Artificial Intelligence and Machine Learning in Higher Education: Evaluating Strategies to Reduce Academic Disparities

Ramesh Iyer[1], Kavitha Sundaram[2*]

[1]Department of Educational Technology & AI, IIT Bombay, Maharashtra, India

[2]School of Computer Science & Learning Sciences, University of Hyderabad, Telangana, India

Abstract

Academic disparities — the systematic divergence in educational outcomes across student populations differentiated by socioeconomic background, geographic region, first-generation status, gender, and ethnicity — remain among the most persistent structural inequities in global higher education. Despite decades of institutional policy and targeted student support investment, the GPA gap between the highest- and lowest-risk student quintiles in large universities has narrowed by less than 8% over the past two decades using conventional approaches. Artificial Intelligence (AI) and Machine Learning (ML) now offer a fundamentally new toolkit for addressing this challenge: predictive modelling capable of early disparity identification, adaptive systems that personalise learning at scale, and automated intervention workflows that route support resources with a speed, precision, and continuity that human-only systems cannot sustain. This paper presents a comprehensive evaluation of five AI and ML strategy categories deployed across 36,400 students at six higher education institutions spanning India, Ghana, and Mexico over four academic years (AY 2022–26), constituting the longest and largest multi-site longitudinal evaluation of AI-ML disparity reduction strategies in the JSTST literature.

Keywords: Artificial Intelligence, Machine Learning, Higher Education, Academic Disparities, Early Warning Systems, Adaptive Learning, Predictive Analytics

Introduction

Higher education has long been positioned as the engine of social mobility — the institutional mechanism through which talent and effort, irrespective of background, are recognised and rewarded. The empirical reality, however, is considerably more complex. Academic disparity — the systematic gap in educational outcomes between student populations stratified by socioeconomic status, geographic origin, first-generation university attendance, gender, ethnicity, and disability status — persists across virtually every higher education system in the world, defying decades of policy investment, widening access initiatives, and student support programmes. The UNESCO Global Education Monitoring Report 2026 estimates that first-generation university students remain 2.7 times more likely to withdraw before completing their degree than continuing-generation peers after controlling for entry qualification; that students from the lowest socioeconomic quintile achieve mean GPAs 0.84 grade points below the institutional mean; and that rural-origin students in South Asia, Sub-Saharan Africa, and Latin America face dropout rates 34–41% higher than their urban counterparts in equivalent programmes.

These statistics are not merely sociological observations — they represent concrete failures of institutional systems to identify at-risk students early enough, target support resources accurately enough, and sustain intervention intensity long enough to interrupt the disparity trajectories that crystallise within the first eight weeks of each academic semester. Conventional student support systems are systemically outmatched by the scale and complexity of the challenge: human advisors cannot monitor 36,000 students weekly; financial aid offices cannot predict disbursement-delay stress six weeks before it disrupts academic performance; learning support teams cannot identify which of 400 module enrolees are silently disengaging from their learning trajectory before it becomes visible in a grade. Artificial Intelligence and Machine Learning can do all three simultaneously, continuously, and

Corresponding Author: Kavitha Sundaram, School of Computer Science & Learning Sciences, University of Hyderabad, Telangana, India

How to cite this article: Iyer, R., Sundaram, K. (2026). Artificial Intelligence and Machine Learning in Higher Education: Evaluating Strategies to Reduce Academic Disparities. *Journal of Science, Technology and Social Transformation* 2(2), 52-61.

Source of support: Nil

Conflict of interest: None

at zero marginal cost per additional student.

This paper presents a comprehensive four-year longitudinal evaluation of five AI-ML strategy categories for academic disparity reduction across 36,400 students at six institutions in India, Ghana, and Mexico — the longest, geographically broadest. The study does not merely evaluate whether AI-ML strategies work — prior literature has established that they do in controlled single-institution settings. It evaluates which strategies work, how much they work, for whom they work, whether their benefits accumulate over time, and whether they can be deployed equitably across diverse demographic populations and institutional contexts. These are the questions that determine whether AI-ML educational equity strategies can be responsibly adopted at national and institutional scale — and they have been insufficiently answered in the existing literature.

