Journal of Applied Data Sciences
Vol. 7, No. 1, January 2026, pp. 786-801

ISSN 2723-6471
786

Analyzing Student Sentiments and Insights on Generative AI for Independent Learning in Universities

Ni Made Satvika Iswari[1,*], I Nyoman Yudi Anggara Wijaya[2], Ni Putu Widya Yuniari[3]

[1,2]Faculty of Information Technology and Design, Universitas Primakara, Denpasar 80226, Indonesia

[3]Faculty of Engineering and Planning, Universitas Warmadewa, Denpasar 80239, Indonesia

(Received: September 1, 2025; Revised: October 20, 2025; Accepted: January 12, 2026; Available online: February 17, 2026)

Abstract

Transformations in higher education brought about by Generative AI have significantly changed how university students' access, comprehend, and develop learning materials. This study explores Indonesian university students' perceptions and experiences regarding the use of Generative AI for independent learning, employing qualitative surveys together with sentiment analysis powered by machine learning. Data were collected from open-ended questionnaires and analyzed using four key algorithms, such as Naive Bayes, Logistic Regression, Random Forest, and Linear SVM, to classify student sentiments towards generative AI technologies. These four classical machine learning models were employed as baseline algorithms commonly used in sentiment analysis to benchmark performance on small, imbalanced educational datasets before applying more complex transformer-based methods. In addition to quantitative analysis, this study also implements thematic analysis of open-ended responses to identify prominent issues, challenges, and student recommendations concerning the use of generative AI in learning. Evaluation results identified Linear SVM as the most consistent model, with the highest weighted F1-score (0.63), although all models showed limitations in detecting negative sentiment due to class imbalance (only three negative samples out of forty responses), which affected model generalization. Key findings indicate that students perceive Generative AI as a supportive tool that accelerates understanding, creativity, and reference searching; however, they remain wary of risks related to dependency, reduced originality, and academic integrity dilemmas. This article recommends the implementation of ethical policy, AI digital literacy training, and enhancement of campus infrastructure to ensure that AI technologies enrich the learning process without compromising student independence and integrity.

Keywords: Multimodal Generative AI, Independent Learning, Sentiment Analysis, Thematic Analysis, Higher Education

1. Introduction

The rapid advancement of generative Artificial Intelligence (AI) has profoundly transformed the landscape of higher education, making AI-powered tools like ChatGPT and similar platforms integral to modern academic activities. In 2025, more than 86% of university students worldwide reported using generative AI in their studies, reflecting a dramatic increase in adoption compared to previous years [1], [2]. These technologies have enabled students to access information instantly, clarify complex concepts, summarize academic literature, and generate ideas for research, fostering a more autonomous and efficient learning environment [3]. By acting as facilitators, generative AI systems empower students to pursue personalized study routines and tailor their academic exploration according to individual needs and interests [2], [4].

Despite the evident benefits, the widespread integration of generative AI into higher education brings significant challenges and risks. One prominent concern is students' growing dependence on AI-powered tools, which can undermine critical thinking, diminish originality, and blur the boundaries of academic integrity [5]. The direct incorporation of AI-generated text in assignments highlights the necessity for clear ethical guidelines and responsible use frameworks across institutions [4]. Furthermore, frequent issues with accuracy, hallucinations, and misinterpretation of prompts may hinder the development of analytical skills and reduce the quality of student learning

*Corresponding author: Ni Made Satvika Iswari (iswari@primakara.ac.id)
DOI: https://doi.org/10.47738/jads.v7i1.1083

outcomes [6], [7]. Ethical implications such as data privacy, algorithmic bias, and educational inequality also pose significant challenges, emphasizing the importance of robust digital literacy programs and institutional support [4], [7], [8].

This study employed a structured methodology to explore university student perceptions and experiences with generative AI in independent learning, beginning with open-ended questionnaire data collection, followed by sentiment labeling using TextBlob and manual methods, preprocessing (including TF-IDF feature extraction), splitting data for training and testing, and training four machine learning models (Naive Bayes, Logistic Regression, Random Forest, Linear SVM), whose performance was compared using standard evaluation metrics; finally, thematic analysis was conducted to identify key challenges and suggestions, which were then interpreted through established learning theories to inform strategic recommendations for responsible and effective generative AI integration in higher education.

The contributions of this study are multifaceted: it provides a comprehensive evaluation of sentiment analysis techniques using multiple machine learning models on actual student feedback about generative AI, highlights the strengths and weaknesses of each approach in class-imbalanced, real-world educational data, and offers a robust thematic mapping of student-identified challenges and suggestions for AI use in independent learning. By grounding thematic interpretations in contemporary learning theories and connecting model performance review to actionable campus-level recommendations, this research not only extends technical knowledge in educational NLP but also informs institutional policy, digital literacy development, and the ethical adoption of generative AI for more effective, equitable, and personalized learning in higher education.

2. Literature Review

2.1. The Growth of Generative AI in Higher Education

Recent developments in generative artificial intelligence, including large language models like ChatGPT, Gemini, and Claude, have transformed academic activities across universities worldwide. A 2025 global survey reported that over 86% of university students now use generative AI tools to support their studies, including accessing information, clarifying complex concepts, summarizing academic papers, and generating new ideas. Systematic reviews confirm that generative AI is driving personalized learning, producing interactive content, and enabling adaptive assessment, which in turn benefit student motivation and learning efficiency[9], [10].

2.2. Sentiment Analysis and Generative AI

Sentiment analysis powered by machine learning has become an essential tool for understanding student perceptions of generative AI adoption in higher education. This research uses several established algorithms for sentiment classification, including Naive Bayes, Logistic Regression, Random Forest, and Linear SVM, each of which offers distinct advantages. Naive Bayes is recognized for its speed and suitability for text classification with large, sparse datasets, while Logistic Regression provides reliable probabilistic interpretations and acts as a strong baseline model. Random Forest, as an ensemble classifier, helps address nonlinearities and imbalanced data commonly found in student feedback. Linear SVM is proven effective in handling high-dimensional text features and routinely outperforms other models in comparable studies. The choice of these algorithms follows patterns seen in contemporary research, where aspect-based sentiment analysis and data augmentation, using both classical machine learning and transformer-based models, are applied to reveal emotional subtleties and educational expectations in User-Generated Content (UGC) [11].

Several foundational studies have explored semantic similarity and multidimensional sentiment categorization in visitor reviews using SBERT and cosine similarity, highlighting the importance of separating review aspects for accuracy and showing that aspect-based approaches can be used to analyse specific categories relevant to generative AI's impact on student creativity and time management [12]. Other studies have mapped global UGC research trends with bibliometric analysis, emphasizing the need for more local studies in Indonesia and supporting the urgency of research on generative AI in universities [13]. Additional research has applied transformer-based models such as BERT and GPT2 for customer sentiment analysis, achieving high accuracy and demonstrating the value of transformer models in understanding complex sentiments [14]. Furthermore, recent work has shown the capability of GPT-based models to perform both sentiment analysis and rating prediction in a commercial context [15]. Collectively, these studies provide a methodological foundation for evaluating the effects of generative AI on student autonomy and learning outcomes. By adapting techniques such as semantic similarity, aspect-based sentiment analysis, and leveraging both classical and transformer models, this research aims to produce a comprehensive picture of how generative AI influences student independence, thereby informing data-driven policies for higher education institutions.

