



“I Didn’t Know That was Plagiarism”: EFL Students' AI-Giarism in a Pesantren-Based Higher Education

Durratul Hikmah^{1*}, Indah Riskiana², Wardatus Zahro³

^{1,2,3}Universitas Nurul Jadid Probolinggo, Indonesia

*Corresponding Author. E-mail: durrohhikmah@unuja.ac.id

Received: 4 August 2026 Revised: 3 September 2026 Accepted: 3 October 2026

Abstrak

Merebaknya alat kecerdasan buatan generatif (Gen-AI) telah mengaburkan batasan etis penulisan akademik dan memicu pergeseran definisi plagiarisme dari bentuk konvensional menuju plagiarisme berbasis AI (AI-giarism). Mahasiswa *English as a Foreign Language* (EFL) menjadi kelompok yang paling rentan terjebak dalam praktik ini akibat kecemasan menulis dalam bahasa kedua yang mendorong ketergantungan pada AI sebagai jalan pintas. Penelitian ini bertujuan mengkaji tingkat pengetahuan, sikap, dan praktik plagiarisme berbasis AI yang dilaporkan mandiri oleh mahasiswa EFL di perguruan tinggi berbasis pesantren. Menggunakan desain survei kuantitatif cross-sectional, data dikumpulkan melalui kuesioner tervalidasi 51 item yang diadaptasi dari Phyo et al. (2023), disebarkan kepada 99 mahasiswa program studi Pendidikan Bahasa Inggris (PBI) pada sebuah perguruan tinggi berbasis pesantren di Jawa Timur, Indonesia, melalui Google Forms. Data dianalisis menggunakan statistik deskriptif, korelasi Spearman, regresi linear berganda, dan regresi logistik biner. Hasil menunjukkan tingkat pengetahuan pada kategori sedang (mean = 9,63/17; 56,6%), dengan kelemahan terbesar pada pemahaman bahwa teks AI yang lolos Turnitin tetap merupakan plagiarisme (hanya 33,3% benar). Sikap menolak plagiarisme berada pada level moderat (mean = 91,24/135; 67,6%), dengan sub-skala membenaran diri paling rendah (55,5%). Sebanyak 50,5% responden pernah melakukan setidaknya satu tindakan AI-giarism. Skor pengetahuan terbukti sebagai satu-satunya prediktor independen signifikan perilaku plagiarisme AI (AOR = 0,764; 95% CI [0,62–0,94]; p = 0,012). Temuan ini mengungkap "Paradoks Pesantren": benturan antara modal moral keagamaan yang kuat dan keterbatasan literasi etis-digital yang membuka celah bagi AI-giarism yang tidak disadari.

Kata Kunci: integritas akademik; kerangka KAP, mahasiswa EFL, AI-giarism; pesantren,

Abstract

The rise of generative AI (Gen-AI) tools has blurred the ethical boundaries of academic writing, shifting the definition of plagiarism from traditional methods to AI-based plagiarism (AI-giarism). English as a Foreign Language (EFL) students are the most vulnerable people to fall into this practice due to second language writing anxiety, which drives dependency on AI as a shortcut. This study aims to examine the levels of knowledge, attitudes, and self-reported AI-based plagiarism practices among EFL students at a pesantren-based higher education institution. Employing a quantitative cross-sectional survey design, data were collected using a validated 51-item questionnaire adapted from Phyo et al. (2023), distributed to 99 English education students at a pesantren-based higher education in East Java, Indonesia, via Google Forms. Data were analyzed using descriptive statistics, Spearman's correlation, multiple linear regression, and binary logistic regression. Result revealed a moderate level of knowledge (mean = 9.63/17; 56.6%), with the greatest weakness in understanding that AI-generated text passing Turnitin still constitutes plagiarism (only 33.3% correct). Attitudes rejecting plagiarism were at a moderate level (mean = 91.24/135; 67.6%), with the self-justification subscale scoring the lowest (55.5%). A total of 50.5% of respondents had engaged in at least one act of AI-giarism. Knowledge score emerged as the sole significant independent predictor of AI plagiarism behavior (AOR = 0.764; 95% CI [0.62–0.94]; p = 0.012). These findings reveal the “Pesantren Paradox”: a conflict



between strong religious moral capital and limited digital-ethical literacy, which creates a loophole for unintended AI-giarism.

Keywords: *academic integrity, AI-giarism, EFL students, KAP framework, pesantren.*

How to cite: Hikmah, D., Riskiana, I. & Zahro, W. (2026). I Didn't Know That was Plagiarism": EFL Students' AI-Giarism in a Pesantren-Based Higher Education. *Journal of Nusantara Education*, 6(1), 124-134. DOI: <https://doi.org/10.57176/nz563a04>

Introduction

The rise of generative artificial intelligence (Gen-AI) tools in the global higher education landscape has fundamentally transformed students' academic writing paradigms. Large Language Model (LLM)-based tools such as ChatGPT, QuillBot, and Grammarly are now massively integrated into students' scholarly activities, ranging from idea generation and automated paraphrasing to text editing (Marzuki et al., 2023; Naidu & Sevnarayan, 2023; Zhao, 2023). While this technology offers tangible efficiencies in language learning, its presence gives rise to serious ethical disruptions (Acosta-Enriquez et al., 2024): the boundaries between legitimate writing assistance and academic misconduct are increasingly blurred, giving rise to an ambiguous space that students actively leverage to legitimize the improper use of AI (Nilufar, 2026; Pahlevi et al., 2024; Sozon et al., 2025).