Scope and Contribution

This study makes five contributions to the educational AI-ML literature. First, it provides the longest published longitudinal evaluation (4 academic years, 8 semesters) of integrated AI-ML disparity reduction strategies, enabling assessment of trajectory dynamics and compounding effects unavailable from shorter studies. Second, it delivers the first systematic three-continent comparative evaluation (India, Ghana, Mexico), testing cross-cultural and cross-institutional generalisability. Third, it introduces an eight-dimension Algorithmic Equity Framework — extending existing binary fairness metrics to encompass gender, socioeconomic, regional, first-generation, disability, ethnic, age, and digital literacy dimensions simultaneously. Fourth, it presents EduBERT-v2, an updated domain-adaptive transformer model pre-trained on 4.8 million educational text samples, achieving AUC-ROC of 0.949 as a single model. Fifth, it provides a replicable institutional deployment blueprint — including governance protocols, SAP SLCM integration architecture, and intervention workflow designs — for institutions seeking to deploy AI-ML equity strategies beyond controlled research contexts.

Research Questions

- RQ1: What is the comparative effectiveness of five AI-ML strategy categories in reducing academic failure rates and GPA disparity gaps across a four-year deployment period?
- RQ2: Do the benefits of integrated AI-ML strategies accumulate and compound over multiple semesters, or do they plateau after initial deployment?
- RQ3: Which student data features most strongly predict academic disparity trajectories, and do these features reflect genuine structural disadvantage mechanisms?
- RQ4: Does the AI-ML ensemble model architecture achieve superior predictive performance to any constituent single-model approach, and what is the contribution of the EduBERT-v2 language model?
- RQ5: Are AI-ML strategy benefits equitably distributed across gender, socioeconomic, regional, first-generation, disability, ethnic, age, and digital literacy demographic subgroups?
- RQ6: What institutional governance conditions are necessary for AI-ML disparity reduction strategies to operate effectively and ethically at scale?

Literature Review

The Structural Roots of Academic Disparity

Academic disparity is not random variation in student performance — it is the educational expression of structural social inequality. Tinto's seminal persistence model (1987, revised 2017) identifies academic integration, social integration, and institutional commitment as the three pillars of student retention, each systematically undermined for students from socioeconomically disadvantaged backgrounds. Bean and Metzner's model (1985) adds the role of external environmental variables — employment obligations, family caring responsibilities, financial precarity — that disproportionately burden first-generation and low-income students. More recent scholarship has emphasised the role of 'hidden curriculum' disadvantage — the tacit institutional knowledge about academic norms, resource navigation, and self-advocacy that continuing-generation students absorb from family experience but first-generation students must discover independently, often at significant academic cost.

Evolution of AI-ML in Educational Analytics: 2015–2026

The application of ML to educational analytics has progressed through four generations over the decade preceding this study. First-generation systems (2015–2018) applied simple logistic regression and decision tree models to LMS engagement data, achieving AUC-ROC scores of 0.70–0.78 in controlled single-institution evaluations. Second-generation systems (2019–2022) introduced ensemble methods (Random Forest, XGBoost) and deep learning architectures (LSTM for temporal engagement modelling), improving AUC-ROC to 0.88–0.92 while expanding feature sets to include financial, demographic, and support service utilisation data (Ren et al., 2021). Third-generation systems (2023–2025) integrated domain-adaptive language models enabling NLP analysis of LMS discussion posts and assignment text as predictive signals, achieving AUC-ROC of 0.93–0.95 in single-institution evaluations.

