Journal of Information System Exploration and Research Vol. 4, No. 3, July 2026, pp. 247-264

SHM PUBLISHER
SURYA HIJAU MANFAAT

Journal of Information System Exploration and Research

https://joiser.shmpublisher.com

p-ISSN 2964-1160 | e-ISSN 2963-6361

Explainable Ensemble Learning for Identifying Digital Citizenship Drivers Among Generative AI Users: A SHAP-Based Behavioral Analysis

Muchammad Thoha[1*], Andy Prasetyo Utomo[2], Zainur Romadhon[3]

[1,2,3]*Department of Information Systems, Universitas Muria Kudus, Indonesia*

DOI: https://doi.org/10.52465/joiser.v4i3.79

Received 05 June 2026; Accepted 18 July 2026; Available online 21 July 2026

Article Info

Keywords:
Digital citizenship;
Generative AI;
Explainable AI;
SHAP;
Ensemble learning

Abstract

The rapid adoption of Generative Artificial Intelligence (Generative AI) in higher education has intensified the need to understand the psychological and behavioral factors driving responsible Digital Citizenship. However, studies have relied on linear statistical or Structural Equation Modeling (SEM) approaches, which fail to capture nonlinear relationships with limited interpretability. To address these limitations, this study proposes an Explainable Ensemble Learning framework combining Explainable Artificial Intelligence (XAI) to predict and explain Digital Citizenship behavior. A nationwide dataset comprising 2,008 students from 41 universities across 22 provinces in Indonesia was analyzed using Random Forest, Extreme Gradient Boosting (XGBoost), and Support Vector Regression. Model performance was evaluated using coefficient of determination (R^2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). XGBoost achieved the highest predictive performance (R^2 = 0.388, MAE = 0.230, RMSE = 0.308). SHAP analysis revealed that technology anxiety, self-efficacy, and attitude toward behavior were the most influential predictors, demonstrating that psychological readiness plays a greater role than demographic characteristics in shaping AI use. By integrating ensemble learning with model explanations, this study provides an interpretable framework advancing Digital Citizenship research and supporting evidence-based AI governance and human-centered educational analytics in higher education.

1. Introduction

The rapid emergence of Generative Artificial Intelligence (Generative AI) has substantially reshaped contemporary educational environments worldwide. AI-powered applications such as ChatGPT,

* ***Corresponding Author:***

Muchammad Thoha,
Department of Information Systems,
Universitas Muria Kudus,
Kudus, Central Java, Indonesia.
Email: 202253002@std.umk.ac.id

Gemini, Perplexity, Gamma, and Canva AI are increasingly embedded in academic activities, including information retrieval, academic writing assistance, idea generation, content development, and communication practice. The accessibility, speed, and adaptive capabilities offered by these technologies have accelerated their adoption among university students, positioning AI as an integral component of modern digital learning ecosystems. Recent international educational policy reports further emphasized that AI-assisted learning environments are transforming higher education systems at an unprecedented pace [1], [2].

As AI integration expands across higher education, university students are no longer merely passive users of technology but active participants within AI-supported learning ecosystems. Nevertheless, the growing dependence on Generative AI also introduces significant concerns related to responsible digital behavior, ethical awareness, academic integrity, and digital literacy. These concerns are closely associated with the concept of digital citizenship, which refers to the responsible, ethical, safe, and productive use of digital technologies in online environments. Existing digital citizenship frameworks further stress the importance of ethical participation, responsible communication, and safe engagement within digitally mediated societies. Digital citizenship has become increasingly critical in the era of artificial intelligence because students are now continuously exposed to automated content generation, algorithmic recommendations, misinformation risks, and ethically complex AI-assisted interactions [3], [4]. Within educational contexts, digital citizenship extends beyond basic digital literacy and encompasses multidimensional competencies, including digital inclusion, information and digital competence, civility and ethics, and digital engagement. Understanding how these dimensions are shaped is therefore important for developing effective digital education strategies, promoting responsible AI adoption, and supporting sustainable educational governance frameworks [5].

Previous studies investigating digital citizenship and educational technology adoption predominantly relied on descriptive statistics, correlation analysis, and linear regression approaches. Although these methods are useful for identifying general behavioral associations, they often struggle to capture nonlinear interactions and complex psychological adaptation patterns associated with AI-supported learning environments. Existing behavioral models also frequently assume simplified technology adaptation mechanisms that may inadequately represent multidimensional emotional and cognitive dynamics. Psychological readiness factors such as self-efficacy, emotional adaptation, and technology anxiety may substantially influence how students interact with AI technologies in educational settings. For example, several educational technology studies grounded in the Technology Acceptance Model (TAM), UTAUT, and related behavioral frameworks predominantly relied on linear regression or SEM-based approaches that assume monotonic behavioral relationships [6]. Such assumptions may oversimplify the dynamic psychological adaptation processes associated with Generative AI usage in educational environments [2].

Recent studies have investigated Digital Citizenship and Generative AI adoption using various analytical approaches, including descriptive statistics, multiple regression, Structural Equation Modeling (SEM), and technology acceptance frameworks such as the Technology Acceptance Model (TAM) and the Unified Theory of Acceptance and Use of Technology (UTAUT). These studies have substantially improved understanding of AI adoption, digital literacy, and responsible technology use in educational contexts. However, their primary focus has generally been limited to identifying linear relationships among behavioral constructs, offering limited insight into complex nonlinear interactions and the relative importance of psychological factors. More recent advances in machine learning have demonstrated superior predictive capabilities, yet many studies emphasize prediction accuracy without providing interpretable explanations of how individual factors influence Digital Citizenship behavior. Consequently, the integration of predictive ensemble learning with Explainable Artificial Intelligence (XAI) remains underexplored in Digital Citizenship research involving Generative AI users.

Despite the growing body of research on Digital Citizenship and Generative AI adoption, important limitations remain unresolved. First, most prior studies continue to rely on linear statistical assumptions or conventional Structural Equation Modeling (SEM)-based frameworks that prioritize general associations over hidden behavioral mechanisms. As a result, relatively little is known about how psychological readiness factors dynamically interact to shape responsible AI behavior within educational ecosystems. Second, many predictive educational AI studies still operate as black-box systems, limiting their transparency and reducing their practical value for evidence-based educational policy development [7]. Consequently, existing educational AI studies may overlook important psychological adaptation mechanisms associated with responsible AI behavior, particularly the nonlinear and context-dependent effects of factors such as technology anxiety and self-efficacy. Third, although recent advances in ensemble learning have demonstrated superior predictive capability for educational analytics, only a limited number of studies have integrated these models with Explainable Artificial Intelligence (XAI) to simultaneously achieve high predictive performance and transparent behavioral interpretation. To address these limitations, ensemble learning approaches such as Random Forest and Extreme Gradient Boosting (XGBoost) provide effective solutions for modeling nonlinear relationships and high-dimensional feature interactions within structured behavioral datasets [8]. However, despite their predictive advantages, these models are frequently criticized for functioning as black-box systems whose internal decision mechanisms remain difficult to interpret[9]. Therefore, integrating ensemble learning with SHAP-based Explainable Artificial Intelligence offers a promising approach for developing predictive models that are both accurate and interpretable, thereby supporting trustworthy educational analytics and evidence-based AI governance. To address these research gaps, this study seeks to answer the following research questions:

RQ1. Which psychological, behavioral, and demographic factors most strongly influence Digital Citizenship among university students using Generative AI?

RQ2. Which ensemble learning model provides the best predictive performance for Digital Citizenship prediction while maintaining model interpretability through SHAP?

RQ3. How can SHAP-based Explainable Artificial Intelligence reveal the nonlinear behavioral mechanisms underlying responsible Generative AI use?