In Indonesia, the academic integrity crisis reached a critical juncture when an identity manipulation and AI-based research fabrication scandal was exposed at the International Symposium on Pneumococci and Pneumococcal Diseases (ISPPD) 2026 in Copenhagen, Denmark (KOMPAS, 2026). This scandal reveals a deeper reality that AI technology has been normalized even at senior reviewer levels as a tool for instant manipulation, representing a phenomenon that unconsciously transforms into a hidden curriculum that legitimizes academic shortcuts among students (Zhai et al., 2024). When the academic ecosystem at higher levels demonstrates such extreme ethical compromise, it is understandable that students rationalize lesser infractions, such as submitting AI-generated text, as commonplace and acceptable.

English as a Foreign Language (EFL) students represent a particularly susceptible group to AI misutilization. This heightened vulnerability stems from L2 writing anxiety, limited linguistic confidence, and the heavy cognitive load required to negotiate complex

syntactic, lexical, and rhetorical standard in academic English (Hyland, 2026). Under intense pressure to produce native-like academic prose within tight deadlines, EFL students frequently view generative AI tools not merely as writing assistants, but as instant remedies to overcome language barriers and mitigate performance anxiety. Malik et al. (2023) and Aminah and Aly (2026) document Indonesian students' highly positive perceptions of AI, particularly for paraphrasing functions and plagiarism management. Many EFL students believe that submitting AI-generated text does not constitute cheating as long as it passes Turnitin detection (Prashar et al., 2024). This definitive shift is defined by Chan (2025) as AI-giarism, which refers to the failure to attribute the role of AI in the cognitive process of scientific writing, representing a category of misconduct that is conceptually distinct and more ambiguous than conventional plagiarism.

Although global literature has begun to explore this issue, a crucial research gap remains. The majority of existing Knowledge-Attitude-Practice (KAP) studies on plagiarism, including Phyo et al. (2023) in Myanmar and Magida et al. (2023) in South Africa, were designed prior to the mass adoption of Gen-AI, thereby failing to capture the dimension of AI-giarism as a phenomenologically distinct category. Furthermore, studies on AI and academic integrity within the EFL context predominantly focus on postgraduate students at Western universities or Gulf Cooperation Council (GCC) countries (Lund et al., 2025), paying very limited attention to undergraduate EFL students in non-Western contexts. Specifically, the present study extends prior KAP research in three ways: (1) by incorporating AI-giarism as a distinct construct absent from previous instrument; (2) by targeting undergraduate EFL students who remain underrepresented in existing Gen-AI literature; and (3) by situating the inquiry within a pesantren-based institutional context that

introduces a unique moral-ethical dimension not captured in Western or GCC-focused studies.

Pesantren-based higher education institutions generally integrate Islamic ethical values, such as truthfulness (*shiddiq*), trustworthiness (*amanah*), and noble character (*akhlakul karimah*), into their academic culture. However, limited digital-ethical literacy and the pressure of EFL writing anxiety create a context-specific dynamic observed in this study: a tension between strong religious moral capital on one hand, and limited digital-ethical literacy on the other, which opens a loophole for unintended AI-giarism. Rather than positing a broad sociological paradox, this empirical phenomenon is defined within this institutional setting as the "Pesantren Paradox": serving as a descriptive analytical lens to examine how localized religious values interact with emerging digital misconduct risks in a single institutional case. The absence of specific institutional guidelines regarding AI-assisted writing in pesantren-based higher education further deepens this gap, as students potentially commit academic integrity violations without fully realizing it.

This study offers distinct conceptual, empirical, and practical contributions. Conceptually, it provides a refined understanding of how the KAP framework operates when applied specifically to generative AI misconduct rather than traditional plagiarism. Empirically, it yields initial baseline data on AI-giarism prevalence, knowledge deficits, and attitude patterns among undergraduate EFL students within a pesantren-based higher education. Practically, the findings inform university administrators and EFL educators in designing targeted AI literacy curricula and formulating explicit institutional AI policies that align modern digital writing standards with localized moral frameworks. Accordingly, this study aims to examine (1) EFL students' levels of knowledge regarding AI-based plagiarism; (2) their attitudes toward AI-giarism; and (3) their self-reported AI-based plagiarism practices.

Method

Research Design

This study employed a quantitative cross-sectional survey design (Creswell, 2014), which is appropriate for measuring the prevalence, distribution, and associations among constructs, namely knowledge, attitudes, and AI-based plagiarism practices of EFL students at a single

point in time. This design directly adapts and extends the approach used by Phyo et al. (2023) in a similar study, with the addition of the AI-giarism dimension and the institutional context of a pesantren. The research was conducted at the English Education Department of a pesantren-based higher education institution in Probolinggo, East Java, Indonesia, an institution that integrates Islamic ethical values into its academic culture.