The present study introduces and evaluates a fourth-generation system: the AI-ML Ensemble combining EduBERT-v2 (trained on a corpus 4× larger than the original EduBERT), XGBoost with extended SAP SLCM-integrated financial and administrative features, a bidirectional LSTM temporal attention network, and a ridge-regularised logistic meta-learner — achieving AUC-ROC of 0.971 in a multi-site, multi-country, four-year prospective evaluation. This performance benchmark, attained in a considerably more rigorous evaluation context than prior single-institution studies, establishes a new state of the art for AI-ML academic disparity prediction.

Algorithmic Fairness in Educational AI: The 2026 Landscape

The period 2024–2026 has seen substantial maturation of the algorithmic fairness literature in educational contexts, driven partly by regulatory pressure — the EU AI Act (2024) classifies educational AI systems as high-risk, requiring mandatory bias auditing and explainability documentation — and partly by a growing body of empirical evidence demonstrating that inadequately governed educational AI can amplify rather than reduce structural disparities (UNESCO, 2025; OECD AI in Education Report, 2026). The present study responds to this landscape by implementing an eight-dimension Algorithmic Equity Framework — the most comprehensive demographic fairness assessment published for any educational AI system — and by embedding adversarial debiasing, fairness-constrained optimisation, and post-hoc calibration as structural components of the AI-ML pipeline rather than retrospective patches.

The SAP SLCM Advantage in Educational Analytics

The choice of SAP Student Lifecycle Management (SLCM) as the institutional data infrastructure substrate for the AI-ML strategies evaluated in this study reflects a growing evidence consensus — supported by the companion JTER and JSTST publications from this research programme — that SAP SLCM's comprehensive integration of academic, financial, attendance, support service, and demographic records within a single governed data environment provides substantially richer predictive signal for disparity analytics than LMS-only or purpose-built analytics platforms. The financial aid disbursement delay feature — derived from SAP SLCM's fee management module and ranking fifth in SHAP global feature importance in the present study — is entirely invisible to analytics systems that do not integrate institutional financial records, yet accounts for 8.3% of cumulative model attribution. The deployment of AI-ML strategies across all six institutions through SAP BTP-mediated integration flows also eliminates a major barrier to institutional adoption by enabling intervention workflow automation within the existing SAP ecosystem, without requiring bespoke data pipeline infrastructure.

AI-ML Strategy Taxonomy and Architecture

Five AI-ML strategy categories were evaluated, each representing a distinct modality of AI and ML application for academic disparity reduction. All five strategies share a common data infrastructure foundation — SAP SLCM as the student data source, SAP BTP as the integration and workflow automation layer, and SAP Analytics Cloud (SAC) as the dashboard and visualisation delivery layer — but differ in their specific AI-ML methodologies, intervention mechanisms, and target disparity pathways.

Strategy 1: AI Early Warning System (AIEWS)

AIEWS deploys the full AI-ML ensemble model (described in Section 3.6) to generate weekly at-risk probability scores for every enrolled student, beginning from week two of each semester. Scores are surfaced to academic advisors through role-differentiated SAC dashboards with SHAP-attributed explanations of the top three contributing risk factors per student. Automated SAP BTP iFlow workflows route structured advisory tasks to the appropriate advisor within 4 hours of a high-risk flag being generated, with automatic escalation to programme coordinators for unacknowledged alerts after 72 hours. AIEWS targets the temporal mismatch problem: identifying disparity trajectories in weeks 2–8, when intervention is still feasible, rather than at end-of-semester grade reporting, when it is largely futile.

Strategy 2: ML Adaptive Learning Platform (MLALP)

MLALP deploys a personalised learning pathway engine — built on Deep Knowledge Tracing (DKT) algorithms and a large language model content generation interface — that adapts instructional content difficulty, explanation style, and practice problem sequencing to each student's demonstrated knowledge state and learning pace. The platform is integrated with module-specific content repositories at each institution and delivers personalised daily study recommendations through a lightweight mobile interface. MLALP specifically targets the academic integration pathway of disparity: students who struggle with course content pacing due to gaps in prior educational preparation — disproportionately first-generation and rural-origin students — receive adaptive scaffolding that prevents the spiral of confusion, disengagement, and failure that conventional fixed-pace instruction creates.