To answer these research questions, this study proposes an Explainable Ensemble Learning framework that integrates ensemble learning algorithms with SHAP-based Explainable Artificial Intelligence to identify and interpret the psychological and behavioral factors influencing Digital Citizenship among university students using Generative AI. By integrating predictive modeling with SHAP-based explainability, the proposed framework not only achieves competitive predictive performance but also provides transparent behavioral interpretations that support trustworthy educational analytics and responsible AI adoption in higher education.[7].

This study makes three main contributions to the existing literature. First, from a theoretical perspective, it extends Digital Citizenship and educational technology research by demonstrating that psychological adaptation toward Generative AI involves nonlinear and context-dependent behavioral mechanisms rather than purely linear relationships. Second, from a methodological perspective, the study proposes an Explainable Ensemble Learning framework that integrates ensemble learning algorithms, SHAP-based explainability, robustness validation, fairness evaluation, and behavioral interpretation within a unified analytical framework, thereby improving both predictive performance and model transparency. Third, from a practical perspective, the findings provide interpretable evidence that can assist universities, educators, and policymakers in developing responsible AI governance, strengthening AI literacy, and promoting human-centered educational analytics for responsible Generative AI adoption in higher education.

The remainder of this paper is organized as follows. Section II describes the dataset, preprocessing procedures, machine learning models, and the proposed explainability framework. Section III presents the experimental results together with the explainability analysis and behavioral interpretation. Finally, Section IV concludes the paper by summarizing the main findings, discussing practical implications, acknowledging research limitations, and outlining future research directions.

This study makes several important contributions to the Digital Citizenship literature. Theoretically, it extends previous research by demonstrating that responsible AI behavior is influenced by nonlinear psychological mechanisms rather than solely linear relationships traditionally examined in behavioral studies. Methodologically, this study proposes an Explainable Ensemble Learning framework that integrates XGBoost, Random Forest, Support Vector Regression, and SHAP to provide accurate prediction, transparent interpretation, robustness evaluation, fairness analysis, and ablation analysis within a unified framework. From a practical perspective, the findings provide interpretable evidence that supports universities and policymakers in designing human-centered AI governance, Digital Citizenship education, and responsible AI adoption strategies.

2. Literature Review

Digital Citizenship refers to the responsible, ethical, safe, and productive use of digital technologies within online environments. In educational settings, Digital Citizenship encompasses multiple dimensions, including digital inclusion, information and digital competence, civility and ethics, and digital engagement. The increasing adoption of Generative Artificial Intelligence (Generative AI) has further expanded the scope of Digital Citizenship, as students increasingly interact with AI-generated content, algorithmic recommendations, and intelligent decision-support systems [10], [11]. Consequently, responsible participation in AI-supported learning environments requires not only digital literacy but also ethical awareness, critical thinking, self-regulation, and responsible engagement with AI-assisted educational technologies [12].

Recent advances in Explainable Artificial Intelligence (XAI) have significantly improved the transparency and interpretability of machine learning systems [13], [14]. Among the most widely adopted XAI techniques are Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP). LIME explains individual predictions by constructing locally interpretable surrogate models, whereas SHAP is based on cooperative game theory and quantifies feature contributions using Shapley values, providing greater theoretical consistency and additive feature attribution[15]. Owing to these advantages, SHAP has become one of the most widely adopted explainability techniques for interpreting complex ensemble learning models in educational analytics. Interpretable AI systems are increasingly recognized as essential for promoting trustworthy, accountable, and human-centered educational AI applications[14].

Previous studies investigating technology adoption and digital behavior have predominantly employed behavioral theories such as the Technology Acceptance Model (TAM), the Unified Theory of Acceptance and Use of Technology (UTAUT), and the Theory of Planned Behavior (TPB), together with descriptive statistics, linear regression, and Structural Equation Modeling (SEM) [16], [17]. These approaches have substantially improved understanding of technology acceptance by explaining behavioral intentions through constructs such as perceived usefulness, perceived ease of use, social influence, attitudes, and facilitating conditions. However, they generally assume linear relationships among behavioral constructs and may therefore overlook complex nonlinear interactions associated with Generative AI usage. Likewise, many machine learning studies emphasize predictive performance while providing limited interpretability regarding the behavioral mechanisms underlying model predictions. Consequently, important psychological factors influencing responsible AI participation may remain insufficiently understood [10].

To overcome these limitations, this study proposes an Explainable Ensemble Learning framework that integrates ensemble learning algorithms with SHAP-based explainability to investigate nonlinear psychological adaptation patterns associated with Digital Citizenship among university students using Generative AI. By combining predictive modeling with transparent model interpretation, the proposed framework complements traditional behavioral theories and provides more comprehensive insights into the psychological mechanisms underlying responsible AI adoption in higher education [13].

3. Method

This study employed a quantitative computational approach to investigate the factors influencing Digital Citizenship behavior among university students using Generative Artificial Intelligence (Generative AI) technologies. The proposed framework integrated machine learning prediction and Explainable Artificial Intelligence (XAI) analysis to identify dominant behavioral factors associated with responsible AI participation [18], [19]. Figure 1 presents the overall research workflow adopted in this study. The framework begins with data acquisition and preprocessing, followed by feature engineering and dataset partitioning for model development. Subsequently, multiple machine learning models are trained and evaluated, after which SHAP-based explainability analysis is conducted to interpret feature contributions. Finally, the findings are translated into behavioral interpretations and educational implications.

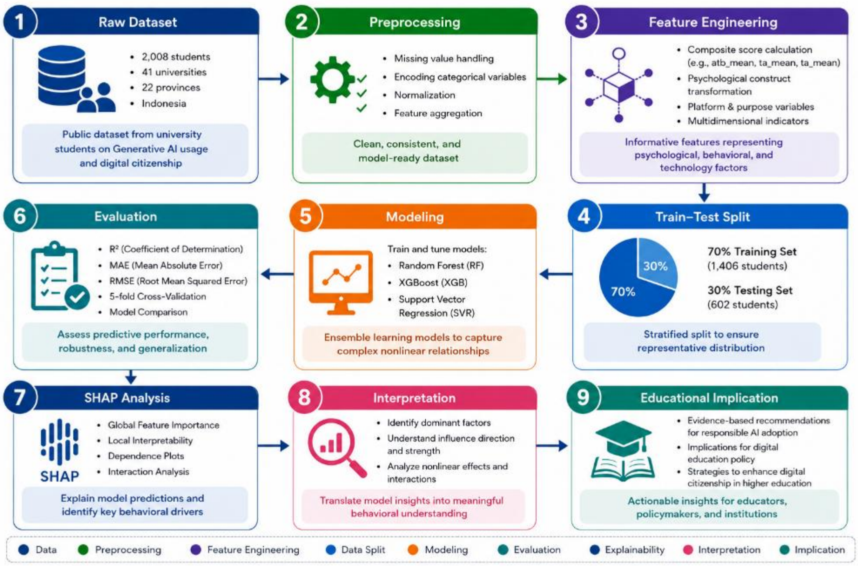


Figure 1. Overview of the proposed explainable ensemble learning framework for identifying digital citizenship determinants among generative AI users.

As shown in Figure 1. the research process begins with data acquisition from a publicly available dataset, followed by preprocessing and feature engineering to prepare the data for modeling. Three ensemble learning models (XGBoost, Random Forest, and Support Vector Regression) were subsequently trained and evaluated using multiple performance metrics. The best-performing model was further analyzed using SHAP to identify feature importance and nonlinear behavioral patterns, providing interpretable insights into the determinants of Digital Citizenship. The dataset used in this study was obtained from the publicly available Mendeley Data repository (DOI: 10.17632/h4gcfmbk5c.3), which contains survey responses from university students actively using Generative Artificial Intelligence (Generative AI) technologies in educational environments [20], [21]. The present study employed the dataset exclusively for secondary data analysis to develop predictive machine learning models and conduct Explainable Artificial Intelligence (XAI) analysis. The final dataset consisted of 2,008 valid respondents collected from 41 universities across 22 provinces in Indonesia. The relatively broad geographical coverage improves the representativeness of behavioral adaptation patterns associated with Generative AI usage in higher education environments. Table 1 summarizes the primary variables utilized in the predictive modeling process.

