R-Square**	**Std. Error of the Estimate**	**Change Statistics**		
					R Square Change	**F Change**	**df1**
1	.381[a]	.145	.084	8.102	.145	2.374	2

Based on the results of the table above, it shows that in the experimental class, the value of R = 0.936 and the value of the determination coefficient R-squared is 0.877. showed that the independent variables X1 (Use of ChatGPT) and X2 (Use of Gemini AI) simultaneously explained 87.7% of the variation in student laziness of thinking in the context of this study. Statistically, this value seems very high. However, in social and educational research, a very high R2 value cannot be directly interpreted as an ideal model strength, but needs to be criticized academically.

The high R^2 value in this study has the potential to be influenced by the homogeneity of respondent characteristics coming from relatively uniform study programs and academic environments. This condition can cause data variation to be limited, so that the relationship between variables appears to be statistically stronger. In addition, the similarity of traits between independent variables, namely ChatGPT and Gemini AI as a form of generative AI that is often used simultaneously, can result in *shared variance* that magnifies the value of R^2.

The R^2 value of 0.877 in this study is not interpreted as evidence of an absolute causal relationship, but as an indicator of the strong empirical relationship between variables in the context and limitations of this study. The interpretation of the results of this study emphasizes caution in drawing conclusions and avoiding excessive generalizations outside the context of the sample being studied.

Meanwhile, in the control class, the value of R = 0.381, and the value of the determination coefficient R-squared was 0.145, which means only 14.5%. Variables X1 and X2 can explain variations in Laziness Thinking, while the remaining 85.5% are influenced by other factors outside the study, such as learning motivation, learning environment, and students' academic habits. This shows that the regression model in the control class is relatively weak.

The significant difference in R^2 values between the experimental class and the control class reflects the differences in the relationship characteristics between the variables in the two groups. In experimental classes that were given AI-based treatments, the relationship between AI use and thinking laziness appeared to be statistically strong. In contrast, in control classes that were not given the treatment of using AI, the relationship between variables tended to be weak to moderate. This condition shows that without the intervention of the use of AI, students' lazy thinking is more influenced by other factors, such as learning motivation, teaching methods, academic environment, and individual characteristics of students.

A very high difference in R^2 values in the experimental class is not necessarily interpreted as absolute causal proof. The high R^2 value in the experimental class needs to be criticized academically, given the possible homogeneity of the respondents and the similarity of the characteristics of the analyzed AI variables. Therefore, the difference between the experimental class and the control class in this study is understood as an empirical indication that the use of AI has a stronger association with student laziness of thinking than with conditions without the use of AI, within the constraints of the context and design of the research used.

5. DISCUSSION

The Effect of the Use of ChatGPT on the Laziness of PAI Students

The results of the data regression analysis conducted by the research (t-test) in tables 14 and 16 showed that in the experimental class, the regression coefficient of using ChatGPT (X1) was 0.033 with a significance value **of 0.938** ($p>0.05$). The *standardized coefficient* (Beta) value is very small, at 0.018.

Based on these figures, it can be concluded that in the context of this study, the intensity of ChatGPT use does not have a statistically significant linear relationship with an increase in students' thinking laziness scores. In other words, the variation in how many college students use ChatGPT does not explain the variation in their level of lazy thinking. Therefore, Hypothesis 1 (H1), which states that there is a significant influence of the use of ChatGPT on students' lazy thinking, is rejected.

These findings are interesting and contradict common assumptions. Theoretically, this can be explained through the lens of *cognitive tools* and *metacognition.* ChatGPT, with its text-based interface, requires users to actively formulate questions (*prompting*), evaluate the relevance of answers, and rearrange them. This process involves cognitive *monitoring* and *regulation*, so it does not necessarily replace the thought process, but rather moves it to a different level (from seeking answers to evaluating answers)." "Empirically, this pattern may be specific to the context of PAI students. The limited use of the free version of ChatGPT and a strong understanding of academic ethics are natural moderators that prevent total dependence."