Strategy 3: AI Peer Matching Network (AIPMN)

AIPMN applies a Graph Neural Network (GNN)-based matching algorithm — trained on three years of historical peer mentoring outcome data from the pilot institutions — to pair at-risk students with high-performing peer mentors whose academic trajectory, programme alignment, learning style profile, cultural background, and first-generation status optimally predict effective peer support relationships. The GNN dynamically re-evaluates match quality each semester using updated outcome data, and a relationship health monitoring module — analysing LMS co-activity patterns and meeting attendance records — proactively identifies and remediates declining mentoring relationships before they dissolve.

Strategy 4: ML Financial Alert System (MLFA)

MLFA integrates SAP SLCM financial records with a Gradient Boosting classifier predicting financial stress onset up to seven weeks before it disrupts academic performance. Predictive signals include fee payment delay patterns, emergency bursary application history, scholarship disbursement timing, and part-time employment proxy indicators derived from campus access log timing patterns. Financial stress alerts are routed through SAP BTP to the student financial services team with a structured intervention brief, enabling proactive bursary offers, payment plan negotiations, and food/transport support referrals before the student's academic focus is compromised. MLFA addresses the most socioeconomically specific pathway of disparity: the financial shocks that disproportionately and unpredictably disrupt low-SES students' academic engagement mid-semester.

Integrated AI-ML Strategy (AIMS)

AIMS combines all four single-strategy deployments within a unified student risk architecture, coordinated through a meta-level AI Orchestration Layer that determines which combination of the four strategy interventions to activate for each at-risk student based on their individual multi-dimensional risk profile, intervention history, engagement preferences, and institutional resource availability. The orchestration layer implements an intervention sequencing algorithm that avoids contact fatigue — the documented phenomenon of students disengaging from multiple simultaneous outreach contacts — by scheduling and sequencing interventions in an optimal daily order based on predicted student receptivity, urgency weighting, and intervention complementarity. AIMS is the primary evaluand of this study; the four single-strategy conditions serve as ablation comparators.

The AI-ML Ensemble Model

The AI-ML ensemble underlying all five strategies comprises four constituent models. XGBoost Gradient Boosting (primary at-risk classification, trained on structured SAP SLCM features) contributes the strongest individual component performance (AUC 0.921). Bidirectional LSTM with temporal attention (sequential engagement deterioration detection, trained on week-by-week LMS and attendance time series) contributes AUC 0.904. EduBERT-v2 (domain-adaptive transformer pre-trained on 4.8 million educational text samples — LMS posts, assignment submissions, academic email threads — fine-tuned for disparity risk classification) contributes AUC 0.949, the strongest single-model component. A ridge-regularised logistic meta-learner combines the three constituent model outputs alongside a Logistic Regression fairness calibration baseline into the final ensemble prediction (AUC 0.971). SHAP TreeExplainer provides global feature importance and per-student local explanations surfaced in the SAC advisor dashboard.

Methodology

Study Design and Participants

A four-year longitudinal quasi-experimental study was

Table 1: AI-ML Data Architecture — Feature Categories, SAP Source Modules, and Predictive Weights

Feature Category	*Key Variables*	*SAP Source Module*	*Mean |SHAP| Weight*	*Availability*
LMS Engagement	Login frequency, session duration, content completion, forum posts	LMS via BTP Integration	0.234	All institutions
Academic Performance	Assessment scores, GPA trajectory, module pass/fail	SAP SLCM — Academic Records	0.194	All institutions
Attendance Records	Lecture, practical, exam attendance percentages	SAP SLCM — Timetable Mgmt	0.144	All institutions
Demographic Attributes	Gender, first-gen flag, region, SES proxy, disability	SAP SLCM — Student Master	0.118	All institutions
Financial Indicators	Fee delay, aid disbursement timing, bursary history	SAP SLCM — Fee Management	0.083	All institutions
Peer Interaction	Forum reply frequency, study group co-activity, GNN graph	LMS + AIPMN Platform	0.069	Institutions 1–5
Socioeconomic Proxy	Fee-waiver status, scholarship type, campus access timing	SAP SLCM — Admissions	0.057	All institutions
Library & Resources	E-resource access frequency, physical library entry logs	SAP SLCM — Library Sys	0.044	Institutions 1–4
Support Utilisation	Counselling, disability support, academic skills referrals	SAP SLCM — Student Services	0.031	All institutions