Table 1. Variable and feature descriptions

Variable	Description	Type	Category
Gender	Respondent gender	Binary	Demographic
Uni_type	Public or private university	Binary	Demographic
Study_year	Academic year level	Ordinal	Demographic
Parent_edu	Parent’s educational background	Ordinal	Demographic
Use_chatGPT	ChatGPT usage	Binary	Platform
Use_gemini	Gemini usage	Binary	Platform
Purpose_info_search	AI used for information search	Binary	Purpose
Atb_mean	Attitude toward behavior score	Continuous	Psychological
Se_mean	Self-efficacy score	Continuous	Psychological
Ta_mean	Technology anxiety score	Continuous	Psychological
Dc_total	Digital citizenship score	Continuous	Target

The multidimensional construction of Digital Citizenship adopted in this study is consistent with established educational frameworks emphasizing responsible, ethical, and productive participation in digital environments [22], [23].

According to the original dataset documentation, the survey data were collected using an online questionnaire distributed through Google Forms to university students from 41 universities across 22 Indonesian provinces. The original study employed a simple random sampling technique to obtain respondents with diverse demographic backgrounds and experiences in using Generative AI for educational purposes environments [20], [21]. The present study utilized the publicly available dataset as secondary data without modifying the original sampling procedure.

The measurement instrument included constructs related to Digital Citizenship, Attitude Toward Behavior, Self-Efficacy, and Technology Anxiety. Based on the original dataset validation, the questionnaire demonstrated satisfactory psychometric properties, with an overall Cronbach's Alpha of 0.883, a Kaiser–Meyer–Olkin (KMO) value of 0.921, and a statistically significant Bartlett's Test of Sphericity ($p < 0.001$), indicating good internal consistency and construct validity

Prior to model development, several preprocessing procedures were conducted to improve data quality and compatibility with machine learning algorithms. These procedures included missing value inspection, categorical feature encoding, feature normalization, and feature aggregation. Categorical variables associated with Generative AI platform usage and AI-assisted learning purposes were transformed using one-hot encoding techniques. Psychological constructs were aggregated into composite scores using mean-based calculations, resulting in variables such as atb_mean, se_mean, and ta_mean. The Digital Citizenship target variable (dc_total) was constructed by aggregating four dimensions: Digital Inclusion (DI), Information and Digital Competence (IDC), Civility and Ethics (CE), and Digital Engagement (DE) [22]. To prevent data leakage, all target-related subdimensions were excluded from the feature set during model training. Feature scaling was additionally applied prior to Support Vector Regression because kernel-based models are sensitive to feature magnitude differences.

Three machine learning algorithms were employed in this study, namely Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR). Random Forest was selected because of its ability to model complex feature interactions while reducing overfitting through bagging mechanisms [24]. XGBoost was employed due to its strong predictive performance and computational efficiency in structured behavioral datasets [24]. Support Vector Regression was utilized as a comparative baseline model to evaluate the relative effectiveness of ensemble learning approaches. The dataset was divided into training and testing subsets using a 70:30 ratio to provide sufficient observations for model training while preserving an independent testing subset for unbiased performance evaluation. Furthermore, five-fold cross-validation was employed to improve the robustness of model performance estimation and reduce variance associated with a single train-test partition.

Model performance was evaluated using the coefficient of determination (R^2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). These metrics were selected to assess explanatory capability, prediction accuracy, and model stability across different validation scenarios. The hyperparameter configurations were determined through iterative empirical tuning guided by previous machine learning studies and repeated experimental evaluation. Candidate parameter values were evaluated to achieve an appropriate balance between predictive accuracy, computational efficiency, and model generalization while minimizing the risk of overfitting. The final hyperparameter configurations are presented in Table 2.

Table 2. Model hyperparameters

Model	Hyperparameter	Value
XGBoost	n_estimators	300
	learning_rate	0.05
	max_depth	6
Random Forest	n_estimators	500
	max_depth	12
SVR	kernel	rbf
	C	10

To improve model interpretability, SHAP (Shapley Additive Explanations) was employed to analyze feature contributions toward Digital Citizenship predictions. SHAP was selected because of its theoretically consistent additive attribution mechanism and its ability to provide both global and local explanations for complex machine learning models [18], [19]. Global interpretability was used to identify dominant psychological and behavioral factors influencing responsible AI participation, whereas local interpretability was utilized to explain individual prediction outcomes. In addition, SHAP dependence plots and interaction analyses were employed to investigate nonlinear behavioral adaptation patterns among psychological variables. This explainability framework enabled transparent interpretation of the predictive models and supported evidence-based educational analytics within AI-assisted learning environments [25].

This study was based exclusively on secondary analysis of an anonymized dataset obtained from the publicly available Mendeley Data repository (DOI: 10.17632/h4gcfmbk5c.3). No additional data were collected from human participants, and no personally identifiable information was accessed during this study. According to the original dataset documentation, ethical approval was granted by the Research Ethics Committee of Universitas Muhammadiyah Surakarta (Approval No. 1520/KEPK-FIK/VIII/2025), and all participants provided informed consent prior to completing the survey [26].

4. Results and Discussion

This section presents the empirical findings of the study and discusses their implications for understanding digital citizenship behavior among Generative AI users. The results are organized into several subsections, beginning with an overview of the dataset characteristics and predictive performance, followed by an interpretation of the key factors influencing digital citizenship based on explainable machine learning analysis.

4.1. Dataset characteristics and predictive performance

This study utilized a national-scale dataset comprising responses of 2,008 Indonesian university students who actively used Generative Artificial Intelligence (Generative AI) platforms in educational contexts. The data were collected from 41 universities distributed across more than 22 provinces in Indonesia, representing both public and private higher education institutions. The preprocessing and validation procedures confirmed that the collected data were sufficiently reliable for predictive behavioral analysis and explainable machine learning modeling.

The demographic distribution showed that female respondents dominated the dataset, accounting for approximately 77.4% of the total participants, while male students represented 22.6%. Regarding institutional background, students from public universities comprised 58.2% of respondents, whereas students from private universities accounted for 41.8%. The analysis of Generative AI platform usage indicated that ChatGPT was the most widely used platform among students, followed by Gemini, Perplexity AI, Microsoft Copilot, Canva AI, and several other AI-assisted tools. This finding demonstrates that conversational AI systems have become increasingly integrated into academic activities such as information retrieval, assignment assistance, academic writing, and idea generation [27].

In addition to demographic variables, the dataset incorporated multidimensional psychological and behavioral constructs, including technology anxiety, self-efficacy, and attitudes toward AI-assisted behavior. These characteristics enabled comprehensive modeling of Digital Citizenship behavior within AI-supported educational environments [28], [29]. The relatively large and geographically diverse dataset provided a useful foundation for investigating behavioral adaptation patterns associated with Generative AI usage in higher education settings.

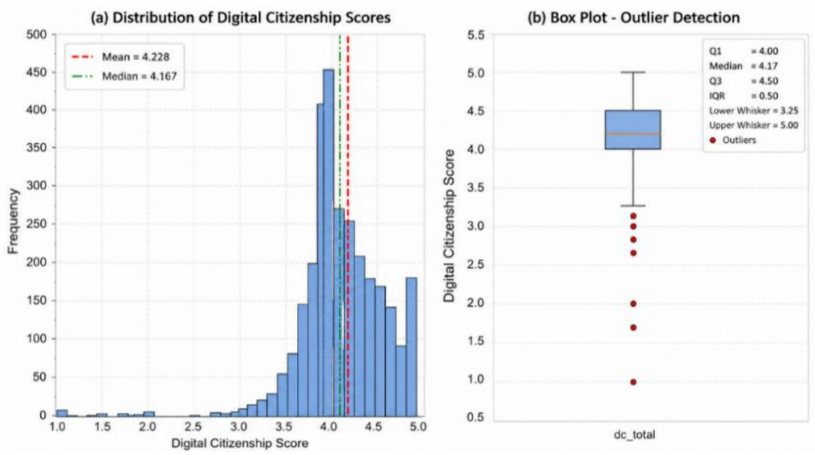


Figure 2. Distribution of digital citizenship scores

Figure 2 illustrates the distribution of the target variable across respondents included in the study. The distribution indicates sufficient variability in Digital Citizenship scores, supporting the suitability of regression-based predictive modeling for identifying behavioral adaptation patterns associated with Generative AI usage.