In the framework of Islamic education, these findings are in line with the concept *of tabayyun* (clarification) in Q.S. Al-Hujurat:6.

يٰٓاَيُّهَا الَّذِيْنَ اٰمَنُوْٓا اِنْ جَاۤءَكُمْ فَاسِقٌۢ بِنَبَاٍ فَتَبَيَّنُوْٓا اَنْ تُصِيْبُوْا قَوْمًاۢ بِجَهَالَةٍ فَتُصْبِحُوْا عَلٰى مَا فَعَلْتُمْ نٰدِمِيْنَ ۝٦

"*O you who have believed, if a wicked person comes to you with important news, then examine the truth so that you do not harm a people because of your ignorance, which causes you to regret your deeds*." (Q.S. Al-Hujurat:6).

Another factor that reinforces these findings is the technical limitations of the free version of ChatGPT used by college students. With a limited number of requests and features, students cannot rely entirely on ChatGPT to complete complex tasks that require in-depth analysis of religious texts. In addition, students' understanding of academic ethics in the Islamic education environment also plays an important role in preventing plagiarism and excessive dependence.

The critical habits of PAI students in verifying information that may also be applied to ChatGPT output are a bulwark against intellectual laziness. Thus, ChatGPT can be seen as a neutral will; The impact depends on the user's intent and *metacognitive* awareness.

The effect of using Gemini AI on the laziness of thinking of Islamic Education students

In contrast, the results for the Gemini AI usage variable (X2) showed a dramatically different pattern. The regression coefficient was 2.078 with a significance **of 0.001** ($p<0.01$). The **value of the standardized coefficient** (Beta) reaches 0.919, which is very large. In regression models, Gemini AI's unique contribution to Y variance is very dominant.

These figures indicate a strong and statistically significant positive relationship between the intensity of Gemini AI use and an increase in thinking laziness scores. The one-unit increase in Gemini AI usage was associated with a 2,078-point increase in the thinking laziness score, with a standard effect (Beta 0.919) approaching 1, showing a near-perfect

linear relationship in this sample. Thus, hypothesis 2 (H2), which states that there is a significant influence of the use of Gemini AI on the laziness of thinking, is accepted.

These findings can be explained through the extreme *theory of Cognitive Offloading* and the technical characteristics of Gemini AI. As an integrated multimodal model, Gemini AI provides more comprehensive solutions (text, images, audio) with minimal *user effort*. This facilitates *lazy thinking* or *heuristic processing* (System 1 Kahneman's quick thinking) at the expense of *analytical thinking* (System 2).

From the point of view of Bloom's Revised Taxonomy, this AI capability that is too "complete" has the potential to bypass *the* student analysis and evaluation stage, directly to the creation or finished work stage, which actually eliminates the essence of HOTS (*Higher-Order Thinking Skills*) learning.

From an Islamic perspective, the high dependence on tools like this has the potential to erode the practices of tafakkur (deep contemplation) and tadabbur (appreciation), which are the core of the development of the intellect (aql). Allah SWT says in Q.S. Ali Imran verses 190-191:

اِنَّ فِيْ خَلْقِ السَّمٰوٰتِ وَالْاَرْضِ وَاخْتِلَافِ الَّيْلِ وَالنَّهَارِ لَاٰيٰتٍ لِّاُولِى الْاَلْبَابِۚ ﴿١٩٠﴾
الَّذِيْنَ يَذْكُرُوْنَ اللّٰهَ قِيَامًا وَّقُعُوْدًا وَّعَلٰى جُنُوْبِهِمْ وَيَتَفَكَّرُوْنَ فِيْ خَلْقِ السَّمٰوٰتِ وَالْاَرْضِۚ رَبَّنَا مَا خَلَقْتَ هٰذَا بَاطِلًاۚ سُبْحٰنَكَ فَقِنَا عَذَابَ النَّارِ ﴿١٩١﴾

What it means: *"Indeed, in the creation of the heavens and the earth and the alternation of night and day there are signs (of Allah's greatness) for the sensible, those who remember Allah while standing, sitting, or lying down, and think about the creation of the heavens and the earth (saying), "O our Lord, You did not create all these things in vain. You are holy. Protect us from the punishment of Hell".* (Q.S. Ali Imran: 190-191).