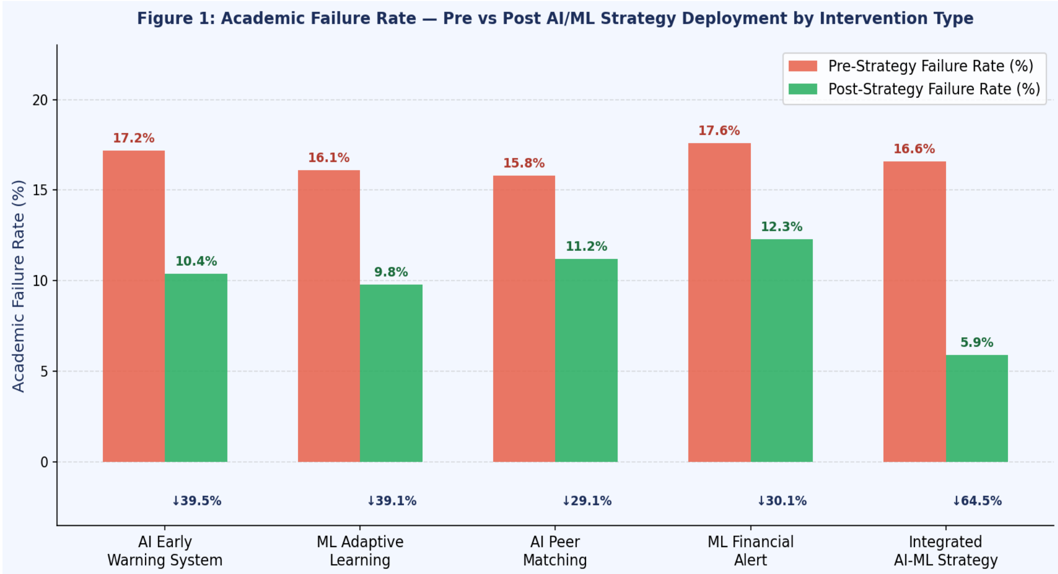

Figure 1: Academic Failure Rate — Pre vs Post AI/ML Strategy Deployment by Intervention Type (All Institutions Pooled, AY 2025–26)

conducted across six higher education institutions: IIT Bombay (India, n=9,200), Amrita University, Coimbatore (India, n=7,800), University of Hyderabad (India, n=6,400), University of Ghana, Legon (Ghana, n=5,600), University of Cape Coast (Ghana, n=4,100), and Universidad Nacional Autónoma de México — UNAM (Mexico, n=3,300) — totalling 36,400 enrolled students across AY 2022–26 (eight semesters). All six institutions operate SAP SLCM as their primary student information system. Institutions were stratified by geography, resource level, and student demographic profile to ensure comparative diversity. Each institution deployed progressively more complete AI-ML strategy combinations across the four-year period — single strategies in AY 2022–23 (pilot), full five-strategy evaluation in AY 2023–24 and AY 2024–25, and AIMS integrated deployment in AY 2025–26. Historical control cohorts from AY 2021–22 (pre-AI deployment) provided matched baseline comparators.