2.3. Benefit and Challenges of Generative AI Integration

Multiple contemporary studies indicate that the use of generative AI in education offers several important benefits, including faster learning processes, greater learner independence, and richer access to adaptive and interactive learning resources. Nevertheless, significant challenges remain, particularly the risk of overreliance on AI, which can weaken students' critical thinking skills and their ability to learn autonomously [4], [16]. In addition, the validity of AI-generated content continues to be a concern due to the potential for hallucinations and data bias, requiring students to further strengthen their information literacy skills [4]. Issues of academic integrity also remain prominent, encompassing concerns about plagiarism, originality, and the ethical use of AI in assignments and assessments [4]. Furthermore, disparities in access to AI technologies persist, especially in developing regions, underscoring the need for inclusive education policies to ensure that the adoption of AI does not exacerbate existing digital divides [17]. Recent literature strongly recommends robust institutional guidelines and ethical frameworks to ensure fair, responsible, and effective use of AI in education. Emerging priorities include digital literacy education, training in prompt engineering, and conscious integration of AI into curriculum design and instructional strategies.

3. Methodology

The research methodology, as shown in figure 1, employed in this study follows a systematic process to ensure a comprehensive analysis of student perceptions and experiences regarding the use of generative AI in learning. The data collection phase begins with gathering an open-ended questionnaire dataset designed to capture students' insights and reflections. Following this, a data labelling process is conducted to assign sentiment categories, such as positive, negative, or neutral, to the textual responses, either through automated tools such as TextBlob or by applying manual and semi-automated approaches to enhance labelling accuracy.

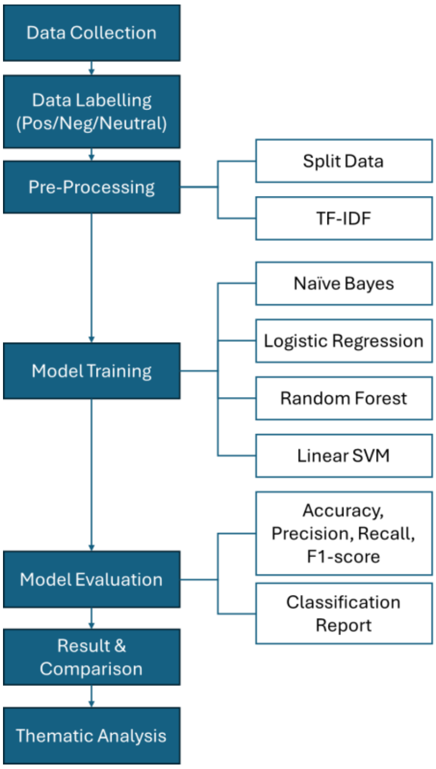

Figure 1. Research Methodology

ISSN 2723-6471

Initial polarity scores were generated with TextBlob, followed by independent verification by two annotators. Both reviewers applied predefined criteria for positive, neutral, and negative sentiments, then discussed any disagreements until full consensus was reached. This procedure was intended to improve the accuracy and reliability of sentiment labels, particularly in context-sensitive responses. TextBlob was employed as a lexicon-based baseline tool to generate preliminary sentiment labels, suitable for small datasets in exploratory studies. Its outputs were then refined through manual verification to enhance accuracy. This hybrid approach provided a transparent and efficient labelling process without the computational demands of deep-learning models. TextBlob was used only as an initial lexicon-based tool to generate baseline polarity scores, after which all labels were independently reviewed and refined by two annotators using predefined sentiment criteria. This hybrid approach ensured labeling transparency while maintaining reliability despite the small dataset size. More advanced tools such as VADER or transformer-based sentiment classifiers were not applied at this stage due to dataset constraints, but they will be incorporated in future studies for comparative labeling once a larger corpus becomes available.

Prior to feature extraction, a structured text-cleaning pipeline was implemented. All responses were lowercased to ensure case-insensitive token matching, followed by the removal of punctuation, numerical characters, and non-alphabetic tokens using regular expression filters. Indonesian stopwords were removed using the default stopword list from widely used NLP libraries to reduce high-frequency, low-information terms. Stemming was deliberately excluded after preliminary inspection showed that morphological reduction (e.g., stemming “membantu” to “bantu”) introduced semantic distortion in short qualitative responses. The final cleaned text ensured consistent tokenization and reliable TF-IDF vector generation for downstream model training.

Before model training, data preprocessing is carried out to organize the dataset effectively. This step includes identifying the relevant columns (text and sentiment label), splitting the dataset into training (80%) and testing (20%) subsets, and extracting textual features using the TF-IDF vectorization technique to convert raw text into numerical representations suitable for machine learning models. TF-IDF was adopted as the feature extraction method in this study due to its suitability for small, sparse text datasets and its compatibility with classical machine learning models. Alternative embedding approaches such as Word2Vec, GloVe, and contextual models like BERT were not employed because these methods generally require larger and more diverse corpora to generate stable semantic representations. To ensure methodological consistency and avoid overfitting on the limited dataset, the analysis was restricted to TF-IDF, with plans to incorporate more advanced embeddings in future work as the dataset expands.

Subsequently, four different machine learning algorithms are trained for sentiment classification, namely Naive Bayes, Logistic Regression, Random Forest, and Linear Support Vector Machine (SVM). To evaluate the performance of these models, predictions are generated using the testing dataset. The evaluation process involves calculating various performance metrics, including accuracy, weighted precision, weighted recall, and weighted F1-score. Additionally, a classification report is provided to assess model performance across each sentiment class in detail. The results are then compared, and the performance of all models is summarized in a comparative table to identify the most effective algorithm.

In this study, four classical machine learning algorithms, such as Naive Bayes, Logistic Regression, Random Forest, and Linear SVM, were employed as the primary classification models. These algorithms were intentionally selected due to their suitability for small to medium-sized datasets, their interpretability, and their strong performance in prior educational NLP studies. Given that the current corpus contains a limited amount of student feedback and presents class imbalance across sentiment categories, transformer-based methods such as BERT were not adopted at this stage. Transformer architectures typically require large, diverse datasets and substantial computational resources to achieve optimal generalization. Therefore, classical models offer a more stable and methodologically appropriate baseline for benchmarking performance in the context of this preliminary investigation.

For all four classical machine learning models, default or minimally adjusted hyperparameters were used to maintain a consistent baseline for comparison. Extensive hyperparameter tuning procedures such as grid search or randomized search were not applied, as the very small dataset (40 samples) increases the risk of overfitting and leads to unstable, non-generalizable parameter configurations. This baseline-focused approach ensured methodological consistency

ISSN 2723-6471

across models, with plans to incorporate systematic optimization techniques in future work once a larger dataset is available.

Beyond sentiment classification, a thematic analysis is performed to uncover recurring themes related to students' challenges and suggestions concerning the use of generative AI in education. These thematic findings are subsequently interpreted in relation to established learning theories to provide a deeper understanding of how generative AI shapes learning behaviours. Finally, the analysis leads to strategic recommendations aimed at enhancing independent learning practices in higher education contexts.

Unsupervised theme extraction techniques such as Latent Dirichlet Allocation (LDA) and Non-negative Matrix Factorization (NMF) were not applied in this study due to the limited size and structure of the textual corpus. The dataset consists of short, heterogeneous student responses that provide insufficient lexical density for stable topic generation, leading to overlapping or semantically incoherent topics during preliminary testing. To ensure the reliability of thematic insights, theme identification was therefore conducted manually through qualitative coding, with plans to incorporate computational topic modeling, such as LDA, NMF, or embedding-based methods like BERTopic, once a larger and more textually rich dataset becomes available.