Table 3. Respondent demographics and AI platform usage

Characteristic	Category	n	%
Gender	Female	1554	77.4
	Male	454	22.6
University Type	Public	1168	58.2
	Private	840	41.8
AI Platform	ChatGPT	1759	87.6
	Gemini	878	43.7
	PerplexityAI	685	34.1
	Microsoft Copilot	353	17.6
	Canva AI	305	15.2

Table 3 summarizes the demographic composition and Generative AI platform usage characteristics of the respondents included in the study. The results indicate that ChatGPT dominated platform usage, followed by Gemini and Perplexity AI, reflecting the increasing adoption of conversational AI systems within higher education environments.

This study further compared the predictive performance of three machine learning algorithms, namely XGBoost, Random Forest, and Support Vector Regression (SVR), for predicting Digital Citizenship behavior among Generative AI users. The evaluation focused on three regression metrics, namely coefficient of determination (R^2), mean absolute error (MAE), and root mean squared error (RMSE).

The experimental results demonstrated that XGBoost achieved the best overall predictive performance with an R^2 score of 0.388, MAE of 0.230, and RMSE of 0.308. Random Forest produced competitive performance with an R^2 value of 0.353, while SVR achieved the lowest predictive performance with an R^2 score of 0.251.

The superior performance achieved by XGBoost suggests that boosting-based ensemble learning may be more effective for modeling complex behavioral adaptation patterns associated with Digital Citizenship in AI-assisted educational environments [30], [31]. Meanwhile, the lower performance observed in SVR indicates that Digital Citizenship behavior cannot be adequately explained using simpler regression boundaries alone. This finding suggests that Digital Citizenship behavior may emerge through multidimensional psychological adaptation mechanisms involving cognitive, emotional, and contextual interactions rather than through isolated linear relationships alone.

Although the proposed XGBoost model achieved the highest predictive performance among the evaluated algorithms, the obtained R^2 value of 0.388 indicates that a substantial proportion of the variance in Digital Citizenship remains unexplained. This outcome is expected because Digital Citizenship is a multidimensional behavioral construct influenced not only by observable demographic and psychological variables but also by contextual, institutional, cultural, socioeconomic, and environmental factors that were beyond the scope of the present dataset. Similar findings have been reported in previous educational AI and behavioral prediction studies, where complex human behaviors are rarely explained by a single set of measurable predictors. Therefore, despite the moderate R^2 value, the proposed model provides meaningful predictive capability while maintaining strong interpretability through SHAP-based explainability analysis.

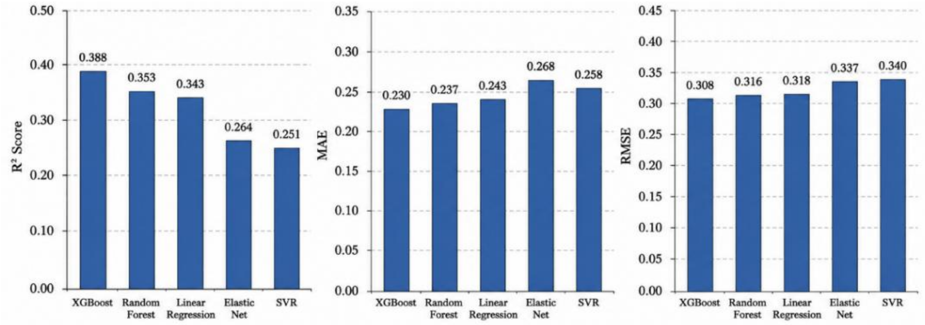


Figure 3. Comparative performance of machine learning models

Figure 3 presents the comparative predictive performance of the evaluated machine learning models across R^2, MAE, and RMSE metrics. The results consistently demonstrate the superiority of XGBoost compared with Random Forest and SVR in capturing behavioral adaptation patterns associated with responsible AI participation.

Table 4. Predictive performance comparison

Model	R^2	MAE	RMSE
XGBoost	0.388	0.230	0.308
Random Forest	0.353	0.237	0.316
Linear Regression	0.343	0.243	0.318
Elastic Net	0.264	0.268	0.337
SVR	0.251	0.258	0.340

Table 4 summarizes the detailed predictive performance results obtained from each evaluated machine learning model. XGBoost achieved the highest explanatory capability and lowest prediction error among the evaluated models, indicating its suitability for modeling Digital Citizenship behavior within AI-assisted educational environments.

Although the predictive performance of XGBoost was superior to the other evaluated models, the obtained R^2 value (0.388) indicates that approximately 38.8% of the variance in Digital Citizenship can be explained by the available predictors. This relatively moderate explanatory power is expected because Digital Citizenship represents a complex behavioral construct influenced by numerous psychological, social, educational, and contextual factors that were not fully captured by the survey variables. Human behavioral phenomena generally exhibit lower R^2 values than engineering or physical prediction problems due to their inherent variability and latent influences. Therefore, the obtained predictive performance remains meaningful for educational behavioral analytics while highlighting opportunities for future studies to incorporate additional contextual variables.

Table 5. Wilcoxon signed-rank test results

Comparison	Statistic	p-value	Interpretation
XGBoost vs Random Forest	86284.5	0.265	Not significant
XGBoost vs SVR	71571.0	<0.001	Significant
XGBoost vs Linear Regression	79113.5	0.005	Significant
XGBoost vs Elastic Net	65390.5	<0.001	Significant

To determine whether the observed differences in predictive performance were statistically meaningful, Wilcoxon signed-rank tests were conducted using the absolute prediction errors of the evaluated models. The results are summarized in Table 5. XGBoost significantly outperformed Support Vector Regression ($p < 0.001$), Linear Regression ($p = 0.005$), and Elastic Net ($p < 0.001$). However, the difference between XGBoost and Random Forest was not statistically significant ($p = 0.265$), indicating that both ensemble learning models achieved comparable predictive capability. These findings suggest that while boosting-based learning consistently provided superior performance over linear and kernel-based approaches, both ensemble methods were similarly effective for modeling Digital Citizenship behavior.

Overall, the performance comparison indicates that ensemble learning approaches were more effective for capturing complex behavioral patterns associated with Digital Citizenship in AI-assisted educational environments. These findings also reinforce the suitability of ensemble learning approaches for educational behavioral analytics involving heterogeneous psychological and contextual variables.

4.2. Global explainability analysis using SHAP

To improve model transparency and interpretability, SHAP (Shapley Additive Explanations) was applied to identify the most influential factors contributing to Digital Citizenship prediction. SHAP enables the interpretation of complex machine learning models by quantifying the contribution of each feature toward prediction outcomes, thereby supporting more transparent and explainable educational analytics.

The SHAP analysis consistently revealed that psychological variables dominated the prediction process across all evaluated machine learning models. In particular, technology anxiety (ta_mean) emerged as the most influential feature. Within the XGBoost model, technology anxiety contributed approximately 36.8% of the total feature importance, followed by attitude toward behavior (atb_mean) and self-efficacy (se_mean). These findings suggest that students' emotional readiness and behavioral adaptation toward AI technologies play a central role in shaping responsible Digital Citizenship behavior [32], [33].