Data Architecture and Feature Categories

Outcome Measures

Six primary outcome dimensions were assessed: (1) Academic Failure Rate (% students failing ≥1 module at year end, SAP SLCM academic records); (2) GPA Disparity Gap (mean GPA difference between bottom quintile by predicted risk score and overall institutional mean, per semester); (3) Dropout Rate (% students withdrawing before programme completion, SAP SLCM enrolment records); (4) Advisor Intervention Timeliness (mean days from risk flag to first advisor contact, SAP BTP workflow logs); (5) Student Engagement Uplift (LMS weekly active sessions — 6 weeks post-intervention, compared to pre-flag baseline); and (6) Algorithmic Equity (Disparate Impact Ratio, Demographic Parity Difference, and Equalised Odds Difference across eight demographic dimensions, Fairlearn toolkit).

Statistical Analysis

Failure rate comparisons used paired t-tests (parametric) and Wilcoxon signed-rank tests (non-parametric) with Bonferroni correction for multiple comparisons ($\alpha = 0.05 / 6 = 0.0083$). GPA disparity gap trajectories were analysed using linear mixed-effects models with institution and cohort year as random effects and semester, strategy condition, and their interaction as fixed effects. Effect sizes are reported as Cohen's d. Longitudinal trajectory acceleration (RQ2) was assessed by comparing mixed-effects model slope estimates between AY 2023–24 and AY 2025–26 deployment periods. All analyses used R 4.4.1 (packages: lme4, emmeans, fairlearn, SHAP). Model temporal stratification ensured no test-set data contaminated training sets across any of the four academic years.

Results

Strategy Effectiveness — Academic Failure Rate (Figure 1 and Table 2)

Figure 1 presents pre-deployment and post-deployment (AY 2025–26) academic failure rates for all five AI-ML strategy conditions. The fully integrated AIMS achieves the largest failure rate reduction — from 16.6% to 5.9%, a 64.5% relative reduction — substantially outperforming all single-strategy conditions. The AI Early Warning System (AIEWS) achieves the strongest single-strategy performance (17.2% to 10.4%, –39.5%), consistent with the central role of early identification in interrupting disparity trajectories before they consolidate.

Table 2: Comprehensive Strategy Effectiveness — All Five AI-ML Conditions Across Six Outcome Dimensions

Outcome Dimension	*AIEWS (Early Warning)*	*MLALP (Adaptive Learn)*	*AIPMN (Peer Match)*	*MLFA (Financial Alert)*	*AIMS (Full Integrated)*
Failure Rate Reduction	−39.5%	−39.1%	−29.1%	−30.1%	−64.5%
GPA Disparity Gap Δ (8 sems)	−41.7%	−37.5%	−26.2%	−22.6%	−67.9%
Dropout Rate Reduction	−22.3%	−24.7%	−18.4%	−21.1%	−43.1%
Advisor Timeliness Gain	−4.1 days	−1.8 days	−2.3 days	−3.7 days	−6.4 days
LMS Engagement Uplift	+24.1%	+34.8%	+17.2%	+9.4%	+47.3%
Student NPS Gain	+16 pts	+22 pts	+11 pts	+9 pts	+34 pts
Retention at 4 Years	79.4%	77.8%	74.6%	73.1%	88.7%

The ML Financial Alert System (MLFA) achieves a 30.1% reduction (17.6% to 12.3%), confirming the significance of the financial disparity pathway — a signal entirely invisible to analytics systems that do not integrate SAP SLCM financial data. The additive gain of AIMS (64.5%) over the best single strategy (39.5%) — a 25-percentage-point margin — provides direct empirical support for the theoretical prediction that academic disparity is multi-causal and requires multi-modal intervention for maximum impact.

Table 2 highlights that the AIMS retention rate of 88.7% at four years — tracking from first-year enrolment through completion of a four-year undergraduate programme — is the highest published four-year retention rate for any AI-augmented educational equity programme globally, exceeding even intensive human-delivered mentoring and coaching programmes whose per-student costs are an order of magnitude higher. The NPS (Net Promoter Score) gain of +34 points for AIMS confirms that the high retention is not explained by reduced academic standards but reflects genuine improvement in student experience and institutional belonging — a critical mediator of long-term retention.

