International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 10 | Oct 2026 www.irjet.net p-ISSN: 2395-0072

"A Comparative Analysis of Student Trust and Acceptance of a Proposed Explainable AI-Based Academic Performance Prediction System Among Colleges in Navi Mumbai with Reference to "DY Patil University and Bharati Vidyapeeth"

Harshdeep Singh[1], Chaitannya Khot[1], Dr. Priya Vij[2]

[1] *Department of MCA, D. Y. Patil College of Humanities and Sciences, D. Y. Patil Deemed to be University, Navi Mumbai, Maharashtra, India*

[2] *Associate Professor, D. Y. Patil Deemed to be University, School of Humanities and Sciences, Belapur, Maharashtra, India.*

Abstract-*Systems that flag students at academic risk can help staff offer timely support, yet they achieve little unless students find the flags believable and are willing to respond. This paper investigates trust and behavioural intention toward a proposed academic performance prediction system based on explainable artificial intelligence (XAI), drawing on students of DY Patil University and Bharati Vidyapeeth in Navi Mumbai. Through a cross-sectional online survey, every respondent was shown an identical fictional high-risk student case, with the warning explained by attendance, assignment completion and earlier assessment marks and paired with safeguards: adviser review and an assurance that grades would not depend on the output. Of the 110 submissions received, 101 respondents consented and formed the analytic sample (56 from DY Patil University and 45 from Bharati Vidyapeeth), and the 9 submissions lacking consent were dropped. Trust was computed as the average of three seven-point items and intention to use as the average of two seven-point items. The analysis consisted of descriptive statistics, one-sample t-tests against the neutral midpoint, Welch independent-samples t-tests, Cronbach's alpha and Spearman's rank correlation. The mean trust score was 4.21 out of 7, not significantly different from the midpoint, $t(100) = 1.43$, $p = .155$, while the mean intention score of 4.42 was significantly above the midpoint, $t(100) = 2.84$, $p = .005$. Neither measure differed significantly across the two institutions, and the two were strongly positively correlated (Spearman's $\rho = .866$, $p < .001$). These results capture self-reported perceptions of one fictional scenario and say nothing definitive about real adoption, prediction accuracy or academic improvement.*

Keywords: Explainable artificial intelligence, Student trust, Behavioural intention, Academic risk prediction, Higher education.

INTRODUCTION

Predicting academic performance is widely seen as a way to find learners who need extra help. Educational analytics can draw on attendance, submitted work, test results and activity on learning platforms to gauge how likely a student is to fall behind, so that help can arrive sooner. Accuracy alone, though, does not make a prediction worth using. Students and teachers must also be able to see why a warning appeared, judge how confident it is, and know how it can be questioned or reviewed.

Explainable artificial intelligence (XAI) covers approaches that make an AI system's output intelligible to people. Techniques such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) estimate the contribution each input feature makes to a model's behaviour, across the whole model or for one prediction. Recent educational data mining studies have increasingly combined machine-learning models with these techniques to make results easier to interpret [1]–[3]. Even so, a model that is technically interpretable does not ensure that students will understand the explanation, believe the prediction or act on it.

This study accordingly examines how students perceive a proposed student-facing academic risk system, rather than how accurately a trained model predicts. All respondents considered one fictional case: a student with 62% attendance, six of ten assignments submitted and a previous assessment average of 45%. The case labelled the student as at high risk of failing at least one upcoming subject, listed the factors behind the warning, conceded that the prediction could be mistaken, and noted that an academic adviser would review the warning and that it would have no bearing on grades.

Responses from two institutional groups, DY Patil University and Bharati Vidyapeeth, are compared on trust, stated intention to use the system, and the relationship between the two. The comparison is exploratory, since the survey cannot show that respondents are representative of the wider student bodies at either institution.

OBJECTIVES

1. To describe how much students trust the system, based on their confidence, trustworthiness and reliability ratings.

2. To measure students' stated intention to use the system and their expected use of academic-risk feedback.

3. To compare trust and intention scores across the two institutional groups.

4. To examine how trust and intention to use are related.

5. To summarise the concerns respondents raised about accuracy, privacy, fairness, context and human oversight.

LITERATURE REVIEW

Explainable AI for educational prediction -Islam et al. [1] trained conventional and boosting classifiers to predict student academic performance and applied several explanation tools, among them SHAP, Shapash, ELI5 and LIME. XGBoost was the strongest model, with 83% accuracy, and the authors described how its outputs could inform recommendations for supporting students. This demonstrates that prediction and explanation can be brought together in practice; however, the study concentrates on model performance and the generation of explanations, and does not investigate how students react to those explanations.is model performance and explanation generation; it does not examine how students themselves respond to the explanations.