Participants

The population comprised all active English Education students at the research institution who had completed at least one academic writing course. This inclusion criterion ensured that all respondents possessed direct experience in completing academic writing assignments relevant to the measured constructs. The sample size was determined using Slovin's formula ($e = 0.05$) through a stratified random sampling technique based on academic cohort (semesters 2, 4, 6, and 8). A total of 99 valid responses were successfully obtained and analyzed, which represents an adequate sample size for binary logistic regression analysis with five predictors (Tabachnick & Fidell, 2013).

Research Instrument

The questionnaire was developed and adapted from Phyo et al. (2023), which drew upon the Attitudes Toward Plagiarism (ATP) Questionnaire by Mavrinac et al. (2010), with substantial modifications tailored to the undergraduate EFL context and the addition of AI-giarism dimensions alongside pesantren values. The instrument consists of 51 items divided into three sections: (1) Knowledge, comprising 17 True/False/Don't Know items with scores ranging from 0 to 17, where correct answers received 1 point; (2) Attitude, comprising 27 5-point Likert scale items ranging from 1 (Strongly Disagree) to 5 (Strongly Agree), featuring reverse scoring on the self-justification subscale (items A8 to A14) with a total score ranging from 27 to 135; and (3) Practice/PSS, comprising 7 5-point Likert scale items adapted from the Plagiarism Severity Scale (PSS) by Phyo et al. (2023), with reverse scoring applied to ethical items (P4, P5, and P7).

Content validity was established through an expert panel consisting of two academic writing lecturers and one measurement specialist ($S-CVI \geq 0.80$). Reliability was tested via a pilot study involving 25 students from an equivalent

institution, yielding results for Section 1 ($\alpha = 0.627$, moderate), Section 2 ($\alpha = 0.861$, excellent, exceeding Nunnally's 1978 threshold of $\alpha \geq 0.70$), and Section 3 ($\alpha = 0.040$, reflecting the heterogeneity of the measured constructs). The questionnaire was distributed via Google Forms, with anonymity fully guaranteed to encourage candid responses on sensitive practice items.

Data Collection

Data were collected between May and June 2026 via Google Forms. The survey link was distributed through semester coordinators and official WhatsApp groups of each academic cohort. A two-week data collection window was set, with a single follow-up reminder sent on Day 7. Informed consent was obtained digitally at the opening page of the questionnaire, which explained the study's purpose, voluntary nature of participation, and guaranteed anonymity. As an incentive, respondents were offered mobile data credit. Responses with incomplete data were excluded, resulting in 99 valid responses for analysis.

Given the sensitive nature of assessing self-reported academic misconduct, comprehensive ethical safeguards were integrated directly into the data collection procedure. Institutional permission to conduct the study was granted by the Head of English Education Program at the participating institution. Informed consent was obtained digitally on the landing page of the questionnaire, which clearly detailed the study's purpose, the voluntary nature of participation, and the unconditional right to withdraw at any point without academic penalty. Explicit digital consent was required to access the survey items. To ensure complete anonymity and eliminate social desirability bias, no personally identifiable information (PII) such as names, student ID numbers, or IP addresses was collected.

Data Analysis

Data were analyzed using IBM SPSS Version 25 through the following stages (Field, 2018): (1) descriptive statistics (mean, SD, and frequency distributions); (2) the Kolmogorov-Smirnov test for normality; (3) Mann-Whitney U and Kruskal-Wallis tests for comparing scores across demographic groups; (4) Chi-square tests to evaluate associations between demographics and the dichotomized PSS (plagiarism vs. non-plagiarism); (5) Spearman's correlation to assess relationships among the three constructs; (6)

multiple linear regression to determine predictors of knowledge and attitude scores; and (7) binary logistic regression to identify predictors of AI plagiarism behavior, yielding Adjusted Odds Ratios (AOR) with 95% confidence intervals (CI). A significance threshold of $p < 0.05$ was applied to all analyses (Phyo et al., 2023). The PSS was dichotomized into Group 0 (never committed AI plagiarism) and Group 1 (committed at least one act of AI plagiarism).

Results and Discussion

1. Demographic Characteristics of Respondents

Of the 99 respondents who met the inclusion criteria, the majority were female (66.7%, $n = 66$). Fourth-semester students constituted the largest group (37.4%, $n=37$), followed by those in semester 2 (30.3%, $n=30$), semester 6 (16.2%, $n=16$), and semester 8 or above (16.2%, $n=16$). In terms of residence, 50.5% resided outside the pesantren, 40.4% lived in the Main Pesantren Complex, and 9.1% resided in Pesantren Satellite Areas. Regarding AI tool utilization, ChatGPT dominated both as a standalone tool (44.4%) and in combination with other applications (28.3%), while Google Gemini was used exclusively by 14.1% of respondents, as presented in Table 1.

Table 1.