4. Research Methodology

4.1. Data Collection

This study employed sentiment analysis to explore university students' perceptions and experiences in using Generative AI as a tool for independent learning. Data were collected from open-ended questions distributed through an online questionnaire, allowing broad participation across different study programs. The number of respondents and their demographic characteristics, including program of study, academic level, and frequency of Generative AI usage, were documented to provide a comprehensive profile of the participants. Although individual response timestamps were automatically recorded during the survey administration, the questionnaire was distributed only once as a single cross-sectional data collection. As all responses were submitted within the same limited time frame, no temporal variation was available to support longitudinal or trajectory-based analyses. Accordingly, sentiment was treated as a static snapshot rather than modeled over time. Figure 2 presents the demographic distribution of the survey participants, including gender, province of origin, current education level, and semester. The sample is dominated by male undergraduate students, with the largest proportion coming from Bali Province and enrolled in Informatics or related study programs.

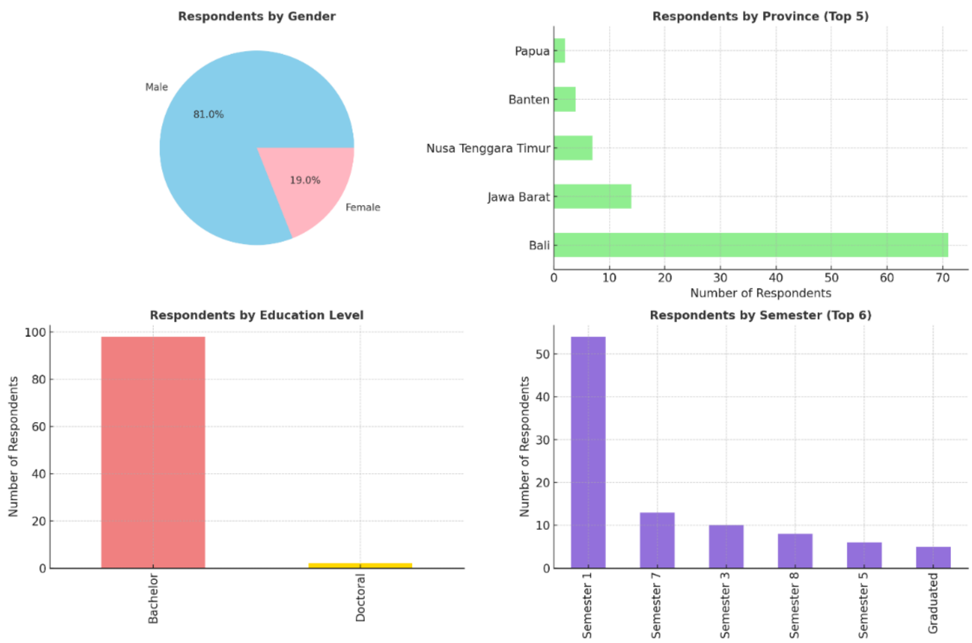

Figure 2. Respondent Demographics Overview

Most respondents are in the early semesters of their studies, particularly Semester 1, while a smaller group represents advanced students or graduates, ensuring a range of perspectives across academic stages. The predominance of first-semester participants suggests that the findings mainly reflect perceptions of students still developing foundational digital literacy and learning independence. Consequently, the generalizability of these insights to upper-level students or graduate cohorts may be limited. Future studies should aim for a more balanced demographic representation to capture a wider spectrum of AI learning experiences.

4.2. Data Labelling

Figure 3 presents the sentiment distribution for two categories: Challenges and Suggestions, based on data labeled using TextBlob in Python. For the "Challenges" category, the counts of Neutral and Positive sentiments are equal, both being the highest, while Negative sentiment is noticeably lower. This indicates that participants' responses regarding challenges are mainly distributed between Neutral and Positive, with fewer expressing Negative sentiment. In contrast, the "Suggestions" category is dominated by Positive sentiment, followed by Neutral, and only a small number register as Negative. This pattern suggests that when providing suggestions, respondents tend to use more positive language, whereas reactions to challenges are more balanced between neutral and positive tones, with negativity less frequently recorded. These results illustrate that TextBlob detected overall optimistic and impartial attitudes in the dataset, particularly in the suggestions provided by participants.

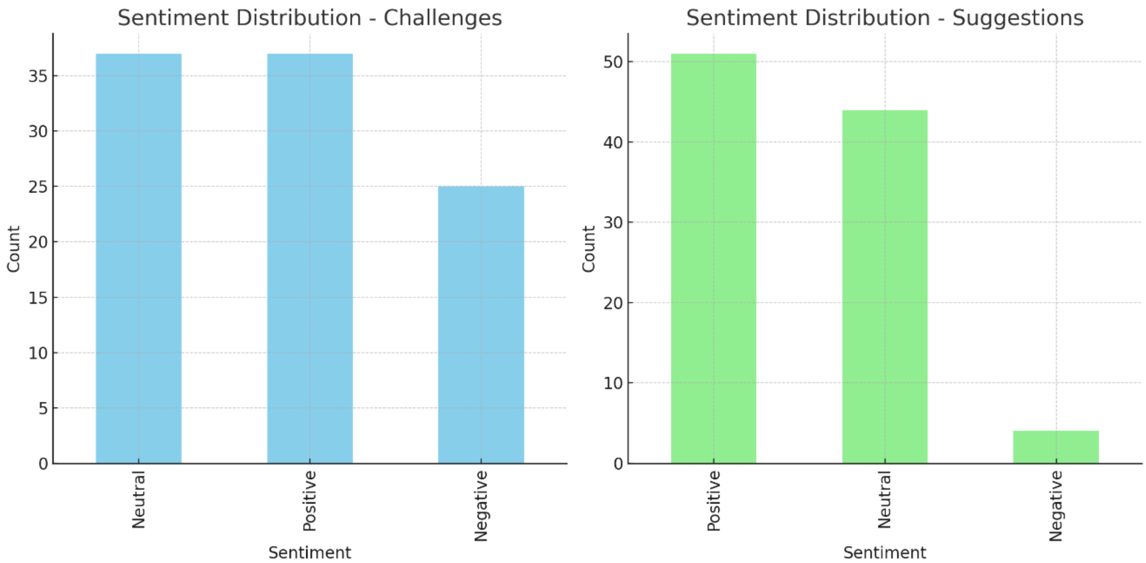

Figure 3. Sentiment Distribution: Challenges and Suggestions in Using Generative AI

4.3. Comparative Analysis of Machine Learning Algorithm

This subsection presents a comparative analysis of the machine learning algorithms employed in the sentiment classification task. Four models, such as Naive Bayes, Logistic Regression, Random Forest, and Linear SVM, were evaluated to determine their effectiveness in identifying student sentiments. The comparison is based on key performance metrics, including accuracy, precision, recall, and F1-score, which provide a comprehensive assessment of predictive capability. By examining the strengths and limitations of each model, this analysis aims to identify the most suitable algorithm for the dataset while highlighting potential challenges such as class imbalance. While this study reports standard weighted evaluation metrics, future work should incorporate macro-averaged scores, Cohen’s Kappa, and class-weighted adjustments to better capture model robustness under class imbalance conditions.