That the signs of Allah's greatness can only be understood by *the ulul albab* (the intellectual), i.e., those who use their intellect to think and meditate, t*his large effect size* is an empirical warning that technology that is too easy can reduce the space for the deep and contemplative thinking process commanded in Islam.

Simultaneous influence of ChatGPT and Gemini AI on the laziness of thinking

The results of the simultaneous test (Test F, Table 15) and the determination coefficient in the experimental class showed a very significant regression model ($p < 0.000$) with a value of R^2 = **0.877**. However, coefficient analysis revealed an unbalanced composition: significant contributions came only from Gemini AI (X2) (Beta 0.919), while ChatGPT (X1) (Beta 0.018) was insignificant. Plus, a high VIF value (9,763) indicates a serious multicollinearity problem between X1 and X2.

This finding contains two important meanings. First, the use of both AI together is indeed very strong in "predicting" thinking laziness (high R^2). Second, and more critically, this predictive power is dominated almost entirely by Gemini AI, while ChatGPT statistically makes no meaningful unique contributions. The high R^2 reflects more of the *shared variance* between X1 and X2 (since students tend to use both) and the strength of X2, rather than the combined contribution of the two independent variables.

Therefore, although Hypothesis 3 (H3) regarding simultaneous influence is technically accepted based on the F-test, its interpretation must be approached with caution. The visible "simultaneous" influence is more of a reflection of *the dominant influence of Gemini AI* that "carries" ChatGPT variables in the model because the two are highly correlated.

The problem of multicollinearity itself elegantly explains this phenomenon. In practice, college students do not strictly separate the use of ChatGPT and Gemini AI; they are very close substitutes for similar academic goals. As a result, in statistical models, the variables X1 and X2 "overlap" with each other. What is measurable as the actual "combined" influence is the influence of a latent construct that we can call "Generative AI Use Intensity", where Gemini AI becomes a stronger proxy due to its more multimodal and integrated capabilities.

From a pedagogical point of view, this reinforces the thesis that it is not just the use of AI, but the type and capabilities of AI that are the determining factors for its cognitive impact. AI that provides overly comprehensive solutions like Gemini is more at risk of causing *cognitive offloading* than more limited and dialogue-based AI like ChatGPT.

From the perspective of Islamic education, these findings require caution so that the use of technology does not shift the goal of education from the development of intellect and character to a mere outcome orientation. The Qur'an reminds people not to follow something without knowledge and consideration of reason, as Allah SWT says in QS. Al-Isra' [17]: 36.

وَلَا تَقْفُ مَا لَيْسَ لَكَ بِهٖ عِلْمٌۗ اِنَّ السَّمْعَ وَالْبَصَرَ وَالْفُؤَادَ كُلُّ اُولٰۤىِٕكَ كَانَ عَنْهُ مَسْـُٔوْلًا ﴿٣٦﴾

What it means: *"Do not follow something you do not know. Indeed, hearing, sight, and conscience, all of these will be held accountable."* (Q.S. Al-Isra': 36)

This verse forbids following the unknown. This high R2 finding could be a metaphor; the reliance on powerful Generative AI can create the illusion of understanding. In the context of AI-based learning, the warning can be interpreted as a necessity for students to keep the process of analytical and reflective thinking active, so that technology functions as a learning tool, not as a substitute for the role of reason in understanding and processing knowledge.

Comparison of experimental classes and control classes

The results of the independent samples t-test are in Table 13. showed a statistically significant difference between the thinking laziness score in the experimental group and the control group, with a significance value of 0.010 (0.<05). The experimental group that used ChatGPT and Gemini AI had an average of 62.80 thinking laziness, while the control class that did not use ChatGPT and Gemini AI had an average of 55.52. This average difference is -7,287 points.