Demographic Characteristics of Respondents (n=99)

Characteristic	Category	n	%
Gender	Female	66	66.7%
	Male	33	33.3%
Semester	2	30	30.3%
	4	37	37.4%
	6	16	16.2%
	8+	16	16.2%
Residence	Outside Pesantren	50	50.5%
	Main Pesantren Complex	40	40.4%
	Satellite Areas	9	9.1%
Primary AI Tool	ChatGPT	44	44.4%
	Gemini	14	14.1%
	Combination	41	41.4%

2. Instrument Reliability

Internal reliability testing was conducted using Cronbach's alpha coefficient. The results indicated that Section 2 (Attitudes) exhibited very high reliability ($\alpha=0.861$), exceeding Nunnally's (1978) standard and comparable to the original instrument by Phyo et al. (2023). Section 1 (Knowledge) yielded $\alpha=0.627$, which is considered acceptable given that knowledge items measure distinct factual dimensions. Section 3 (Practices) produced a low internal consistency ($\alpha=0.040$). This result does not stem from instrument instability, but rather reflects the structural heterogeneity of the measured constructs. The PSS items encompass a broad spectrum of non-linear behaviors, spanning from severe ethical infractions (e.g., text generation, reference fabrication) to appropriate practices (e.g., seeking instructor permission, grammar editing). Because engaging in one practice does not logically imply engaging in another, these items do not form a unidimensional latent scale. Consequently, Cronbach's alpha inherently understates the instrument's utility. To mitigate this measurement limitation, the PSS was subsequently dichotomized for regression analyses, though correlations involving the continuous PSS composite score should still be interpreted cautiously. **Table 2** summarizes the results of the reliability testing.

Table 2.

Instrument Reliability Test Result (Cronbach's Alpha)

Section	Construct	Number of Items	Cronbach's α	Remark
Section 1	Knowledge	17	0.627	Acceptable
Section 2	Attitude	27	0.861	Very Good
Section 3	Practice (PSS)	7	0.040	Low*

3. Knowledge Level Regarding AI-Based Plagiarism

The mean knowledge score was 9.63 out of a maximum of 17 (56.6%, $SD = 3.09$), indicating a moderate category, as presented in **Table 3**. This aligns with Magida et al. (2023) regarding undergraduate students in South Africa, yet is lower than Phyo et al. (2023), who reported 72.6% among postgraduate students with formal research ethics training, a reasonable difference given that this instrument intentionally targets AI-giarism as a novel construct. The normality test revealed that the knowledge score data were

not normally distributed ($KS\ p = 0.000$); hence, comparative analyses were conducted using non-parametric tests.

Table 3.

Knowledge Scores by Sub-Dimension

Dim.	Knowledge Area	Mean	%Max
A	Basic Understanding of AI (K1-K6)	3.63	60.5%
B	Ethical Boundaries and Source (K7-K12)	3.24	54.0%
C	Institutional and Technical Rules (K13-K17)	2.76	55.2%

Sub-dimension A (Basic Understanding of AI) achieved the highest attainment rate (60.5%), indicating that students are relatively more familiar with concrete and overt AI-giarism scenarios. Conversely, Sub-dimension B (Ethical Boundaries) recorded the lowest attainment rate (54.0%), revealing a weak understanding regarding AI attribution obligations and intellectual ownership of machine-generated content.

The most critical finding emerged from item K6, where only 33.3% of respondents understood that AI-generated text passing Turnitin still constitutes plagiarism. This implies that two-thirds of the students still operate under a dangerous logic of equivalence: "*passing Turnitin equals non-plagiarism*." Chan (2025) conceptualizes this as technological false legitimacy, wherein the absence of algorithmically detected violations is misinterpreted as an absence of ethical infractions. This condition creates a severe structural loophole: as long as students believe that integrity standards can be fully delegated to technological detection tools, any attempt to reform behavior without first correcting this misconception will reach a dead end. Items K15 (34.3%, regarding AI-driven language style modifications) and K16 (36.4%, regarding AI-driven research data manipulation) also demonstrated profound confusion regarding the boundary between neutral technical AI assistance and substantive integrity violations.

For item K15 (34.3%), the high error rate was driven by the gray area between legitimate language editing (language polishing) and writing ethics violations. Most students assumed that adjusting the writing tone from casual to academic using AI constitutes plagiarism, even though the underlying text was their own original work. This reflects an over-perception or misplaced concern (AI anxiety), wherein students fail to clearly distinguish between technical tools for stylistic enhancement and the misappropriation of substantive ideas.

Meanwhile, for item K16 (36.4%), the majority of students failed to recognize that manipulating interview data to appear cleaner or more aesthetically pleasing using AI constitutes a serious scientific breach of ethics (data fabrication/falsification). This confusion stems from the assumption that AI functions merely as a formatting or narrative-structuring tool, without realizing that AI intervention in raw research data can alter factual substance and compromise scientific data integrity. Overall, the low performance on both items underscores the urgent need for in-depth education regarding the ethical boundaries of Generative AI usage in academic writing.