In this study, evaluation relied on weighted accuracy, precision, recall, and F1-score due to the small dataset size and the extreme imbalance between sentiment classes, particularly the negative class with only three samples. Under these constraints, macro-averaged metrics and Cohen’s Kappa were not included because both measures become highly unstable and difficult to interpret in settings where one class contains fewer than five instances. Likewise, class-weighted adjustments were not applied so that all four classical machine learning models could be compared under consistent baseline conditions. This approach ensures methodological comparability while acknowledging that additional evaluation metrics would provide a more balanced assessment once a larger dataset is available. Computational outputs are presented in tabular form to provide clear quantitative comparisons, while thematic findings are illustrated through cluster diagrams to visually summarize conceptual patterns.

4.3.1. Naive Bayes

Table 1 presents classification report for Naive Bayes Algorithm. Naive Bayes achieved an accuracy of 65% with a weighted F1-score of 0.61. The model performed relatively well on the positive class (F1-score 0.75) due to a high recall (0.87), indicating its ability to correctly identify most positive opinions. However, its performance on the neutral class was moderate (F1-score 0.50), and it completely failed to classify the negative class (F1-score 0.00). This limitation is mainly attributed to the extremely small number of negative samples (only three), which prevented the model from learning a reliable distribution. Oversampling and other imbalance handling techniques were not applied because the negative class contained only three samples. Under such conditions, resampling would likely replicate noise and lead to overfitting, rather than enabling the model to learn meaningful patterns. Importantly, this decision reflects a deliberate methodological choice to preserve the integrity and interpretability of the results. The findings therefore remain valid for the observed data distribution and provide meaningful insights into the dominant sentiment patterns, while also highlighting the need for future studies with more balanced datasets to further validate and extend these results.

Table 1. Classification Report for Naïve Bayes

	Precision	Recall	F1-Score	Support
Negative	0.00	0.00	0.00	3
Neutral	0.60	0.43	0.50	14
Positive	0.67	0.87	0.75	23
Accuracy			0.65	40
Micro Avg	0.42	0.43	0.42	40
Weighted Avg	0.59	0.65	0.61	40

4.3.2. Logistic Regression

Table 2 presents classification report for Logistic Regression Algorithm. Logistic Regression produced an accuracy of 62.5% and a weighted F1-score of 0.61. The model demonstrated a more balanced performance across neutral (F1-score 0.56) and positive (F1-score 0.71) classes, although its recall for positive sentiment (0.70) was slightly lower than that of Naive Bayes. Similar to the previous model, Logistic Regression was unable to detect the negative class, underscoring the strong impact of class imbalance.

Table 2. Classification Report for Logistic Regression

	Precision	Recall	F1-Score	Support
Negative	0.00	0.00	0.00	3
Neutral	0.50	0.64	0.56	14
Positive	0.73	0.70	0.71	23
Accuracy			0.62	40
Micro Avg	0.41	0.45	0.42	40
Weighted Avg	0.59	0.62	0.61	40

4.3.3. Random Forest

Table 3 presents classification report for Random Forest Algorithm. This algorithm recorded the lowest accuracy (57%) with a weighted F1-score of 0.56. Interestingly, the model achieved high recall for the neutral class (0.79), suggesting strong sensitivity in identifying neutral responses. Random Forest's high recall for neutral sentiment can be attributed to its ensemble averaging process, which tends to favor dominant or ambiguous classes in small datasets. The model captures a wide range of features but generalizes mid-intensity expressions as 'neutral,' reflecting its sensitivity to overlapping feature distributions between neutral and positive responses. However, this came at the cost of lower precision and weaker performance in the positive class (F1-score 0.62), where several positive samples were misclassified as neutral. Like the other models, Random Forest also failed to recognize the negative class.

Table 3. Classification Report for Random Forest

	Precision	Recall	F1-Score	Support
Negative	0.00	0.00	0.00	3
Neutral	0.46	0.79	0.58	14
Positive	0.75	0.52	0.62	23
Accuracy			0.57	40
Micro Avg	0.40	0.44	0.40	40
Weighted Avg	0.59	0.57	0.56	40

4.3.4. Linear SVM

Table 4 presents classification report for Linear SVM. This algorithm delivered the best overall results, with 65% accuracy and the highest weighted F1-score (0.63). The model achieved a good balance between precision and recall, especially for the positive (F1-score 0.73) and neutral (F1-score 0.61) classes. Although it also failed to classify the negative sentiment, the model's consistent performance across the two dominant classes makes Linear SVM the most effective choice for this dataset.

Table 4. Classification Report for Linear SVM

	Precision	Recall	F1-Score	Support
Negative	0.00	0.00	0.00	3
Neutral	0.53	0.71	0.61	14
Positive	0.76	0.70	0.73	23
Accuracy			0.65	40
Micro Avg	0.43	0.47	0.44	40
Weighted Avg	0.62	0.65	0.63	40

4.3.5. Overall Comparison

Table 5 presents the performance evaluation of four machine learning algorithms, such as Naive Bayes, Logistic Regression, Random Forest, and Linear SVM, applied to sentiment classification. The models were assessed using four standard metrics, namely accuracy, precision, recall, and F1-score, which collectively provide a comprehensive view of classification effectiveness. While accuracy indicates the overall proportion of correct predictions, precision and recall highlight the models' ability to correctly identify sentiment classes, and the F1-score balances these two measures. This comparison allows for a clearer understanding of the relative strengths and weaknesses of each algorithm in handling the dataset.

ISSN 2723-6471

Among the four algorithms, Linear SVM proved to be the most consistent and robust, achieving the best weighted F1-score. Naive Bayes stood out for its high recall in the positive class, while Random Forest excelled in detecting neutral responses but at the expense of overall accuracy. Logistic Regression offered more balanced performance but slightly lagged behind SVM. The models' inability to detect negative sentiments was primarily caused by severe class imbalance (three negative samples). Although techniques such as SMOTE or data augmentation were not applied in this exploratory phase to preserve data authenticity, future studies should investigate these strategies alongside transformer-based approaches to improve minority class detection.

Table 5. Performance Comparison for Naïve Bayes, Logistic Regression, Random Forest, and Linear SVM

Model	Accuracy	Precision	Recall	F1-Score
Naive Bayes	0.650	0.593333	0.650	0.608962
Logistic Regression	0.625	0.593182	0.625	0.605764
Random Forest	0.600	0.607097	0.600	0.581858
Linear SVM	0.650	0.622306	0.650	0.630303

The results indicate that none of the classical machine learning models successfully detected the negative sentiment class. This outcome is largely attributable to the severe imbalance in the dataset, where negative responses constitute only a very small fraction of the total corpus. While techniques such as SMOTE or oversampling could theoretically improve recall for the minority class, their use in this study was avoided due to the risk of introducing artificial or misleading patterns into a small, qualitative dataset. This limitation underscores the need for a larger and more balanced collection of student feedback in future research to enable more representative model training and evaluation.