This statistically significant finding ($p < 0.010$) provides strong evidence that there is a real difference between the two groups. However, in social and educational research, statistical significance alone is not enough. Therefore, *an effect size analysis* was carried out using Cohen's d to find out the magnitude of the difference practically. The calculation of

effect size yields a value of d = 0.72; this value is included in the medium to large *effect category*.

This suggests that the differences between the experimental and control groups are not only statistically significant but also have important practical implications. The magnitude of *the effect size* of 0.72 indicates that the use of AI has a substantial impact on increasing the tendency of laziness in students' thinking. In the context of education, an effect of this magnitude can be considered to have serious consequences on the quality of the student learning process.

This significant difference is in line with *the theory of cognitive offloading*. Experimental groups that have access to ChatGPT and Gemini AI tend to offload or transfer their cognitive load to external systems (AI). In contrast, control groups that did not use AI had to rely on their own internal cognitive resources.

In the experimental group, the ease of access to instant solutions from AI reduced the need to perform independent thinking processes, deep analysis, and information synthesis. This is in line with the findings that the lazy score thinks they are higher. In the control group, the unavailability of "instant" AI tools requires students to go through a more complete cognitive process, ranging from information search, analysis, evaluation, to synthesis. Although this process is more challenging, it *actually trains higher-order thinking skills* (HOTS).

Reflection and synthesis of researchers

In the series of analysis and discussion above, the researcher understands that the dynamics of the influence of artificial intelligence on the laziness of PAI students' thinking are complex and not monolithic. The study's findings do not support the simple narrative that "AI must make you lazy to think". Instead, what was revealed was an impact differentiation that was heavily influenced by the technical characteristics of the AI tool (such as Gemini's multimodal capabilities versus the dialogical nature of ChatGPT) and most likely by user characteristics (such as the tabayyun value internalized by PAI students).

As a researcher, I interpret the very high R^2 findings (0.877) in the experimental class not as proof of a perfect prediction model, but as a warning light of how powerful generative AI, especially multimodal ones, can be in "gripping" student learning patterns. *The effect size* of the difference in groups in the mid-large category (d = 0.72) confirms that this warning has a real empirical basis. However, ChatGPT does not necessarily partially provide a gap in optimism, that not all AI technologies have the same impact, and that there is room for healthier utilization if the tool is designed or used with the right approach.

This research raises a paradox of the digital era in Islamic education. On the one hand, artificial intelligence allows unprecedented access and efficiency of knowledge, which is in line with the spirit of Islam in the pursuit of knowledge. On the contrary, it can interfere with *the process of tafakkur* and *tadabbur*, which are at the core of a strong religious understanding. The result that the control group (without artificial intelligence) had a lower score of thinking laziness is a reminder of the important value of cognitive effort, process-based learning, and the intellectual struggle of *jihad al-'aql*.

As a researcher, I am fully aware of the limitations of this study. Posttest-only design and perception-based instruments limit the claims of causality and depth of behavioral analysis. Therefore, I interpret these findings not as the end of a debate, but rather as paving the way for a richer discussion. This study managed to map the symptoms and identify risk patterns, especially related to Gemini AI, but the question of "how should it be?" AI being integrated into the PAI curriculum that still adheres to the values of fikr and tazkiyah requires further exploration, perhaps with a *mixed-methods* approach that involves direct observation and in-depth interviews."

My position as a researcher and aspiring educator of Islam led me to see these findings as a call to action in a pedagogical manner. The identified risks must be balanced with intelligent learning engineering. The future of Islamic religious education in the era of AI does not lie in the rejection of technology, but in the critical mastery of it, namely the ability to make AI an assistant in the process of tafakur, not as a substitute for reason, which is a divine grace. This is the main task of educators in the future; this research can contribute by providing initial data to start the task.