From the logistic regression model, knowledge score was the only significant independent predictor of AI plagiarism behavior (AOR=0.764; 95% CI [0.62–0.94]; $p=0.012$), where each one-point increase in knowledge score decreased the odds of engaging in AI-giarism by 23.6%. This finding contrasts with Roig (2015) and Phyo et al. (2023), who identified attitude as the sole predictor in the context of conventional plagiarism. This discrepancy reflects the unique nature of AI-giarism as a cognitively more ambiguous form of misconduct: students do not necessarily violate rules out of apathy, but rather because they genuinely do not know that their actions constitute a violation. This serves as the empirical essence of the title of this study. Joshi et al. (2025) and Hutson (2024) corroborate that the most effective interventions for AI-giarism must prioritize raising conceptual literacy, rather than merely updating regulations or deploying detection technologies.

4. Students' Attitudes Toward AI-Based Plagiarism

The mean attitude score was 91.24 out of a maximum of 135 (67.6%, $SD=12.95$), reflecting a moderate rejection of plagiarism. The

Kolmogorov-Smirnov test yielded $p=0.200$, indicating a near-normal distribution. The proportionality of this score (67.6%) is identical to that reported by Phyo et al. (2023) in Myanmar, suggesting that moral ambivalence toward plagiarism represents a consistent phenomenon across different cultural contexts and educational levels.

Table 4.

Attitude Scores per Sub-scale (n=99)

Sub-scale Code	Sub-scale	Mean	SD	% Max
A	Clash with Pesantren Values	24.96	5.32	71.3%
B	Self-Justification (reversed)	19.41	5.57	55.5%
C	Perception of AI Function	25.93	5.29	74.1%
D	Environment and Policy	20.94	4.07	69.8%

In Table 4, Subscale C (Perception of AI Function: 74.1%) recorded the highest score, indicating that students generally recognize that AI should serve as an auxiliary tool rather than a substitute writer, while also understanding the negative impacts of AI dependency on critical thinking skills. Conversely, Subscale B (Self-Justification: 55.5% after reverse scoring) was the lowest, reflecting a tendency to rationalize AI usage based on deadline pressures, limitations stemming from EFL writing anxiety, and the absence of written regulations. Memon (2020), Saeed et al. (2011), Sozon et al. (2025) and Prashar et al. (2024) confirm that time constraints and language limitations represent the two most consistent factors driving the rationalization of plagiarism in developing countries.

Subscale A (Clash with Pesantren Values: 71.3%) confirms that the values of *shiddiq* (truthfulness) and *amanah* (trustworthiness) act as tangible moral brakes in students' perceptions. Nevertheless, 50.5% of students still reported engaging in AI plagiarism, representing the "Pesantren Paradox" that constitutes the novel contribution of this study. This paradox can be explained through three mutually reinforcing mechanisms: (1) the absence of an operational translation of pesantren values into concrete regulations regarding AI use; (2) intense pressure from EFL writing anxiety that drives the adoption

of AI as a quick fix (Hyland, 2026); and (3) the social normalization of AI usage within the campus environment, where individual moral values face difficult conformity pressure when peers engage in similar behavior. Ajzen (1991), through the Theory of Planned Behavior, posits that under high academic pressure, supportive subjective norms, and low perceived behavioral control, a nominally anti-plagiarism attitude is insufficient to prevent the behavior. Gender emerged as a significant predictor of attitude scores in the linear regression model ($B=5.586$; $p=0.032$; $R^2=0.239$), with female students demonstrating higher plagiarism rejection scores, a finding consistent with moral socialization literature identifying stronger honesty orientations among females (Bandura, 1997).

5. Self-Reported AI-Based Plagiarism Practices

The mean PSS score was 19.95 out of a maximum of 35 ($SD=2.93$). The Kolmogorov-Smirnov test yielded $p=0.092$ (>0.05), indicating a near-normal distribution. When the PSS was dichotomized, 50.5% ($n=50$) of respondents reported having committed at least one act of AI-giarism. This rate exceeds the conventional plagiarism prevalence reported by Phyo et al. (2023) at 37.6%, and should be interpreted as a lower bound of the actual prevalence, considering the under-reporting bias inherent in self-reported surveys on sensitive behaviors.

Table 5.

Response Distribution Per Practice Item (n=99)

Item	Statement	N%	S%	O%	A%
P1	Copy-pasting ChatGPT into essays	20.2	9.4	12.1	2.0
P2	Paraphrasing AI to bypass Turnitin Assignments	21.2	33.3	11.1	6.1
P3	100% produced by AI	18.2	41.4	12.1	1.0
P4	Citing AI in the references	43.4	24.2	9.1	5.1
P5	Using AI solely for grammar checking	4.0	44.4	20.2	8.1

P6	Fabricating references via AI	40.4	27.3	11.1	1.0
P7	Asking lecturer's permission before using AI	25.3	27.3	15.2	6.1

**ethical items (reverse scoring); N=Never, S=Sometimes, O=Often, A=Always*

In Table 5, item P3 (submitting assignments that are 100% AI-generated: 54.5% reporting "sometimes" to "always") emerged as the most commonly reported behavior, indicating that the total substitution of the writing process by AI has become a relatively prevalent practice. Item P2 (using AI paraphrasing to evade Turnitin: 50.5% reporting "sometimes" to "always") is equally alarming, as it no longer represents mere AI utilization, but rather the active employment of AI as an instrument to conceal conventional plagiarism. Acosta-Enriquez et al. (2024) conceptualize this as "plagiarism laundering," which constitutes one of the most severe and elusive forms of AI-giarism.