To further contextualize these performance results, several representative examples illustrate the types of responses that were frequently misclassified across the four models. Sentences that combined positive remarks with subtle critique, such as "AI helps me understand faster, but the explanations are sometimes unclear", were often predicted as positive, despite carrying mixed sentiment. Polite or softened expressions of dissatisfaction, which are common in Indonesian communication, for example, "Hasilnya kurang tepat, tapi masih bisa dipakai"("The results are not very accurate, but they are still usable"), were typically classified as neutral or positive, indicating that the models tended to under-detect mild negative cues embedded within polite phrasing. Very short or context-dependent responses, such as "Lumayan membantu" ("Quite helpful") or "Kadang pas, kadang tidak" ("Sometimes it works, sometimes it doesn't"), were also frequently labeled neutral due to the limited lexical signals available for sentiment inference. These examples illustrate how ambiguity, indirect wording, and minimalistic expressions contribute to misclassification, reinforcing the challenges classical machine learning models face when applied to small, imbalanced, and linguistically nuanced datasets.

In addition to the performance comparison, this study did not include feature-importance or token-level interpretability analyses such as SHAP or LIME. These methods were considered; however, the small dataset size and the exploratory baseline focus of this work limited their applicability. With only 40 samples and highly imbalanced sentiment classes, explainability outputs would likely be unstable and difficult to generalize. Nonetheless, the absence of these analyses is acknowledged as a methodological limitation, as identifying influential words or phrases could provide deeper insight into how the models interpret student responses. Future studies with larger and more balanced datasets will incorporate SHAP, LIME, or similar explainability tools to better understand token-level contributions and enhance model transparency.

This study did not include statistical significance testing to compare model performance. Although tests such as paired t-tests, McNemar's test, or bootstrap-based confidence intervals can help determine whether observed performance differences are meaningful, the small dataset size and severe class imbalance limit the reliability of such inferential analyses. Significance tests generally require sufficiently large and balanced datasets to produce stable estimates, conditions that are not met in this preliminary exploratory study. As a result, model comparisons in this work are presented descriptively, with plans to incorporate appropriate statistical significance testing once additional data are collected.

Ensemble techniques such as bagging, boosting, or stacking were not applied in this study to maintain a clear baseline comparison among the four classical machine learning models. Given the very small dataset and severe class imbalance, implementing ensemble or hybrid architectures would substantially increase model complexity and risk overfitting. To ensure methodological consistency and avoid unstable performance estimates, the analysis in this exploratory phase was restricted to individual classical models, with plans to incorporate ensemble approaches once a larger dataset is available.

Cross-validation techniques such as 5-fold or 10-fold CV were not applied in this study due to the small dataset and the very limited number of negative instances. Even when using stratified folds, several folds would contain zero or near-zero negative samples, resulting in unstable training behavior and unreliable fold-level metrics. To avoid these issues and maintain methodological consistency in this exploratory baseline, the evaluation was conducted using a single 80/20 train–test split, with plans to adopt cross-validation once a larger and more balanced dataset is available.

4.4. Challenges of Using Generative AI in Independent Learning

The analysis of student responses revealed a range of challenges in utilizing Generative AI for independent learning. These challenges extend beyond technical limitations and reflect deeper epistemic, cognitive, and behavioral issues, which can be framed in relation to established learning theories and pedagogical frameworks.

Quantitative frequencies were incorporated to complement the qualitative thematic analysis of student-reported challenges. The most frequently mentioned concern was overreliance on AI (n = 11), where students expressed fear that extensive use could reduce critical thinking and independent learning. Issues related to accuracy and hallucinations (n = 9) were also prevalent, reflecting students' experiences with unclear or incorrect outputs. Several respondents highlighted contextual limitations (n = 7), particularly in technical or mathematical tasks that require deeper domain understanding. Difficulties with prompt engineering (n = 6) and ethical concerns such as plagiarism and reduced originality (n = 6) further illustrate gaps in AI literacy. Other students noted the lack of structured explanations (n = 5) and technical barriers (n = 4). These frequency indicators help clarify the relative prominence of each challenge while maintaining alignment with the qualitative depth of the thematic findings. Tabel 6 presents frequency of themes identified in student-reported challenges and figure 4 presents thematic cluster diagram of student challenges in using Generative AI.

Table 6. Frequency of Themes Identified in Student-Reported Challenges

Theme	Frequency (n)	Description
Overreliance and reduced independence	11	Students express concern about depending too heavily on AI, reducing critical thinking and autonomy.
Accuracy and hallucination issues	9	Challenges related to incorrect, unclear, or misleading outputs generated by AI tools.
Contextual or domain limitation	7	AI struggles with technical, mathematical, or context-heavy academic tasks.
Prompt engineering difficulty	6	Students report challenges in crafting effective prompts to obtain accurate responses.
Lack of structured learning pathways	5	AI answers perceived as fragmented, lacking step-by-step guidance.
Technical or access limitations	4	Barriers such as paywalls, usage limits, unstable internet access.
Ethical concerns (plagiarism, dependency)	6	Worries about academic integrity, originality, and responsible use.

4.4.1. Overreliance and Dependency

A significant concern among students is the risk of becoming overly dependent on Generative AI. This overreliance can reduce critical thinking, self-directed problem-solving, and motivation to consult primary learning resources. Recent studies stress that excessive dependence on AI may compromise students' development of autonomy and essential academic skills, challenging the principles of self-regulated learning and cognitive load management. These findings highlight the need to balance the convenience of AI with opportunities for genuine, independent learning [18].

4.4.2. Accuracy and Information Validity

Questions regarding the accuracy and reliability of AI-generated content are prominent. Students describe encountering “hallucinations”, which refer to instances where AI provides convincing but incorrect answers. Recent research warns that these issues undermine trust and make it crucial for students to develop strong information literacy skills. Effective verification and critical assessment of AI outputs are essential in preventing the spread of misinformation within educational settings [4].

4.4.3. Contextual and Linguistic Limitations

Students often report that AI cannot always grasp contextual, linguistic, or symbolic nuances, especially in mathematics or specialized disciplines. New studies affirm that effective AI integration requires context-sensitive approaches that recognize the diversity of learning situations, in line with situated learning theory. Prompt refinement can mitigate, but not entirely resolve, these contextual challenges [19].