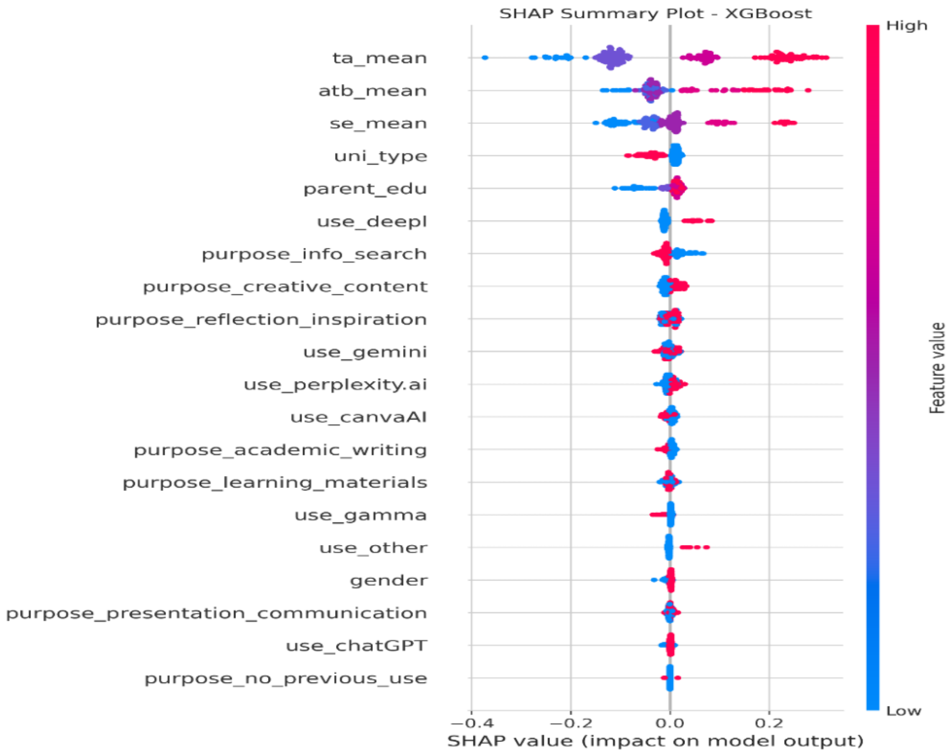


Figure 4. SHAP summary plot for XGBoost model

Figure 4 illustrates the global SHAP summary plot generated from the XGBoost model. The visualization presents both the relative importance and directional contribution of each feature toward Digital Citizenship prediction outcomes. Features positioned at the top of the ranking contributed more substantially to prediction performance, while the color gradient represents the magnitude of feature values.

Table 6. SHAP feature importance ranking across models

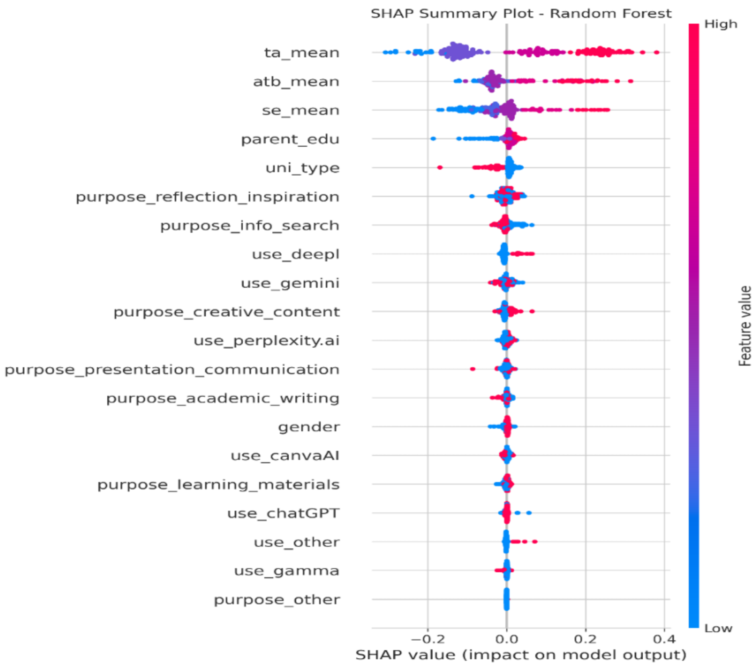
Rank	Feature	XGBoost(%)	Random Forest(%)	SVR(%)	Mean Rank	Consensus Rank
1	Technology Anxiety	36.8	38.9	27.0	1.00	1
2	Attitude Toward Behavior	16.3	16.7	14.5	2.00	2
3	Self-Efficacy	14.0	14.5	10.3	3.00	3
4	University Type	5.1	4.3	4.9	4.33	4

Table 6 summarizes the comparative SHAP feature importance rankings obtained from XGBoost, Random Forest, and SVR models. Technology anxiety consistently achieved the highest importance score across all evaluated models, followed by attitude toward behavior and self-efficacy.

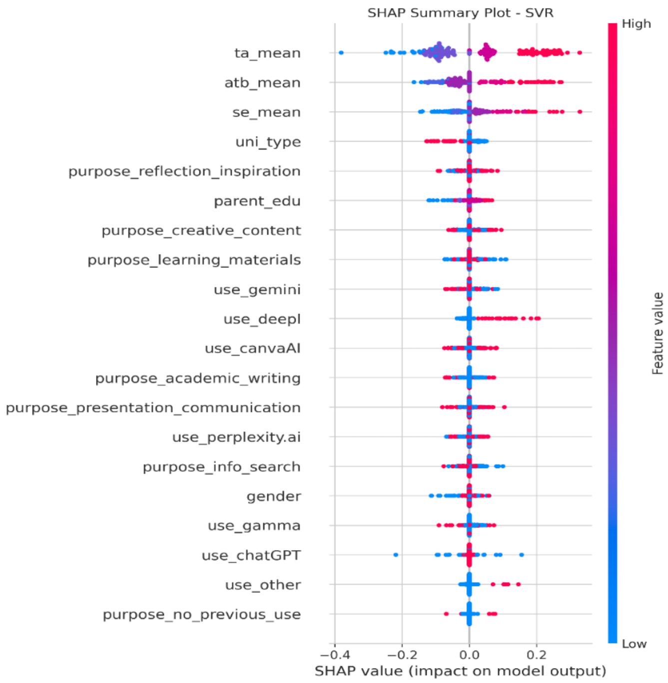
Similar patterns were also observed in Random Forest and SVR models, indicating relatively strong consistency in the explainability results. This observation highlights the stronger influence of emotional adaptation, confidence, and behavioral attitudes toward AI technologies compared with demographic background or platform preference in shaping responsible Digital Citizenship behavior. The dominance of behavioral and emotional factors identified in this study is also aligned with previous educational AI research emphasizing the importance of emotional readiness and adaptive technology engagement in responsible AI participation [34].

Interestingly, demographic variables such as university type and parental education demonstrated substantially lower importance scores compared with psychological constructs. This pattern indicates that Digital Citizenship behavior in the era of Generative AI may depend more strongly on internal behavioral readiness than on socioeconomic conditions alone [28]. Furthermore, platform-related variables such as ChatGPT, Gemini, and Perplexity AI usage contributed relatively smaller predictive effects, suggesting that responsible AI participation is shaped more substantially by students' adaptive interaction with AI technologies than by platform-specific technological preferences.

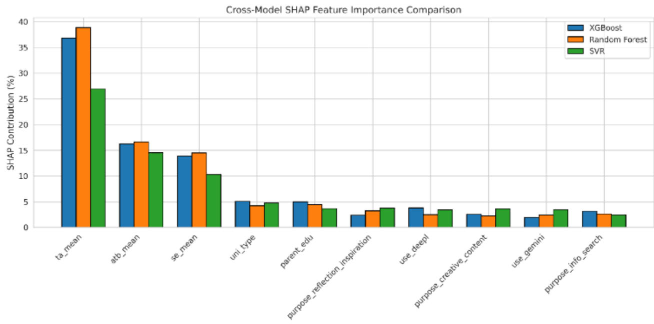
To further evaluate the reliability of the explainability results, feature importance rankings were compared across XGBoost, Random Forest, and SVR models. The comparison demonstrated relatively stable feature rankings, with technology anxiety consistently remaining the most influential predictor, followed by attitude toward behavior and self-efficacy.