The discrepancy between P3 (54.5% submitting AI-generated assignments) and P4 (only 38.4% acknowledging AI in their reference lists) serves as an empirical measure of the transparency gap, wherein students utilize AI without disclosing it to their instructors. This pattern mirrors the primary characteristic of AI-giarism identified by Naidu and Sevnanarayan (2023). Furthermore, item P6 (fabricating references via AI: 39.4% reporting "sometimes" to "always") presents additional concerns given its broader implications for the credibility of academic literature.

The Chi-square test revealed a paradoxical finding: residence was the only variable significantly associated with AI plagiarism behavior ($\chi^2=7.115$; $df=2$; $p=0.029$). Students residing in Satellite Areas exhibited the highest proportion of plagiarism (88.9%), substantially exceeding those in the Main Pesantren Complex (40.0%) and Outside the Pesantren (52.0%). This finding challenges the conventional assumption that the propensity for academic misconduct increases linearly with physical distance from the pesantren environment. Instead, the data reveal a non-linear pattern that highlights the heightened vulnerability of a transitional or liminal zone: Institutional Ambiguity, the Satellite Area operates as an intermediate space, remaining under the umbrella of pesantren culture while

physically detached from centralized supervision. Consequently, students in this zone experience double-bind pressure: they face heavy academic workloads alongside restricted time and literature access characteristic of traditional students, yet lack both the dense social surveillance of the core complex and independent digital ethics education. This combination of a governance void and limited AI ethics understanding transforms the Satellite Area into a high-risk hotspot for pragmatic and instant AI usage. Conversely, Socio-Institutional Inhibition in the Core Complex, students residing in the Main Pesantren Complex exhibited the lowest plagiarism prevalence (40.0%). This suggests that direct socio-institutional surveillance and communal oversight serve as more effective inhibitors of unethical behavior than reliance on individual self-regulation.

This outcome confirm previous research by Zhai et al. (2024) regarding the socio-institutional environment exerts a far stronger deterministic influence over digital ethical behavior than personally held values.

6. Relationship Among Constructs and Behavior Predictors

Spearman's correlation revealed a theoretically critical pattern: knowledge correlated strongly with attitude ($r=0.559$; $p<0.001$), attitude correlated significantly and negatively with the PSS ($r = -0.239$; $p = 0.017$), but knowledge showed no direct significant correlation with the PSS ($r = -0.103$; $p = 0.310$).

This three-construct pattern confirms the foundational premise of the Theory of Planned Behavior (Ajzen, 1991), by demonstrating that attitude operates as a key mediator between knowledge and behavior. More importantly, it extends TPB's applicability to digital academic misconduct, illustrating that while AI literacy directly informs student attitudes, these attitudes subsequently shape self-reported AI-giarism practices. TPB's applicability to digital academic misconduct, illustrating that while AI literacy directly informs student attitudes, these attitudes subsequently shape self-reported AI-giarism practices."which posits that knowledge influences behavior through attitude formation rather than through a direct pathway.

Table 6.

Binary Logistic Regression: Predictors of Dichotomous PSS

Predictor	B	SE	p	AOR (95% CI)
Knowledge Score	-0.270	0.107	0.012	0.764 [0.62-0.94]
Attitude Score	-0.036	0.021	0.084	0.964 [0.93-1.01]
Gender	-0.001	0.512	0.998	0.999 [0.91-2.34]
Semester	0.380	0.240	0.113	1.462 [0.91-2.34]
Residence	0.304	0.365	0.405	1.355 [0.66-2.77]

* $p<0.05$; $Model\chi^2 = 22.864$; $df= 5$; $p<0.001$; Nagelkerke $R^2 = 0.167$

In Table 6, the logistic regression model demonstrates that knowledge score was the only significant independent predictor of AI plagiarism behavior (AOR=0.764; 95% CI [0.62–0.94]; $p=0.012$). Although the attitude score exhibited the expected direction (AOR=0.964) and approached statistical significance ($p=0.084$), it did not cross the $p<0.05$ threshold, likely due to substantial multicollinearity between knowledge and attitude ($r=0.559$). None of the demographic variables remained significant in this multivariable model, confirming that the effect of residence observed in the Chi-square test did not persist when controlled alongside the primary constructs.

The results of the multiple linear regression complement this picture: knowledge score ($B=0.112$; $p<0.001$) and gender ($B=-1.582$; $p=0.009$) served as significant predictors of knowledge scores ($R^2=0.266$); whereas knowledge score ($B=2.040$; $p<0.001$) and gender ($B=5.586$; $p=0.032$) functioned as significant predictors of attitude scores ($R^2=0.239$). Both models indicate that knowledge and attitude reciprocally predict one another, aligning with the assumptions of the KAP framework that the three constructs are mutually interdependent (Launiala, 2009). The practical implication is that interventions successfully improving knowledge regarding AI-giarism will likely exert a spillover effect on enhancing attitudes as well, and vice versa, thereby supporting an integrated academic integrity education approach.