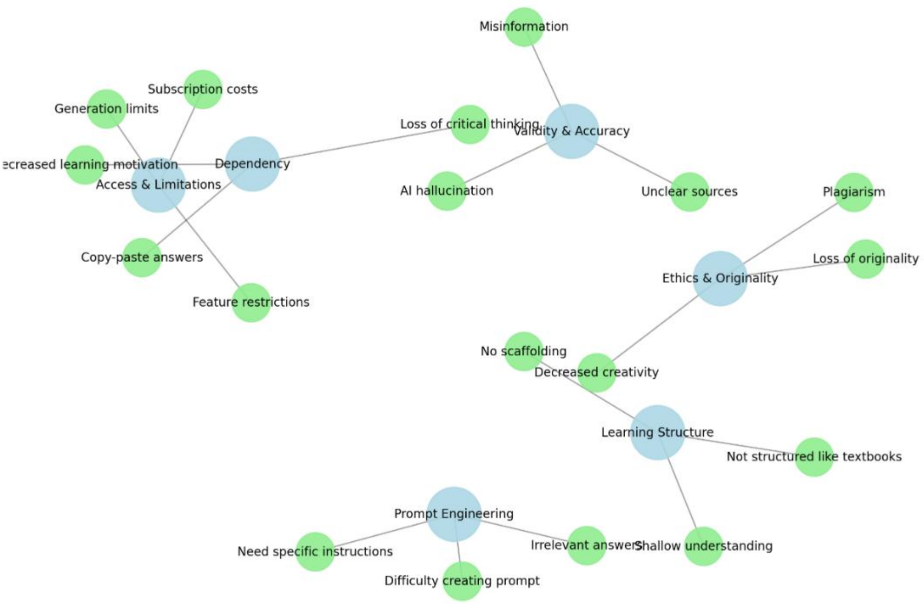

Figure 4. Thematic Cluster Diagram: Student Challenges in Using Generative AI

4.4.4. Prompt Engineering Skills

The ability to design clear, precise prompts is now viewed as a vital digital literacy skill. Research published in 2024 emphasizes that prompt engineering combines both technical ability and metacognitive awareness, aligning with broader frameworks of digital literacy and self-regulated learning. Students increasingly see this skill as a learnable challenge rather than a technical flaw [18], [19]. While prompt engineering is increasingly recognized as a vital digital literacy skill, not all students have equal access to the resources and training required to develop this competence. This disparity may reinforce digital inequities within higher education, underscoring the need for institutions to provide inclusive AI literacy programs that ensure equitable participation and skill development across diverse student groups.

4.4.5. Lack of Structured Learning Pathways

Generative AI frequently generates fragmented or instant answers that lack the structured scaffolding found in textbooks or formal curricula. Recent literature calls for a pedagogy-aware design in AI systems that supports deep, progressive learning rather than superficial recall or comprehension. Absence of clear sequencing can leave knowledge acquisition shallow and unintegrated [18].

4.4.6. Technical and Access Limitations

Technical barriers, including paywalls, usage limits, and inequities in access, continue to shape students' adoption and sustained use of AI tools. Recent studies highlight these practical concerns, noting that equitable and inclusive access to AI-driven learning remains a fundamental policy and design challenge [4].

4.4.7. Ethical Concerns: Plagiarism and Creativity

Ethical challenges such as plagiarism, loss of originality, and stifled creativity are central concerns in the use of Generative AI. Research emphasizes the need for strong academic integrity frameworks and creative learning paradigms to ensure AI complements rather than undermines genuine intellectual growth [4], [18]. Institutions play a critical role in mitigating AI-related plagiarism and misuse by developing transparent AI-integrity policies, strengthening plagiarism detection mechanisms, and training educators to guide ethical use. However, inconsistent enforcement or lack of awareness about these policies may undermine their effectiveness and perpetuate irresponsible AI practices among students. Therefore, institutional governance and continuous ethics education are key to ensuring that AI tools are used responsibly and constructively in academic contexts.

4.5. Student Suggestion for the use of Generative AI in Higher Education

Thematic analysis of student feedback regarding the use of Generative AI in higher education demonstrates several key themes that reflect current opportunities and ongoing concerns about its role in learning. Recent studies confirm that students largely support the presence of AI in academic settings but stress the importance of institutional guidance and ethical practice for effective implementation. Figure 5 presents thematic cluster diagram of student suggestion in using Generative AI. While several student suggestions correspond to previously identified challenges, they represent constructive and solution-oriented responses. To clarify this distinction, student recommendations have been categorized into actionable domains, such as AI as Support Tool for Learning, Ethics and Regulation, Infrastructure and Accessibility, reflecting proactive strategies to address institutional and pedagogical gaps.

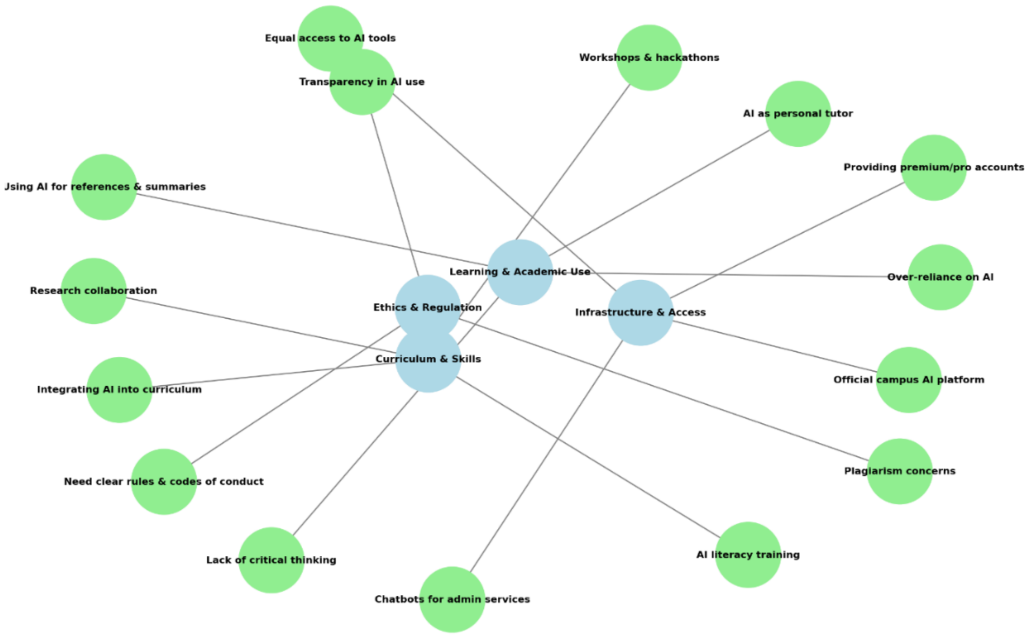

Figure 5. Thematic Cluster Diagram: Student Suggestion in Using Generative AI

For student suggestions, AI as a supportive learning tool was the most frequently referenced theme (n = 12), highlighting students' appreciation for AI's ability to simplify concepts, generate explanations, and speed up learning. Many students recommended the development of ethical guidelines (n = 10) and the implementation of AI literacy or prompt-engineering training (n = 9) to promote responsible and effective usage. Suggestions for curriculum integration (n = 8) reflect students' interest in structured and pedagogically meaningful use of AI within academic settings. Additionally, several respondents emphasized the need for better infrastructure and access (n = 6), including campus-

ISSN 2723-6471

supported tools and AI laboratories. These frequency counts strengthen the data-driven nature of the thematic analysis while acknowledging the interpretive depth of the qualitative responses. Table 7 presents frequency of themes identified in student-reported challenges.

Table 7. Frequency of Themes Identified in Student-Reported Challenges

Theme	Frequency (n)	Description
AI as a support tool (clarification, summarization, idea generation)	12	Students emphasize AI's role in accelerating understanding and providing learning support.
Ethical guidelines and academic integrity policies	10	Requests for clear campus rules to avoid plagiarism and misuse.
AI literacy and prompt engineering training	9	Students suggest workshops, courses, or official training to improve effective use of AI.
Integration into curriculum and pedagogy	8	Interest in structured use of AI in courses, assessments, and academic projects.
Improved infrastructure and access	6	Requests for campus-provided tools, premium access, or dedicated AI labs.

4.5.1. AI as Support Tool for Learning

A dominant theme is the recognition of AI as a support tool for learning rather than a replacement for active participation. Students appreciate AI for assisting with reference searches, summarization, conceptual clarification, and providing interactive learning support. This supports the principles of Constructivist Learning Theory, reinforced by the findings of [20], [21], which highlight that learning is most effective when students construct knowledge proactively and AI acts as scaffolding to facilitate, rather than replace, critical engagement. Recent research further acknowledges that AI tools can help with planning and monitoring learning (Self-Regulated Learning), while maintaining the centrality of student agency.