(a)



(b)



(c)

Figure 5. Comparison of SHAP-based explainability results across machine learning models, including (a) Random Forest SHAP summary plot, (b) Support Vector Regression (SVR) SHAP summary plot, and (c) cross-model feature importance ranking

The observed consistency indicates that the identified behavioral relationships were not merely artifacts produced by a specific algorithm. Instead, the explainability results revealed relatively robust behavioral patterns embedded within the dataset. Such consistency across different algorithms is particularly important in educational Explainable AI research because unstable explanations may reduce stakeholder trust and limit practical adoption within educational decision-making environments.

The relatively consistent feature rankings across models further strengthen the reliability of the proposed explainable ensemble learning framework. Moreover, the repeated dominance of psychological variables across different algorithms reinforces the interpretation that Digital Citizenship behavior is fundamentally influenced by students' emotional and cognitive adaptation toward AI technologies. Collectively, these findings indicate that psychological readiness may represent a relatively stable behavioral foundation underlying responsible AI participation across multiple predictive modeling approaches.

4.3. Nonlinear behavioral and robustness analysis

To further investigate the behavioral mechanisms underlying Digital Citizenship in AI-assisted educational environments, additional analyses were conducted, including nonlinear behavioral modeling, robustness evaluation, ablation analysis, and fairness assessment. These analyses were performed to examine whether the identified behavioral patterns remained stable under different analytical perspectives and validation procedures.

The nonlinear relationship between technology anxiety and Digital Citizenship behavior was investigated using linear regression, quadratic regression, and Generalized Additive Model (GAM) spline analysis. The linear regression model produced an R^2 value of 0.261 and identified a general negative relationship between technology anxiety and Digital Citizenship behavior. The quadratic regression model slightly improved predictive performance and indicated the presence of nonlinear behavioral dynamics. More importantly, the GAM spline analysis achieved the highest explanatory performance with an R^2 value of approximately 0.305 and identified two important inflection points at technology anxiety scores of approximately 2.63 and 4.96.

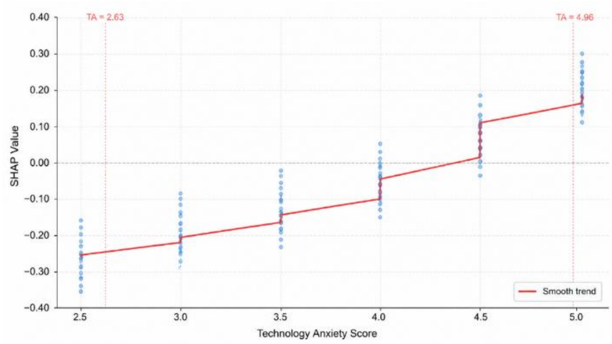
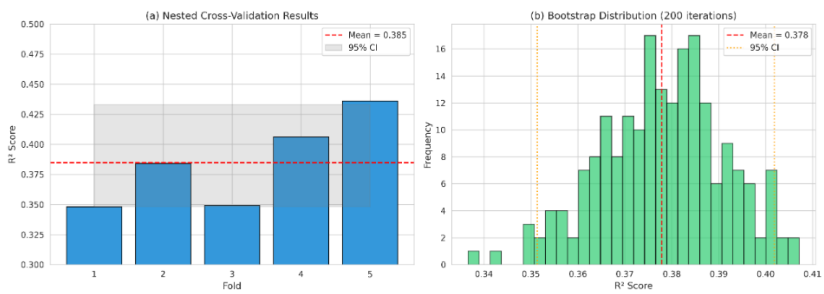


Figure 6. Technology anxiety dependence plot

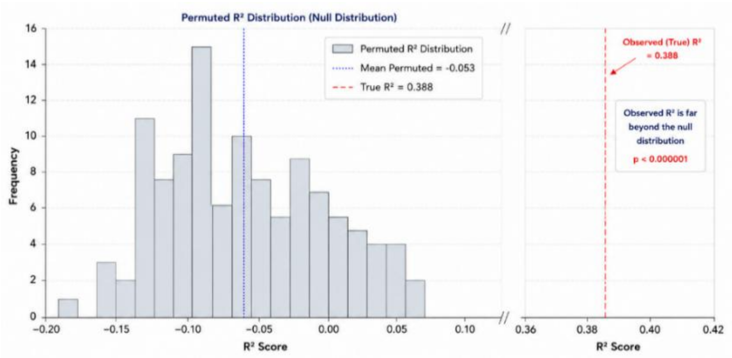
Figure 6 illustrates the nonlinear relationship between technology anxiety and Digital Citizenship behavior identified through GAM-based dependence analysis. The results indicate that the behavioral influence of technology anxiety is not purely linear. Moderate levels of technology anxiety may encourage more cautious, reflective, and ethically aware interaction with AI systems, thereby indirectly supporting responsible digital behavior. However, excessive anxiety may reduce confidence, increase cognitive overload, and discourage meaningful digital participation, ultimately weakening responsible AI engagement.

This nonlinear pattern partially extends previous educational technology studies that primarily assumed linear associations between technology anxiety and digital behavior [32]. Unlike many earlier studies that treated technology anxiety as a uniformly negative factor, the present findings indicate that its behavioral influence may vary across different stages of AI adaptation processes. More broadly, the results demonstrate that psychological adaptation toward Generative AI technologies may involve dynamic and context-sensitive mechanisms rather than simplified linear behavioral responses [27], [35].

To evaluate model reliability and generalizability, several robustness analyses were conducted. Nested cross-validation produced an average R^2 value of approximately 0.385 with relatively small variance across folds, indicating stable predictive performance under different validation scenarios. Bootstrap analysis further demonstrated strong model consistency by generating a narrow confidence interval and relatively stable R^2 distributions across 200 bootstrap iterations. In addition, permutation testing produced a significance value of $p < 0.000001$, confirming that the observed predictive performance was statistically significant and unlikely to be generated by random chance.



(a)



(b)

Figure 7. Robustness and statistical validation results of the proposed explainable ensemble learning framework, including (a) nested cross-validation and bootstrap analysis and (b) permutation test for statistical significance validation

Table 7. Robustness evaluation results

Validation Method	Result
Nested Cross-Validation Mean R^2	0.385
Bootstrap Mean R^2	0.378
Bootstrap Iterations	200
Permutation Test p-value	$p < 0.000001$
Bootstrap (95% Confidence Interval)	[0.352, 0.401]

As shown in Figure 7(a), the nested 5-fold cross-validation yielded R^2 values ranging from 0.348 to 0.436, with a mean of 0.385, indicating stable predictive performance across different validation folds. The bootstrap analysis further confirmed this stability by producing a mean R^2 of 0.378 and a relatively narrow 95% confidence interval. Furthermore, Figure 7(b) demonstrates that the observed R^2 (0.388) lies well beyond the null distribution generated through permutation testing, resulting in a highly significant p-value ($p < 0.000001$). Table 7 summarizes the numerical results of the robustness evaluation, showing consistent performance across the nested cross-validation, bootstrap, and permutation analyses. Collectively, these findings demonstrate that the proposed explainable ensemble learning framework achieves robust, reliable, and statistically significant predictive performance under repeated sampling and randomized validation conditions, supporting its practical application in educational analytics and AI-assisted learning environments. An ablation study was subsequently conducted to evaluate the relative contribution of different feature groups toward predictive performance. The results demonstrated that removing psychological variables caused the most substantial performance degradation among all evaluated feature categories.