XAI models and performance metrics -Johora et al. [2] worked with 872 undergraduate records, compared seven machine-learning classifiers against a custom stacking ensemble, and interpreted the predictions with SHAP and LIME; the ensemble reached 86.38% accuracy. Findings of this kind show that explainable prediction is technically feasible, but accuracy says little about whether students would accept or trust a system built on such a model.

Stability and changing learning contexts- Tiukhova et al. [3] examined how stable student-success prediction models remain when learning contexts change, and reported that the importance given to individual features can move as the context shifts; deployed models therefore need continuing monitoring. This is relevant to tools aimed at students, since an explanation cannot be presumed to stay valid when the student population, the course design or learning behaviour changes.

Transparency, trust and acceptance -Work on algorithmic transparency suggests that revealing more information does not by itself raise trust; the outcome partly turns on whether the algorithm's result met users' expectations [4]. Shin [5] likewise distinguishes an explanation being on offer from it being understood, and studies how explainability and causability shape users' trust in and acceptance of AI-driven services.

Study	Method / technology	Data or setting	Reported contribution	Limitation for the present study
Islam et al. (2025) [1]	Decision tree, random forest, gradient boosting, XGBoost; SHAP, LIME and related tools	Public student-performance dataset	XGBoost reached 83% accuracy; explanations gave feature-level interpretation	Technical study; no assessment of student trust
Johora et al. (2025) [2]	Seven classifiers and a stacking ensemble; SHAP and LIME	872 undergraduate records	Stacking ensemble reached 86.38% accuracy	Accuracy does not indicate trust or real-world adoption
Tiukhova et al. (2024) [3]	Stability analysis of prediction models using XAI	Student-success prediction across learning contexts	Feature importance can change with context; models need monitoring	Does not address student trust or institutional comparison
Kizilcec (2016) [4]	Transparency manipulation in an algorithmic interface	Online peer-assessment course	Trust depended on context and on whether outcomes met expectations	Different application; this survey uses one fixed scenario
Shin (2021) [5]	Analysis of explainability, causability, trust and acceptance	AI-driven service context	Separates having an explanation from understanding it	This survey records recognition,notdemonstrated comprehension

Note: Because the reviewed studies use different datasets, algorithms and outcomes, their accuracy figures cannot be compared with each other or with the present study, which does not train or evaluate any predictive model.

RESEARCH METHODOLOGY

Research design

Type of research: The research is quantitative, descriptive and comparative, with an exploratory component. It takes a cross-sectional survey form: a single online questionnaire completed at one point in time, where ratings were analysed statistically and one open-ended item was summarised by theme. Because the analysis is exploratory, causal conclusions cannot be drawn.

This is a quantitative, cross-sectional, comparative survey with one open-ended feedback item. Respondents rated a shared fictional scenario at one point in time; no follow-up, intervention, real system use or academic outcome was observed.

Primary data

The authors gathered primary data themselves through an online Google Forms questionnaire answered by students of DY Patil University and Bharati Vidyapeeth in Navi Mumbai. The form captured institution, programme, year of study, familiarity with AI, five seven-point ratings (confidence, trust, reliability, intention to use and expected use), recognition of the explanation, and open-ended comments.

Secondary data

Secondary data came from the published journal and conference sources cited in the Literature Review [1]–[7], which cover explainable AI for educational prediction, the stability of prediction models, and transparency, trust and acceptance, including trust in e-learning recommendations. They supplied background and context, such as the accuracy figures reported in [1] and [2]. They were neither pooled nor re-analysed, and no existing dataset entered the statistical analysis.

Source of secondary data. The secondary sources were published in Telematics and Informatics Reports [1], Array [2], Decision Support Systems [3], the proceedings of ACM CHI 2016 [4], the International Journal of Human-Computer Studies [5] and the proceedings of ACM IUI 2022 and 2023 [6], [7]. Each is identified by its DOI in the References.