4.5.2. Ethics and Regulation

Ethics and regulation are also prominent in student recommendations. Many suggest the adoption of clear campus guidelines to prevent plagiarism, excessive dependence on AI, and to distinguish ethical study support from inappropriate task completion. Work by [22], [23] demonstrates that academic integrity frameworks are evolving to address AI's role in learning and assessment, emphasizing transparency and the importance of original work. Work by [24] highlights the risk of plagiarism and the need for robust policies that promote responsible AI use in academia.

Integration of AI into curriculum and pedagogy forms another crucial aspect. Suggestions include AI literacy courses, dedicated workshops, hackathons, and collaborative research between faculty and students. Studies by [25] provide updated models of TPACK (Technological Pedagogical Content Knowledge) for AI, recommending alignment of technology, pedagogy, and content knowledge to optimize student outcomes using AI.

4.5.3. Infrastructure and Accessibility

Infrastructure and accessibility are additional areas of concern. Students propose expanding official campus AI platforms, providing premium AI tools, and dedicated AI labs to support learning and administration. Studies by [23] stresses the necessity of ensuring equitable access to AI technologies across all student groups, reflecting the importance of digital equity and the diffusion of technological innovations within higher education.

Sentiment analysis of student viewpoints reveals a largely positive perspective, emphasizing AI's ability to accelerate and enrich learning, research, and administrative processes. However, many remain cautious, urging safeguards against AI replacing critical thinking and personal initiative. Constructive feedback highlights the need for ethical literacy, transparent regulation, and equal resource provision rather than resistance to technological progress.

In summary, contemporary student feedback aligns with recent academic studies, confirming that the successful integration of Generative AI in education requires not only enthusiasm for innovation but robust institutional guidelines, ethical training, curricular alignment, and equitable infrastructure. Students express readiness to engage constructively with AI, underlining its position as a valuable learning scaffold rather than a substitute for genuine intellectual activity.

As this study is based on open-ended self-reported feedback rather than instrumented interaction tracing, the analysis focuses on students' perceived experiences with generative AI tools rather than system-level behavior. Consequently, references to issues such as hallucinations, unstructured outputs, and interface limitations reflect subjective user accounts rather than objective usability metrics or log-based evidence. This scope is consistent with the survey-driven design of the research, while acknowledging that future work incorporating interaction logs or task-level usability analytics would offer a more comprehensive view of how these limitations manifest during actual use.

5. Conclusion

The integration of Generative AI in autonomous learning at the university level has brought significant advantages, revolutionizing how students acquire knowledge and engage with academic content. The findings of this study indicate that students largely perceive AI tools as valuable aids for enhancing understanding, generating summaries, conducting reference searches, and fostering academic creativity. While the enthusiasm for adopting these technologies is high, the potential risks cannot be overlooked. Student have raised concerns about excessive dependence on AI, the decline of critical thinking skills, and challenges to maintaining academic integrity. In this context, educational institutions play a crucial role in equipping students with sufficient digital literacy and creating a learning environment that strikes a balance between the use of technology and the cultivation of independent competencies. Such efforts are essential to ensure that AI's presence enhances, but does not overshadow, genuine intellectual development.

From a technical perspective, the comparative analysis of four machine learning algorithms for sentiment classification yielded valuable insights. Linear SVM emerged as the most consistent and robust algorithm, achieving an accuracy of 65% and the highest weighted F1-score of 0.63 on the student sentiment dataset. Naive Bayes followed closely, also reaching 65% accuracy but with a slightly lower weighted F1-score of 0.61; it demonstrated strong recall for positive sentiments (recall: 0.87), though it failed to recognize negative ones due to the limited number of samples. Logistic Regression achieved 62.5% accuracy and a weighted F1-score of 0.61, offering balanced performance across sentiment classes, while Random Forest attained an accuracy of 57% and a weighted F1-score of 0.56, excelling in recognizing neutral responses but misclassifying some positives. Despite these strengths, none of the models could successfully classify negative sentiments, underscoring the need for future work on data balancing and augmentation strategies to address class imbalance challenges.

A key limitation of this study is the absence of transformer-based classifiers such as BERT, RoBERTa, or DistilBERT, which represent the current state of the art in sentiment analysis. The primary reason for this exclusion lies in the dataset size and domain specificity; the current corpus is not yet sufficiently large or balanced to train or fine-tune transformer models effectively. Future work will focus on expanding the dataset by collecting additional student feedback across multiple semesters and courses. This will enable the integration and comparative evaluation of transformer-based architectures, allowing for deeper contextual understanding and potentially higher classification accuracy. The planned extension will also explore domain-adaptive pretraining and hybrid feature-based approaches to assess performance improvements over the baseline classical models presented in this study.

Practically, these results highlight the necessity for institutions to develop clear ethical policies, ongoing AI literacy training, and equitable infrastructure so that all academic members can benefit from such technology. Generative AI should be positioned as an empowering support tool that enhances learning processes while respecting disciplinary contexts. Its integration must be tailored to the epistemological and ethical frameworks of each field to ensure appropriate and responsible use. Campuses must implement practical regulations, enhance human resource capacity, and build inclusive AI ecosystems to ensure effective and meaningful integration of these technologies. Concurrently, upholding strong principles of academic integrity is vital so that digital innovation fosters, rather than inhibits, the creativity and autonomy of students. In conclusion, synergy between technology, policy, and student character will be critical to creating higher education systems that adapt effectively to the age of AI, while remaining anchored to authentic learning values.

6. Declarations

6.1. Author Contributions

Conceptualization: N.M.S.I., I.N.Y.A.W., and N.P.W.Y.; Methodology: N.M.S.I.; Software: I.N.Y.A.W.; Validation: N.M.S.I., I.N.Y.A.W., and N.P.W.Y.; Formal Analysis: N.M.S.I., I.N.Y.A.W., and N.P.W.Y.; Investigation: N.M.S.I.; Resources: I.N.Y.A.W.; Data Curation: N.M.S.I.; Writing Original Draft Preparation: N.M.S.I., I.N.Y.A.W., and N.P.W.Y.; Writing Review and Editing: N.M.S.I., I.N.Y.A.W., and N.P.W.Y.; Visualization: N.P.W.Y.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors gratefully acknowledge the financial support provided by the Direktorat Penelitian dan Pengabdian kepada Masyarakat Direktorat Jenderal Riset dan Pengembangan Kementerian Pendidikan Tinggi, Sains, dan Teknologi through the Research Grant Contract No. 129/C3/DT.05.00/PL/2025; 2166/LL8/AL.04/2025; 009/WRIV/DRPM/VI/2025. The authors also extend their sincere appreciation to Universitas Primakara and Universitas Warmadewa for providing research infrastructure and access to relevant datasets. Special thanks are conveyed to the reviewers for their constructive feedback and insightful comments, which have significantly enhanced the quality of this manuscript.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Anara, "AI in higher education statistics: The complete 2025 report." Accessed: Sep. 29, 2025. [Online]. Available: https://anara.com/blog/ai-in-education-statistics

[2] A. Giugliano, "2025 Student AI Survey Insights: AI Tools for Students in Higher Education." Accessed: Sep. 29, 2025. [Online]. Available: https://www.thesify.ai/blog/2025-student-ai-survey-insights-ai-tools-for-students-in-higher-education

[3] M. I. Baig and E. Yadegaridehkordi, "ChatGPT in the higher education: A systematic literature review and research challenges," *Int J Educ Res*, vol. 127, no. 1, pp. 1-24, 2024, doi: 10.1016/j.ijer.2024.102411.