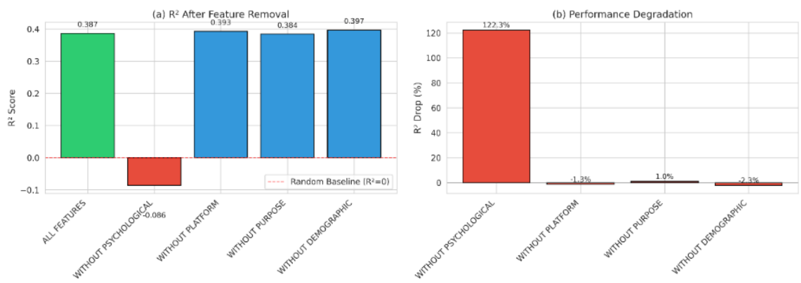


Figure 8. Ablation study results showing (a) the effect of removing each feature group on the model R^2 score and (b) the relative performance degradation (R^2 drop) compared with the full model.

As shown in Figure 8(a), excluding the psychological variables caused the R^2 score to decrease dramatically from 0.387 to 0.086, representing a performance decline of approximately 122.3%. This substantial degradation indicates that psychological variables provide the dominant explanatory information associated with responsible AI participation. In contrast, removing platform-, purpose-, or demographic-related variables resulted in only marginal changes in predictive performance, as illustrated in Figure 8(b). These findings reinforce the interpretation that psychological constructs constitute the primary drivers of Digital Citizenship behavior among Generative AI users. Furthermore, the ablation results are highly consistent with the SHAP explainability analysis, in which Technology Anxiety, Self-Efficacy, and Attitude Toward Behavior were consistently identified as the most influential predictors. This agreement between ablation analysis and SHAP interpretation provides additional evidence for the robustness and reliability of the identified behavioral mechanisms underlying responsible AI participation.

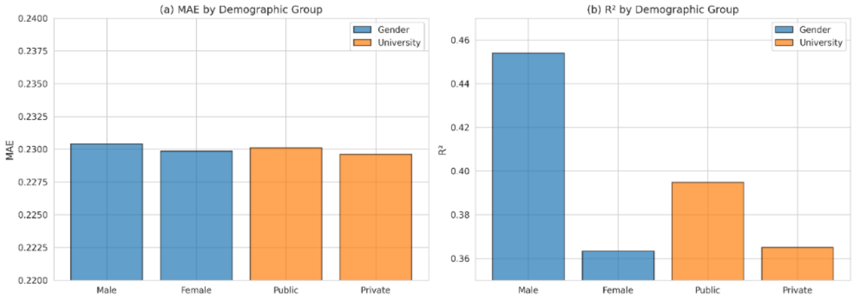


Figure 9. Fairness analysis across demographic groups: (a) comparison of MAE values and (b) comparison of R^2 values for gender and university type.

Table 8. Fairness evaluation results

Group	MAE	R^2
Male	0.230	0.453
Female	0.229	0.363
Public University	0.230	0.395
Private University	0.229	0.365

As shown in Figure 9(a), the MAE values remained highly comparable across all demographic groups, ranging from 0.229 to 0.230, indicating consistent prediction errors regardless of gender or university type. Figure 9(b) presents the corresponding R^2 scores, where male respondents achieved a higher predictive performance ($R^2 = 0.453$) than female respondents ($R^2 = 0.363$), while students from public universities also showed slightly higher predictive performance than those from private universities ($R^2 = 0.395$ vs. 0.365). Although moderate differences in R^2 were observed, the consistently similar MAE values suggest that the proposed framework maintains reasonably stable predictive behavior across the evaluated demographic groups. These subgroup variations may reflect contextual educational differences rather than substantial model bias, warranting further investigation using more diverse datasets and fairness metrics.

4.4. Discussion of research contributions and educational implications

The findings of this study contribute to the growing literature on Digital Citizenship and AI-assisted learning environments in several important ways. First, the results extend existing Digital Citizenship research into the context of Generative AI ecosystems, where students continuously interact with AI-generated content, automated recommendations, and adaptive learning systems. The study broadens current understanding of Digital Citizenship beyond conventional digital literacy frameworks by emphasizing the importance of responsible participation within AI-mediated educational environments [28], [33].

The observed dominance of psychological variables is also consistent with established behavioral theories, including the Technology Acceptance Model (TAM), the Unified Theory of Acceptance and Use of Technology (UTAUT), and the Theory of Planned Behavior (TPB). These theoretical frameworks emphasize that attitudes, perceived behavioral control, self-efficacy, and individual beliefs substantially influence technology adoption and responsible technology use. The present findings extend these theories by demonstrating that psychological readiness toward Generative AI involves nonlinear and context-dependent behavioral mechanisms that can be effectively captured through explainable ensemble learning models rather than conventional linear analytical approaches.

Second, the findings demonstrate that psychological adaptation toward AI technologies may involve nonlinear behavioral dynamics rather than purely linear relationships. The identified inflection points associated with technology anxiety indicate that moderate levels of anxiety may encourage reflective and responsible AI usage, whereas excessive anxiety may weaken ethical participation and digital engagement. This observation suggests that psychological adaptation to AI technologies is context-dependent and may evolve throughout the AI adoption process [35].

Self-Efficacy also emerged as one of the dominant predictors across all evaluated machine learning models. This finding is consistent with Social Cognitive Theory, which argues that individuals possessing stronger confidence in their capabilities are more likely to engage in adaptive learning behaviors and responsible technology use. Students with higher AI-related self-efficacy are therefore better equipped to critically evaluate AI-generated information, use Generative AI ethically, and integrate AI technologies into academic activities without excessive dependence on automated recommendations.

Third, this study demonstrates the usefulness of Explainable Artificial Intelligence (XAI) for educational analytics. By integrating ensemble learning models with SHAP-based explainability analysis, the proposed framework provides not only predictive capability but also interpretable insights into the behavioral mechanisms underlying responsible AI participation. Such transparency is particularly important for supporting trustworthy educational analytics and evidence-based decision making [27], [34].

From an academic perspective, this study contributes to the emerging field of Explainable Artificial Intelligence in educational analytics by demonstrating that behavioral prediction should not rely solely on predictive accuracy but also on model transparency and interpretability. The integration of ensemble learning with SHAP-based explainability enables researchers to uncover hidden nonlinear behavioral mechanisms that remain difficult to identify using traditional statistical approaches such as linear regression or Structural Equation Modeling. Consequently, the proposed framework contributes both methodological advancement and theoretical insight into Digital Citizenship research within AI-mediated educational environments.

From a practical perspective, the findings indicate that psychological readiness toward AI technologies may play a more influential role than technological accessibility alone in shaping responsible Digital Citizenship behavior among university students. Higher education institutions should therefore prioritize AI literacy programs that strengthen students' ethical awareness, self-efficacy, emotional readiness, critical thinking, and responsible AI engagement rather than focusing exclusively on technological accessibility [29], [31].

The identified nonlinear relationship between technology anxiety and Digital Citizenship behavior further highlights the importance of balanced educational interventions. Moderate levels of technology anxiety may encourage more cautious and reflective AI interaction, whereas excessive anxiety may reduce confidence and discourage meaningful digital participation. This finding suggests that universities should promote balanced AI adaptation strategies that combine technological competence with emotional readiness and ethical awareness [34], [35].

In addition, explainable educational analytics may support more transparent, accountable, and evidence-based AI governance frameworks within higher education environments. Universities may utilize psychological readiness indicators, such as technology anxiety and self-efficacy, to identify students who require additional support in adapting to AI-assisted learning systems and responsible AI practices. Furthermore, the dominance of psychological variables observed in this study emphasizes the importance of human-centered AI education policies that integrate digital ethics, emotional adaptation, responsible AI usage, and critical AI literacy into university curricula [30], [33].

Overall, the findings demonstrate that Explainable Artificial Intelligence can function not only as a predictive framework but also as a human-centered analytical approach for understanding responsible Digital Citizenship behavior in the era of Generative AI. By integrating predictive performance with transparent behavioral interpretation, the proposed framework supports trustworthy educational analytics, evidence-based AI governance, and responsible AI adoption within higher education ecosystems.