Sample size

Data were gathered with an online Google Forms questionnaire that began with an age and consent screening item. In total the form received 110 submissions. Of these, 101 respondents confirmed they were 18 or older and agreed voluntarily to take part, whereas 9 did not consent. The 9 non-consenting submissions were removed from every analysis, leaving an analytic sample of 101 respondents: 56 from DY Patil University and 45 from Bharati Vidyapeeth (Figure 1).

Sampling method

The number of students invited and the sampling frame were not recorded, so the sample is regarded as a non-probability convenience sample and a response rate cannot be computed. The identity and uniqueness of respondents were not independently checked.

Scenario and measures

The scenario explained a fictional high-risk prediction using attendance (62%), assignments submitted (6 of 10) and previous assessment average (45%). It also said that the prediction could be wrong, would be reviewed by an academic adviser and would not decide grades. Confidence in the system, trust in it and its perceived reliability were each measured with a seven-point item, and intention to use the system and expected use were measured with two further seven-point items. A recognition question asked whether the example explicitly explained why the prediction was made, and an open-ended item invited concerns or suggestions.question asked whether the example explicitly explained why the prediction was produced, and an open-ended item invited concerns or suggestions.

Variable construction

Trust was calculated as the mean of the three trust-related items, and intention to use as the mean of the two intention-related items; higher values denote more positive ratings. The neutral midpoint of the seven-point scale is 4. The items are study-specific measures, so the results do not establish full construct validity.

Statistical analysis

Means, standard deviations and medians were computed for the five rating items and the two composites. One-sample t-tests set each composite mean against the neutral midpoint of 4, and Welch independent-samples t-tests compared the two institutional groups. Cronbach's alpha was computed for the three trust items. Because the underlying items are ordinal ratings, Spearman's rank correlation was used for the trust–intention association.

The five inferential tests were not formally corrected for multiplicity; still, the two significant results ($p = .005$ and $p < .001$) would stay significant under a conservative Bonferroni threshold of .01, and the other tests were non-significant. All analyses are descriptive and exploratory rather than causal.

Hypotheses and Objectives

No.	Objective	Null hypothesis (H0)	Alternative hypothesis (H1)
1	To assess whether students' overall trust in the proposed explainable AI-based academic risk system differs from a neutral rating.	$H0_1$: The population mean trust score equals 4 (the neutral midpoint).	$H1_1$: The population mean trust score differs from 4.
2	To assess whether students' stated intention to use the system differs from a neutral rating.	$H0_2$: The population mean intention-to-use score equals 4 (the neutral midpoint).	$H1_2$: The population mean intention-to-use score differs from 4.
3	To compare trust between students of DY Patil University and Bharati Vidyapeeth.	$H0_3$: Mean trust does not differ between the two institutional groups.	$H1_3$: Mean trust differs between the two institutional groups.
4	To compare intention to use between students of the two institutions.	$H0_4$: Mean intention to use does not differ between the two institutional groups.	$H1_4$: Mean intention to use differs between the two institutional groups.
5	To examine the association between trust and intention to use.	$H0_5$: There is no monotonic association between trust and intention to use.	$H1_5$: Trust and intention to use are associated.

All five hypotheses are tested at $\alpha = .05$. H_{01} and H_{02} use one-sample comparisons against the neutral midpoint; H_{03} and H_{04} use Welch's t-tests to compare the institutional groups; H_{05} uses Spearman's rank correlation. As the sample is one of convenience, these tests are exploratory and their results should be read with that limitation in mind.

Data Analysis and Interpretation

Sample characteristics

Figure 1 presents programme composition by institution together with self-reported AI familiarity in this sample. MCA students formed the majority (59.4%), and 79 of 101 respondents (78.2%) rated themselves as moderately, very or extremely familiar with AI. Programme information was unclear or missing for 27 respondents (26.7%), and the programme mix differed between the institutions, so differences between institutions cannot be fully separated from differences in programme composition.

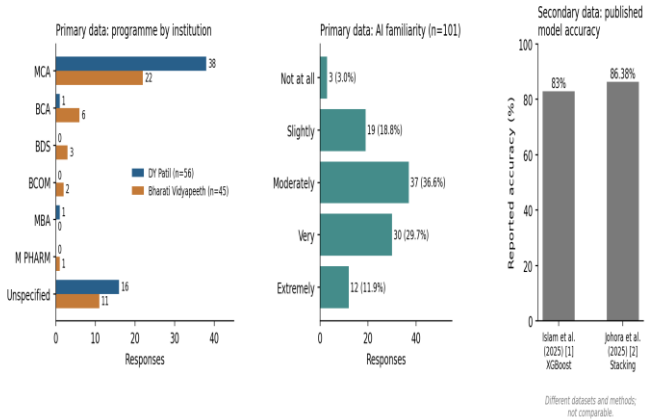


Figure 1. Programme composition by institution and self-reported AI familiarity (analytic sample, n = 101).