6. CONCLUSION

This study found that the use of Gemini AI partially had a significant positive effect on the laziness of PAI students' thinking ($p<0.01$, beta=0.919), while ChatGPT was insignificant ($p>0.05$). Simultaneously, both predicted 87.7% of the laziness variance ($R^2=0.877$, $p<0.01$) in the experimental group, with Cohen's effect size d=0.72 showing a moderate-to-large difference in the control group. These findings confirm the differential impact of Gemini's multimodal AI through cognitive offloading that reduces HOTS and internalization of fikr-tafakur-tazkiyah, especially in the thesis assignment of UIN STS Jambi.

However, limitations include posttest-only designs that limit absolute causality, subjective perception instruments, and high multicollinearity between AIs that affect simultaneous interpretation. Suggestions for further research include mixed-methods with longitudinal observations and a broader sample to test the moderation of Islamic ethics. Practically, the implications demand ethical guidelines: PAI lecturers integrate AI as a reflection prompter (not an instant solution), implement QS-based self-verification. Al-Hujurat:6, and UIN's institutional policy for a hybrid curriculum that strengthens jihad al-aql to prevent spiritual degradation in the digital era [17].

REFERENCES

[1] Y. K. Dwivedi *et al.*, "Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy," *Int. J. Inf. Manag.*, vol. 57, p. 101994, Apr. 2021, doi: 10.1016/j.ijinfomgt.2019.08.002.

[2] F. Gandasari, A. S. Koeswinda, A. K. Putri, D. A. P. Kumala, and N. Muftihah, "Etika Pemanfaatan Teknologi Artificial Intelligence dalam Penyusunan Tugas Mahasiswa," *EDUKATIF J. ILMU Pendidik.*, vol. 6, no. 5, pp. 5572–5578, Aug. 2024, doi: 10.31004/edukatif.v6i5.7036.

[3] Alya Resti Saraswati, Vasya Ayu Karmina, Maharani Putri Efendi, Zahrina Candrakanti, and Nur Aini Rakhmawati, "Analisis Pengaruh ChatGPT Terhadap Tingkat Kemalasan Berpikir Mahasiswa ITS Dalam Proses Pengerjaan Tugas," *J. Pendidik. Bhs. Dan Budaya*, vol. 2, no. 4, pp. 40–48, Oct. 2023, doi: 10.55606/jpbb.v2i4.2223.