5. Conclusion

This study proposed an Explainable Ensemble Learning framework that integrates XGBoost, Random Forest, Support Vector Regression, and SHAP-based Explainable Artificial Intelligence (XAI) to identify the key determinants of Digital Citizenship among university students using Generative Artificial

Intelligence (Generative AI). The proposed framework represents a methodological contribution by combining predictive modeling with explainable machine learning, enabling both accurate prediction and transparent interpretation of behavioral factors within AI-assisted educational environments. Among the evaluated models, XGBoost achieved the best predictive performance, while SHAP analysis consistently identified Technology Anxiety, Attitude Toward Behavior, and Self-Efficacy as the most influential predictors, with nonlinear analysis further revealing that Technology Anxiety exhibits context-dependent effects in shaping responsible AI participation. Robustness, fairness, and ablation analyses confirmed the reliability, stability, and trustworthiness of the proposed framework across multiple validation scenarios. These findings provide practical implications for universities, educators, and policymakers by highlighting the importance of strengthening AI literacy, ethical awareness, self-efficacy, emotional readiness, critical thinking, and responsible AI engagement to support evidence-based educational analytics and transparent AI governance. Nevertheless, this study is limited by its reliance on a cross-sectional survey of Indonesian university students and self-reported behavioral measures, which may restrict the generalizability of the findings and prevent causal interpretation. Future research should validate the proposed framework using longitudinal and cross-cultural datasets, incorporate additional contextual and multimodal educational variables, and explore causal machine learning approaches, advanced Explainable AI techniques, and Large Language Model-assisted educational analytics to further enhance the assessment of Digital Citizenship in AI-enhanced learning environments.

Credit Authorship Contribution Statement

Muchammad Thoha: Conceptualization, Methodology, Software, Data curation, Formal analysis, Visualization, Writing – original draft. **Andy Prasetyo Utomo**: Supervision, Methodology, Validation, Writing – review & editing. **Zainur Romadhon**: Investigation, Validation, Resources, Writing – review & editing.

Declaration of Competing Interests

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.

Data Availability

The dataset analyzed in this study is publicly available in the Mendeley Data repository at DOI: 10.17632/h4gcfmbk5c.3. This study used the dataset solely for secondary data analysis. No new data were generated during this study.

Use of Artificial Intelligence

The authors used Generative Artificial Intelligence (ChatGPT, OpenAI) solely to assist with language refinement, grammar checking, and improving manuscript readability. All scientific content, research design, data analysis, interpretation of the results, and conclusions were developed, verified, and approved by the authors. The authors take full responsibility for the content of this manuscript.

References

[1] M. Bond, O. Zawacki-Richter, and M. Marin, "A Meta Systematic Review of Artificial Intelligence in Higher Education: A Call for Increased Ethics, Collaboration, and Rigour," *Int. J. Educ. Technol. High. Educ.*, vol. 21, 2024, doi: 10.1186/s41239-019-0171-0.

[2] E. Kasneci, "ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education," *Learn. Individ. Differ.*, vol. 103, 2023.

[3] A. Dabis, "AI and Ethics: Investigating the First Policy Responses of Higher Education Institutions to the Challenge of Generative AI," *Humanit. Soc. Sci. Commun.*, 2024, doi: 10.1145/2939672.2939785.

[4] S. Cotton, "Chatting and Cheating: Ensuring Academic Integrity in the Era of ChatGPT," *Innov. Educ. Teach. Int.*, vol. 61, no. 2, pp. 228–239, 2024.

[5] O. Zawacki-Richter, "Systematic Review of Research on Artificial Intelligence Applications in Higher Education – Where Are the Educators?," *Int. J. Educ. Technol. High. Educ.*, 2019, doi: 10.1186/s41239-019-0171-0.
[6] M. Farrokhnia, "A SWOT Analysis of ChatGPT: Implications for Educational Practice and Research," *Innov. Educ. Teach. Int.*, vol. 61, no. 3, pp. 460–474, 2024.
[7] A. B. Arrieta, "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, 2020.
[8] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," 2016, doi: 10.1145/2939672.2939785.
[9] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," 2017.
[10] Y. K. Dwivedi, "So What if ChatGPT Wrote It?," *Int. J. Inf. Manage.*, 2023, doi: 10.1016/j.ijinfomgt.2023.102642.
[11] A. Tlili, "What if the Devil Is My Guardian Angel," *Smart Learn. Environ.*, 2023, doi: 10.1186/s40561-023-00237-x.
[12] D. Long, "What Is AI Literacy? Competencies and Design Considerations," 2020. doi: 10.1145/3313831.3376727.
[13] C. Meske, "Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities," *Inf. Syst. Manag.*, vol. 39, no. 1, pp. 53–63, 2022.
[14] H. Khosravi, "Explainable Artificial Intelligence in Education," *Comput. Educ. Artif. Intell.*, vol. 3, 2022, doi: 10.1016/j.caeai.2022.100074.
[15] D. V Carvalho, "Machine Learning Interpretability: A Survey on Methods and Metrics," *Electronics*, 2019, doi: 10.3390/electronics8080832.
[16] M. M. M. Abbad, "Using the UTAUT Model to Understand Students' Usage of E-Learning Systems in Developing Countries," *Educ. Inf. Technol.*, vol. 26, pp. 7205–7224, 2021.
[17] L. Xue and J. Churchill, "The Unified Theory of Acceptance and Use of Technology (UTAUT) in Higher Education: A Systematic Review," *SAGE Open*, vol. 14, no. 1, 2024.
[18] A. Adadi, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," *IEEE Access*, 2018, doi: 10.1109/ACCESS.2018.2870052.
[19] C. Molnar, *Interpretable Machine Learning*. 2022.
[20] N. Selwyn, *Should Robots Replace Teachers? AI and the Future of Education*. 2019.
[21] W. Holmes, *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. 2019.
[22] M. Ribble, *Digital Citizenship in Schools*. 2011.
[23] K. Mattson, "Digital Citizenship Education," 2018. doi: 10.4324/9781315628110-13.
[24] A. Geron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. 2022.
[25] UNESCO, "Guidance for Generative AI in Education and Research," 2023.
[26] Y. A. Prasetiyo, "Dataset of Digital Citizenship, Technology Anxiety, Self-Efficacy, Attitude Toward Behavior, and Generative Artificial Intelligence Usage Among Indonesian University Students," 2025, *Mendeley Data, Amsterdam, Netherlands*. doi: 10.17632/h4gcfmbk5c.3.
[27] D. Lee, "The Impact of Generative AI on Higher Education Learning and Teaching: A Study of Educators' Perspectives," *Comput. Educ. Artif. Intell.*, 2024, doi: 10.1016/j.caeai.2024.100221.
[28] T. K. Ng, "Conceptualizing AI Literacy: An Exploratory Review," *Comput. Educ. Artif. Intell.*, 2021, doi: 10.1016/j.caeai.2021.100041.
[29] R. Luckin, *Intelligence Unleashed: An Argument for AI in Education*. 2016.
[30] F. Ouyang, "Artificial Intelligence in Education: The Three Paradigms," *Comput. Educ. Artif. Intell.*, 2021, doi: 10.1016/j.caeai.2021.100020.
[31] X. Zhai, "A Review of Artificial Intelligence (AI) in Education from 2010 to 2020," *Complexity*, 2021, doi: 10.1155/2021/8812542.
[32] A. Bandura, *Self-Efficacy in Changing Societies*. 1995.
[33] O. Almatrafi, "A Systematic Review of AI Literacy Conceptualization, Constructs, and Implementation and Assessment Efforts (2019–2023)," *Comput. Educ. Open*, 2024.
[34] W. Holmes, "State of the Art and Practice in AI in Education," *Eur. J. Educ.*, 2022, doi: 10.1111/ejed.12533.
[35] M. Giannakos, "The Promise and Challenges of Generative AI in Education," *Behav. Inf. Technol.*, 2025, doi: 10.1080/0144929X.2024.2394886.