[4] I. M. García-López and L. Trujillo-Liñán, "Ethical and regulatory challenges of Generative AI in education: a systematic review," *Front Educ (Lausanne)*, vol. 10, no.6, p.1-12. 2025, doi: 10.3389/feduc.2025.1565938.

[5] A. M. Hasanein and A. E. E. Sobaih, "Drivers and Consequences of ChatGPT Use in Higher Education: Key Stakeholder Perspectives," *Eur J Investig Health Psychol Educ*, vol. 13, no. 11, pp. 2599–2614, Nov. 2023, doi: 10.3390/ejihpe13110181.

[6] M. Pikhart and L. H. Al-Obaydi, "Reporting the potential risk of using AI in higher Education: Subjective perspectives of educators," *Computers in Human Behavior Reports*, vol. 18, no. 1, pp. 1-13, May 2025, doi: 10.1016/j.chbr.2025.100693.

[7] L. Fan, K. Deng, and F. Liu, "Educational impacts of generative artificial intelligence on learning and performance of engineering students in China," *Sci Rep*, vol. 15, no. 1, pp. 1-21, Jul. 2025, doi: 10.1038/s41598-025-06930-w.

[8] A. A. Quba, M. N. A. Guba, A. Awad, and A. Traish, "University students' reactions to ChatGPT: Enhancing academic English writing and addressing challenges," *International Journal of Innovative Research and Scientific Studies*, vol. 8, no. 1, pp. 2732–2740, 2025, doi: 10.53894/ijirss.v8i1.5048.

[9] S. Lee, H. Choe, D. Zou, and J. Jeon, “Generative AI (GenAI) in the language classroom: A systematic review,” *Interactive Learning Environments*, vol. 2025, no. May, pp. 1–25, May 2025, doi: 10.1080/10494820.2025.2498537.

[10] R. Chugh et al., “The Promise and Pitfalls: A Literature Review of Generative Artificial Intelligence as a Learning Assistant in ICT Education,” *Computer Applications in Engineering Education*, vol. 33, no. 2, pp.1-12, Mar. 2025, doi: 10.1002/cae.70002.

[11] N. F. B. Casillano, “Education in the ChatGPT Era: A Sentiment Analysis of Public Discourse on the Role of Language Models in Education,” *Journal Evaluation in Education (JEE)*, vol. 5, no. 4, pp. 144–154, Oct. 2024, doi: 10.37251/jee.v5i4.1151.

[12] N. M. S. Iswari, N. Afriliana, E. M. Dharma, and N. P. W. Yuniari, “Enhancing Aspect-based Sentiment Analysis in Visitor Review using Semantic Similarity,” *Journal of Applied Data Sciences,* vol. 5, no. 2, pp. 724–735, May 2024, doi: 10.47738/jads.v5i2.249.

[13] N. Afriliana, N. M. S. Iswari, and Suryasari, “Sentiment Analysis of User-Generated Content: A Bibliometric Analysis,” *Journal of System and Management Sciences*, vol. 12, no. 6, pp. 1-12, 2022, doi: 10.33168/JSMS.2022.0634.

[14] N. P. W. Yuniari, N. M. S. Iswari, and I. M. S. Kumara, “Environment Sentiment Analysis of Bali Coffee Shop Visitors Using Bidirectional Encoder Representations from Transformers (BERT) and Generative Pre-trained Transformer 2 (GPT2) Model,” *Journal of Applied Data Sciences*, vol. 5, no. 4, pp. 1765–1781, Dec. 2024, doi: 10.47738/jads.v5i4.302.

[15] D. Magdaleno, M. Montes, B. Estrada, and A. Ochoa-Zezzatti, “A GPT-Based Approach for Sentiment Analysis and Bakery Rating Prediction,” in *Advances in Computational Intelligence, MICAI 2023 International Workshops*: WILE 2023, HIS 2023, and CIAPP 2023, vol. 2024, no. 1, pp. 61–76, 2024. doi: 10.1007/978-3-031-51940-6_7.

[16] C. Costa and M. Murphy, “Generative artificial intelligence in education: (what) are we thinking?,” *Learn Media Technol*, vol. 2025, no. Jun., pp. 1–12, Jun. 2025, doi: 10.1080/17439884.2025.2518258.

[17] A. L. V G and P. K. N S, “Generative Artificial Intelligence in Enhancing English Language Skills: A Systematic Review,” *Stud Media Commun*, vol. 13, no. 4, pp. 256-273, Aug. 2025, doi: 10.11114/smc.v13i4.7763.

[18] M. Giannakos et al., “The promise and challenges of generative AI in education,” *Behaviour & Information Technology*, vol. 44, no. 11, pp. 2518–2544, Jul. 2025, doi: 10.1080/0144929X.2024.2394886.

[19] D. T. K. Ng, E. K. C. Chan, and C. K. Lo, “Opportunities, challenges and school strategies for integrating generative AI in education,” *Computers and Education: Artificial Intelligence*, vol. 8, no. 1, pp. 1-13, Jun. 2025, doi: 10.1016/j.caeai.2025.100373.

[20] H. Guo, W. Yi, and K. Liu, “Enhancing Constructivist Learning: The Role of Generative AI in Personalised Learning Experiences,” in *International Conference on Enterprise Information Systems, ICEIS - Proceedings, Science and Technology Publications, Lda*, vol. 2024, no. 1, pp. 767–770, 2024. doi: 10.5220/0012688700003690.

[21] D. Gibson, V. Kovanovic, D. Ifenthaler, S. Dexter, and S. Feng, “Learning theories for artificial intelligence promoting learning processes,” *British Journal of Educational Technology*, vol. 54, no. 5, pp. 1125–1146, Sep. 2023, doi: 10.1111/bjet.13341.

[22] A. Ateeq, M. Alzoraiki, M. Milhem, and R. A. Ateeq, “Artificial intelligence in education: implications for academic integrity and the shift toward holistic assessment,” *Front Educ (Lausanne)*, vol. 9, no. Oct., pp. 1-12, 2024, doi: 10.3389/feduc.2024.1470979.

[23] Y. Jin, L. Yan, V. Echeverria, D. Gašević, and R. Martinez-Maldonado, “Generative AI in higher education: A global perspective of institutional adoption policies and guidelines,” *Computers and Education: Artificial Intelligence*, vol. 8, no. 1, pp. 1-18, Jun. 2025, doi: 10.1016/j.caeai.2024.100348.

[24] M. Sozon, O. H. Mohammad Alkharabsheh, P. W. Fong, and S. B. Chuan, “Cheating and plagiarism in higher education institutions (HEIs): A literature review,” *F1000Res*, vol. 13, no.1, pp. 788-796, Sep. 2024, doi: 10.12688/f1000research.147140.2.

[25] D. Petko, P. Mishra, and M. J. Koehler, “TPACK in context: An updated model,” *Computers and Education Open*, vol. 8, no.1, pp. 1-24, Jun. 2025, doi: 10.1016/j.caeo.2025.100244.