Figures 2 to 6 reproduce the unfiltered Google Forms summaries, so they include all 110 submissions. Figure 2 gives the outcome of the consent screening: 101 respondents (91.8%) consented while 9 (8.2%) did not. Figures 3 to 5 present institution, year of study and AI familiarity for all submissions, and Figure 6 displays how responses to the confidence item were distributed. Every inferential result in the sections on item-level statistics, hypothesis tests and internal consistency rests only on the 101 consenting respondents. Second-year students were the biggest year-of-study group (66.4% of all submissions).

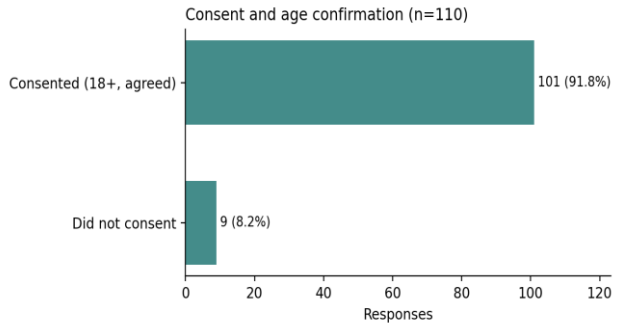


Figure 2. Consent and age confirmation (all submissions, n = 110).

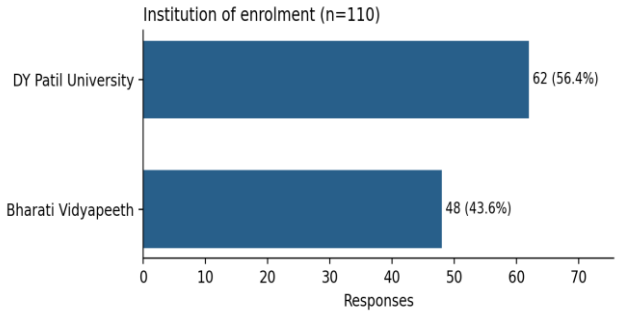


Figure 3. Institution of enrolment (all submissions, n = 110).

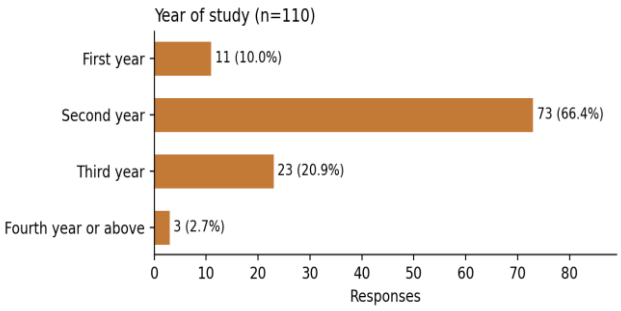


Figure 4. Year of study (all submissions, n = 110).

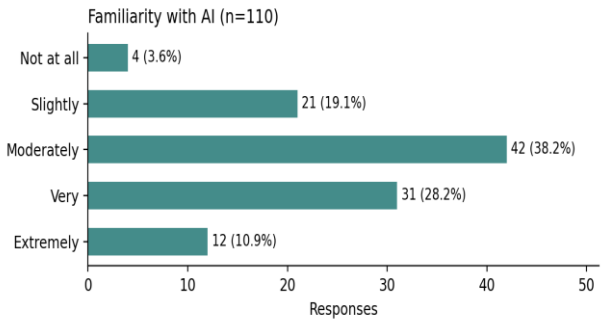


Figure 5. Self-reported familiarity with AI (all submissions, n = 110).