[4] M. H. Alfurqan and W. Wirdati, “Analisis Penggunaan Website Berbasis AI dalam Mengerjakan Tugas Akademik pada Mahasiswa Pendidikan Agama Islam Universitas Negeri Padang,” *ISLAMIKA*, vol. 6, no. 4, pp. 1727–1739, Oct. 2024, doi: 10.36088/islamika.v6i4.5309.
[5] Miftahul Huda and Irwansyah Suwahyu, “Peran Artificial Intelligence (Ai) Dalam Pembelajaran Pendidikan Agama Islam,” *REFERENSI Islam. J. Studi Islam*, vol. 2, no. 2, pp. 53–61, Aug. 2024, doi: 10.61220/ri.v2i2.005.
[6] Y. K. Dwivedi *et al.*, "Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy," *Int. J. Inf. Manag.*, vol. 57, p. 101994, Apr. 2021, doi: 10.1016/j.ijinfomgt.2019.08.002.
[7] Alya Resti Saraswati, Vasya Ayu Karmina, Maharani Putri Efendi, Zahrina Candrakanti, and Nur Aini Rakhmawati, “Analisis Pengaruh ChatGPT Terhadap Tingkat Kemalasan Berpikir Mahasiswa ITS Dalam Proses Pengerjaan Tugas,” *J. Pendidik. Bhs. Dan Budaya*, vol. 2, no. 4, pp. 40–48, Oct. 2023, doi: 10.55606/jpbb.v2i4.2223.
[8] Miftahul Huda and Irwansyah Suwahyu, “Peran Artificial Intelligence (Ai) Dalam Pembelajaran Pendidikan Agama Islam,” *REFERENSI Islam. J. Studi Islam*, vol. 2, no. 2, pp. 53–61, Aug. 2024, doi: 10.61220/ri.v2i2.005.
[9] F. Gandasari, A. S. Koeswinda, A. K. Putri, D. A. P. Kumala, and N. Muftihah, "Etika Pemanfaatan Teknologi Artificial Intelligence dalam Penyusunan Tugas Mahasiswa," *EDUKATIF J. ILMU Pendidik.*, vol. 6, no. 5, pp. 5572–5578, Aug. 2024, doi: 10.31004/edukatif.v6i5.7036.
[10] B. Capili and J. K. Anastasi, "An Introduction to Types of Quasi-Experimental Designs," *AJN Am. J. Nurs.*, vol. 124, no. 11, pp. 50–52, Nov. 2024, doi: 10.1097/01.NAJ.0001081740.74815.20.
[11] A. S. Sitanggang *et al.*, "An Analysis of the Impact of ChatGPT Usage on University Student's Creativity and Critical Thinking in the AI Era," in *Proceedings of the 8th International Conference on Informatics, Engineering, Science & Technology (INCITEST 2025)*, vol. 287, L. Warlina and S. Luckyardi, Eds., in Advances in Engineering Research, vol. 287. , Dordrecht: Atlantis Press International BV, 2025, pp. 201–210. doi: 10.2991/978-94-6463-924-7_17.
[12] F. Gandasari, A. S. Koeswinda, A. K. Putri, D. A. P. Kumala, and N. Muftihah, "Etika Pemanfaatan Teknologi Artificial Intelligence dalam Penyusunan Tugas Mahasiswa," *EDUKATIF J. ILMU Pendidik.*, vol. 6, no. 5, pp. 5572–5578, Aug. 2024, doi: 10.31004/edukatif.v6i5.7036.
[13] Alya Resti Saraswati, Vasya Ayu Karmina, Maharani Putri Efendi, Zahrina Candrakanti, and Nur Aini Rakhmawati, “Analisis Pengaruh ChatGPT Terhadap Tingkat Kemalasan Berpikir Mahasiswa ITS Dalam Proses Pengerjaan Tugas,” *J. Pendidik. Bhs. Dan Budaya*, vol. 2, no. 4, pp. 40–48, Oct. 2023, doi: 10.55606/jpbb.v2i4.2223.
[14] Alya Resti Saraswati, Vasya Ayu Karmina, Maharani Putri Efendi, Zahrina Candrakanti, and Nur Aini Rakhmawati, “Analisis Pengaruh ChatGPT Terhadap Tingkat Kemalasan Berpikir Mahasiswa ITS Dalam Proses Pengerjaan Tugas,” *J. Pendidik. Bhs. Dan Budaya*, vol. 2, no. 4, pp. 40–48, Oct. 2023, doi: 10.55606/jpbb.v2i4.2223.
[15] A. S. Sitanggang *et al.*, "An Analysis of the Impact of ChatGPT Usage on University Student's Creativity and Critical Thinking in the AI Era," in *Proceedings of the 8th International Conference on Informatics, Engineering, Science & Technology (INCITEST 2025)*, vol. 287, L. Warlina and S. Luckyardi, Eds., in Advances in Engineering Research, vol. 287. , Dordrecht: Atlantis Press International BV, 2025, pp. 201–210. doi: 10.2991/978-94-6463-924-7_17.
[16] M. H. Alfurqan and W. Wirdati, “Analisis Penggunaan Website Berbasis AI dalam Mengerjakan Tugas Akademik pada Mahasiswa Pendidikan Agama Islam Universitas Negeri Padang,” *ISLAMIKA*, vol. 6, no. 4, pp. 1727–1739, Oct. 2024, doi: 10.36088/islamika.v6i4.5309.
[17] A. N. Afifah and I. Ermiana, "The Effects of the Jigsaw Type Cooperative Learning Model and Self-Efficacy on Elementary School Students' Scientific Literacy Skills," vol. 4, 2025.