Conclusion

This study reveals three primary, interrelated findings regarding AI-based plagiarism among undergraduate EFL students at the research institution. First, the overall knowledge level fell into the moderate category (56.6%), with the most severe structural

deficiency stemming from the misconception that passing Turnitin is equivalent to being free of plagiarism (33.3% correct), representing a cognitive deficit empirically associated with AI-giarism behavior (AOR = 0.764; $p = 0.012$). Second, the moderate attitude of rejecting plagiarism (67.6%) is trapped within the "Pesantren Paradox," where religious moral values act as tangible moral brakes (71.3%) yet fail to effectively prevent rationalization driven by academic pressure and the absence of clear regulations (self-justification subscale: 55.5%). Third, half of the respondents (50.5%) reported having engaged in AI-giarism, with residence being the only significant demographic predictor, as students situated furthest from the supervision of the main pesantren complex exhibited the highest prevalence rate (88.9%).

Several limitations of this study should be acknowledged. First, the collected sample size ($n = 99$) fell short of the initial target ($n = 150$) due to time constraints during data collection; nevertheless, this number still fulfills the minimum requirement for binary logistic regression analysis with five predictors (minimum of 10 cases per predictor = 50), thereby maintaining the statistical validity of the findings (Tabachnick & Fidell, 2013). Second, the low Cronbach's alpha coefficient for Section 3 (Practice, $\alpha = 0.040$) does not reflect instrument failure, but rather stems from the underlying construct heterogeneity, as this subscale intentionally encompasses a spectrum of diametrically opposed behaviors ranging from severe integrity violations (copy-pasting AI text, fabricating references) to fully ethical practices (seeking instructor permission, using AI solely for grammar checking), making high internal consistency unexpected. A similar limitation was acknowledged by Phyo et al. (2023) regarding the original PSS instrument. Third, the cross-sectional design precludes causal inferences, meaning the identified relationships remain

associative. Fourth, self-reported data are inherently susceptible to social desirability bias; thus, the 50.5% prevalence rate likely represents a lower bound of the actual figure.

Practically, these findings carry direct pedagogical implications for both course instructors and university administrators. Academic Writing lecturers must pivot from passive surveillance toward proactive AI literacy, explicitly integrating technical sessions that dissect the boundaries of ethical AI use, the fallibility of Turnitin, and proper citation protocols for generative tools directly into the curriculum. Concurrently, the institution must establish a formal AI Academic Policy that bridges traditional pesantren values (akhlakul karimah) with clear institutional guidelines. Rather than relying solely on abstract moral appeals, institutional governance must deploy targeted academic writing support and structured digital monitoring which particularly targeting satellite campuses and high-pressure academic periods to effectively eliminate the rationalization of AI-giarism.

The most theoretically significant finding is that in the context of AI-giarism, knowledge, rather than attitude, serves as a more direct and significant predictor of behavior, contrasting with conventional plagiarism dynamics (Phyo et al., 2023). This implies that the most effective interventions must prioritize explicitly enhancing AI-giarism literacy within academic writing instruction, accompanied by the development of an AI Academic Policy that integrates pesantren moral values into a modern academic integrity framework. Future research should explore these causal relationships through longitudinal designs, expand the sample across other pesantren-affiliated institutions to evaluate the generalizability of the Pesantren Paradox, and design pedagogical interventions based on these empirical findings.

References

- Acosta-Enriquez, B. G., Arbulú Ballesteros, M. A., Arbulu Perez Vargas, C. G., Orellana Ulloa, M. N., Gutiérrez Ulloa, C. R., Pizarro Romero, J. M., Gutiérrez Jaramillo, N. D., Cuenca Orellana, H. U., Ayala Anzoátegui, D. X., & López Roca, C. (2024). Knowledge, attitudes, and perceived Ethics regarding the use of ChatGPT among generation Z university students. *International Journal for Educational Integrity*, 20(1), 10. <https://doi.org/10.1007/s40979-024-00157-4>
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211.