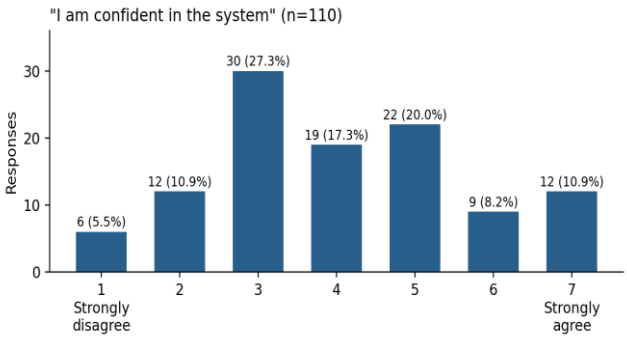


Figure 6. Distribution of ratings for "I am confident in the system" (all submissions, n = 110; 1 = strongly disagree, 7 = strongly agree).

Item-level descriptive statistics

Table 1. Descriptive statistics for the five rating items (n = 101)

Item	Mean	SD	Median
Confidence	4.08	1.70	4.0
Trust	4.22	1.54	4.0
Reliability	4.35	1.60	4.0
Intention to use	4.30	1.51	4.0
Expected use	4.54	1.58	5.0

Note. All items were rated on a seven-point scale (1 = lowest, 7 = highest).

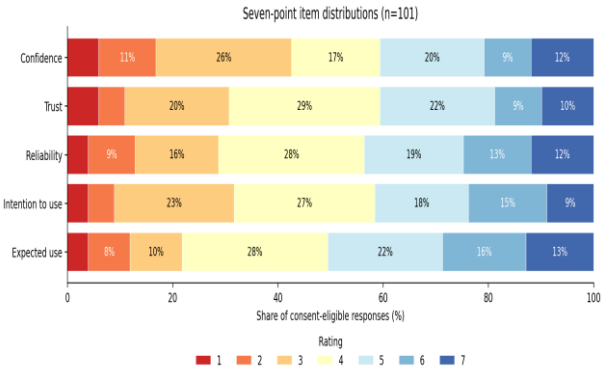


Figure 7. Distribution of ratings for the five seven-point items (analytic sample, n = 101).

Each of the five item means sits slightly above the neutral midpoint, and expected use is the highest. The large standard deviations indicate that opinions were divided, so

a mean above 4 should not be taken as general approval of the proposed system.

Composite scores and hypothesis tests

Table 2. Results of hypothesis tests

Test	Result	p	Decision (α = .05)
Trust vs. neutral midpoint (4)	M = 4.215, SD = 1.505; t(100) = 1.432	.155	Fail to reject H_{01}
Intention vs. neutral midpoint (4)	M = 4.421, SD = 1.488; t(100) = 2.842	.005	Reject H_{02}
Trust: DY Patil vs. Bharati Vidyapeeth	4.292 vs. 4.119; Welch t = 0.567	.572	Fail to reject H_{03}
Intention: DY Patil vs. Bharati Vidyapeeth	4.509 vs. 4.311; Welch t = 0.662	.510	Fail to reject H_{04}
Trust–intention association	Spearman's ρ = .866	< .001	Reject H_{05}

Mean trust was not significantly different from the neutral midpoint. Intention to use was significantly above the midpoint, though the gap is small in absolute terms and reflects stated intention rather than observed behaviour. Neither comparison between institutions reached significance. The strong positive association means that respondents with higher trust also tended to state stronger intention to use the system; like any correlation, it does not show that trust causes intention.

Internal consistency

The three trust items gave a Cronbach's alpha of approximately .927 in the analytic sample, which indicates high internal consistency among them. This alone does not show that the scale is valid, unidimensional or generalisable, and similar item wording and shared response tendencies may inflate the coefficient. No reliability coefficient was calculated for the two-item intention composite.

Explanation recognition and open-ended feedback

Among the 101 respondents, 63 noticed that the scenario gave reasons for the prediction. Noticing an explanation is not the same as understanding it: the questionnaire did not check whether respondents grasped how the prediction was produced, and the example was fictional, not the output of a trained model.

The open-ended comments mixed perceived benefits with concerns. Some respondents considered early identification useful for targeting support, while others questioned the accuracy and fairness of such predictions, raised privacy worries, or doubted that attendance alone reflects learning. Suggestions included review by an adviser, transparent reasons for every warning, a chance for students to explain personal circumstances, protection of student data and bias checks. Many entries were very short, unclear or placeholders, so the comments illustrate themes and cannot be used to estimate how common each concern is.