- [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Aminah, M., & Aly, A. H. (2026). University EFL Students' Perceptions of Generative AI in Academic Writing: A Qualitative Case Study in Indonesia. *AL-ISHLAH: Jurnal Pendidikan*, 18(1). <https://doi.org/10.35445/alishlah.v18i1.9461>
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H. Freeman.
- Chan, C. K. Y. (2025). Students' perceptions of 'AI-giarism': Investigating changes in understandings of academic misconduct. *Education and Information Technologies*, 30(6), 8087–8108. <https://doi.org/10.1007/s10639-024-13151-7>
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed). SAGE Publications.
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th edition). SAGE.
- Hutson, J. (2024). Rethinking Plagiarism in the Era of Generative AI. *Journal of Intelligent Communication*, 3(2). <https://doi.org/10.54963/jic.v3i2.220>
- Hyland, K. (2026). Writing in the AI era: Rethinking writing, research and teaching. *Journal of Second Language Writing*, 101302. <https://doi.org/10.1016/j.jslw.2026.101302>
- Joshi, K., Singhal, S., & Tandon, A. (2025). Plagiarism in the Age of Artificial Intelligence: A Call for Ethical Vigilance. *NMO Journal*, 19(1), 1–2. https://doi.org/10.4103/JNMO.JNMO_25_25
- KOMPAS. (2026). *Kronologi Terbongkarnya Dugaan Skandal Riset Palsu di Denmark yang Seret Alumni UNY*. <https://www.kompas.com/tren/read/2026/05/28/131500765/kronologi-terbongkarnya-dugaan-skandal-riset-palsu-di-denmark-yang-seret>
- Launiala, A. (2009). How much can a KAP survey tell us about people's knowledge, attitudes and practices? Some observations from medical anthropology research on malaria in pregnancy in Malawi. *Anthropology Matters*, 11(1). <https://doi.org/10.22582/am.v11i1.31>
- Lund, B., Mannuru, N. R., Teel, Z. A., Lee, T. H., Ortega, N. J., Simmons, S., & Ward, E. (2025). Student perceptions of AI-assisted writing and academic integrity: Ethical concerns, academic misconduct, and use of generative AI in Higher education. *AI in Education*, 1(1), 2. <https://doi.org/10.3390/aieduc1010002>
- Magida, N., Cross, K., Baloyi, K., Mitrovich, T., Muyengwa, L., Mupambirei, P., Van Der Mescht, M., & Graham, M. (2023). Knowledge, attitudes, and practices of plagiarism by undergraduate healthcare science in Gauteng. *Journal of University Teaching and Learning Practice*, 20(6). <https://doi.org/10.53761/1.20.6.4>
- Malik, A. R., Pratiwi, Y., Andajani, K., Numertayasa, I. W., Suharti, S., Darwis, A., & Marzuki. (2023). Exploring artificial intelligence in academic essay: Higher education student's perspective. *International Journal of Educational Research Open*, 5, 100296. <https://doi.org/10.1016/j.ijedro.2023.100296>
- Marzuki, Widiati, U., Rusdin, D., Darwin, & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education*, 10(2), 2236469. <https://doi.org/10.1080/2331186X.2023.2236469>
- Mavrinac, M., Brumini, G., Bilić-Zulle, L., & Petrovečki, M. (2010). Construction and Validation of Attitudes Toward Plagiarism Questionnaire. *Croatian Medical Journal*, 51(3), 195–201. <https://doi.org/10.3325/cmj.2010.51.195>
- Memon, A. R. (2020). Similarity and plagiarism in scholarly journal submissions: Bringing clarity to the concept for authors, reviewers and editors. *Journal of Korean Medical Science*, 35(27), e217. <https://doi.org/10.3346/jkms.2020.35.e217>
- Naidu, K., & Sevnarayan, K. (2023). ChatGPT: An ever-increasing encroachment of artificial intelligence in online assessment in distance education.

- Online Journal of Communication and Media Technologies*, 13(3), e202336.
<https://doi.org/10.30935/ojcmnt/13291>
- Nilufar, M. (2026). Academic plagiarism in higher education causes consequences and prevention strategies. *International Journal of Pedagogics*, 06(02), 237–239.
<https://doi.org/10.37547/ijp/Volume06Issue02-55>
- Nunnally, J. C. (1978). *Psychometric theory* (2. ed). McGraw-Hill.
- Pahlevi, R., Thamrin, S., Ahmad, I., & Setiani, R. (2024). Plagiarism in higher education and impacts. *International Journal of Humanities Education and Social Sciences (IJHESS)*, 4(2).
<https://doi.org/10.55227/ijhess.v4i2.1132>
- Phyo, E. M., Lwin, T., Tun, H. P., Oo, Z. Z., Mya, K. S., & Silverman, H. (2023). Knowledge, attitudes, and practices regarding plagiarism of postgraduate students in Myanmar. *Accountability in Research*, 30(8), 672–691.
<https://doi.org/10.1080/08989621.2022.2077643>
- Prashar, A., Gupta, P., & Dwivedi, Y. K. (2024). Plagiarism awareness efforts, students' ethical judgment and behaviors: A longitudinal experiment study on ethical nuances of plagiarism in higher education. *Studies in Higher Education*, 49(6), 929–955.
<https://doi.org/10.1080/03075079.2023.2253835>
- Roig, M. (2015). *Avoiding plagiarism, self-plagiarism, and other questionable writing practices: A guide to ethical writing*.
- Saeed, S., Aamir, R., & Ramzan, M. (2011). Plagiarism and its implications on higher education in developing countries. *International Journal of Teaching and Case Studies*, 3(2/3/4), 123.
<https://doi.org/10.1504/IJTCS.2011.039552>
- Sozon, M., Pok, W. F., Sia, B. C., & Alkharabsheh, O. H. M. (2025). Cheating and plagiarism in higher education: A systematic literature review from a global perspective, 2016–2024. *Journal of Applied Research in Higher Education*, 17(5), 1728–1742.
<https://doi.org/10.1108/JARHE-12-2023-0558>
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6. ed., international ed). Pearson.
- Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, 11(1), 28.
<https://doi.org/10.1186/s40561-024-00316-7>
- Zhao, X. (2023). Leveraging artificial intelligence (AI) technology for english writing: Introducing wordtune as a digital writing assistant for EFL writers. *RELC Journal*, 54(3), 890–894.
<https://doi.org/10.1177/00336882221094089>