Supplementary exploratory cluster analysis

An exploratory K-means analysis was run on the five standardised rating items, trying two to five clusters. The highest silhouette score, 0.463, occurred at k = 2, and the first two principal components accounted for 90.4% of the standardised item variance. The two clusters largely distinguish lower from higher ratings. This is an unsupervised illustration of response patterns, not a student-risk model, a validated typology or evidence of institutional categories. The same records were used to choose and to display the clusters, and no independent validation was carried out.

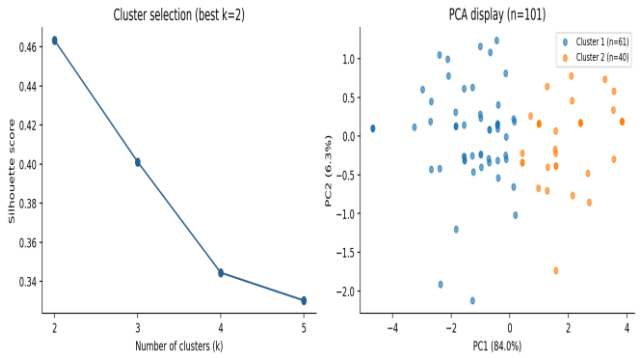


Figure 8. Exploratory K-means cluster selection and PCA display (analytic sample, n = 101).

Findings

1. Of the 110 form submissions, 101 consenting respondents were analysed, 56 from DY Patil University and 45 from Bharati Vidyapeeth.

2. The mean trust score was 4.21 out of 7, not significantly different from the neutral midpoint (p = .155).

3. The mean intention to use was 4.42 out of 7, significantly above the neutral midpoint (p = .005).

4. No significant difference between institutions emerged for trust (p = .572) or intention to use (p = .510).

5. Trust and intention showed a strong positive association (Spearman's ρ = .866, p < .001).

6. The trust items were highly consistent with one another (Cronbach's α ≈ .927), although construct validity was not established.

LIMITATIONS

1. The 101 consenting respondents form a non-probability convenience sample, so the results cannot be treated as representative estimates for either university, and no response rate can be computed.

2. Programme information was unclear for 27 respondents (26.7%), and the groups differ in programme composition, which confounds the comparison between institutions.

3. Every respondent rated the same fixed, fictional scenario, so the findings may not transfer to other cases or to a real, trained prediction system.

4. Trust and intention are self-reported perceptions and do not demonstrate real use or adoption of the system.

5. The trust and intention measures were designed for this study, the intention composite contains only two items, and construct validity was not established.

6. The exploratory cluster analysis lacks independent validation and should be treated only as an illustration of response patterns.

CONCLUSION

This study presents an exploratory comparison of how students at DY Patil University and Bharati Vidyapeeth perceive a proposed explainable AI-based academic-risk prediction system. Among 101 consenting respondents, mean trust was 4.21 out of 7 and not significantly different from the neutral midpoint, whereas mean intention to use was 4.42 and significantly above it. The two institutions did not differ significantly on either measure, and trust and intention were strongly related. The results suggest some willingness to use explained risk feedback, but trust stays mixed rather than uniformly positive. They cannot establish real adoption, model accuracy, the causal effect of explainability or better academic outcomes. The convenience sample, unclear programme information and dependence on one fictional scenario restrict how widely the findings can be generalised.

REFERENCES

1. Islam, M. M., Sojib, F. H., Mihad, M. F. H., Hasan, M., & Rahman, M. (2025). The integration of explainable AI in educational data mining for student academic performance prediction and support system. Telematics and Informatics Reports, 18, Article 100203. https://doi.org/10.1016/j.teler.2025.100203

2. Johora, F. T., Hasan, M. N., Rajbongshi, A., Ashrafuzzaman, M., & Akter, F. (2025). An explainable AI-based approach for predicting undergraduate students' academic performance. Array, 26, Article 100384. https://doi.org/10.1016/j.array.2025.100384

3. Tiukhova, E., Vemuri, P., López Flores, N., Islind, A. S., Óskarsdóttir, M., Poelmans, S., & Snoeck, M. (2024). Explainable learning analytics: Assessing the stability of student success prediction models by means of explainable AI. Decision Support Systems,182,Article114229. https://doi.org/10.1016/j.dss.2024.114229

4. Kizilcec, R. F. (2016). How much information? Effects of transparency on trust in an algorithmic interface. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 2390–2395). Association for Computing Machinery. https://doi.org/10.1145/2858036.2858402

5. Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. International Journal of Human-Computer Studies,146, Article102551. https://doi.org/10.1016/j.ijhcs.2020.102551