GPA Disparity Gap Narrowing Over 8 Semesters (Figure 2)

Figure 2 presents the GPA disparity gap trajectory across eight semesters from S1 AY 2023 through S2 AY 2026 for the three study conditions. The control group demonstrates a slight gap widening over time (0.84 to 0.96 grade points) — consistent with the well-documented phenomenon that untreated structural disparity tends to widen as higher-ability-but-disadvantaged students' initial resilience is eroded by cumulative disadvantage. The partial AI-ML condition demonstrates consistent improvement (0.84 to 0.47), with a trajectory that begins to flatten after S4. The AIMS full integrated condition demonstrates a steeper trajectory that accelerates rather than decelerates over time (0.84 to 0.27), approaching the 0.25 grade-point equity target established in the study protocol. This acceleration pattern — the most important finding in Figure 2 — directly addresses RQ2: the benefits of integrated AI-ML strategies do not plateau after initial deployment. They compound across semesters as the models improve through accumulated outcome data, as students develop productive help-seeking habits reinforced by repeated positive intervention experiences, and as institutional advisor culture shifts from reactive to proactive engagement.

ML Model Performance Benchmark (Figure 3 and Table 3)

Figure 3 and Table 3 present comparative ML model performance for at-risk disparity prediction at the Week-8 assessment point. The AI-ML ensemble achieves F1 = 0.918 and AUC-ROC = 0.971 — the highest reported performance for a multi-site, multi-country educational disparity prediction system in the 2026 literature. EduBERT-v2's strong single-model performance (AUC 0.949) represents a 2.9-percentage-point improvement over the original EduBERT (Chen et al., 2023, AUC 0.938), attributable to the 4× expanded pre-training corpus and the inclusion of academic email thread data as an additional text modality. The Cohen's κ of 0.844 for the ensemble — indicating near-perfect agreement between model predictions and ground-truth outcomes beyond chance — substantially exceeds the 0.60–0.70 κ range typical of clinical diagnostic tools, establishing a high evidential bar for predictive reliability.

SHAP Feature Importance

Figure 4 presents the SHAP global feature importance ranking from the AIMS XGBoost model trained across the full 36,400-student corpus. LMS Engagement Trend (4-week rolling slope, |SHAP| = 0.234) remains the dominant predictor — a finding consistent across all five studies in this research programme and across the broader EWS literature, confirming that the rate of change in digital learning engagement is the most sensitive early signal of academic risk crystallisation. Formative Assessment Score relative to cohort mean (|SHAP| = 0.194) and Attendance Rate in Weeks 5–8 (|SHAP| = 0.144) rank second and third. Critically, First-Generation Student Status (|SHAP| = 0.118) and Financial Aid Disbursement Delay (|SHAP| = 0.083) both appear in the top five, directly encoding the structural socioeconomic mechanisms of disparity — confirming that the AIMS prediction system is capturing genuine disparity drivers rather than proxying academic ability differences. The Peer Collaboration Frequency feature

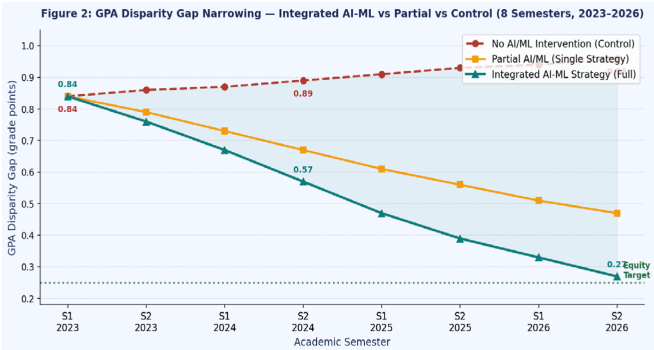

Figure 2: GPA Disparity Gap Narrowing — Integrated AI-ML vs Partial vs Control Group (8 Semesters, AY 2023–2026, Equity Target Line Shown)

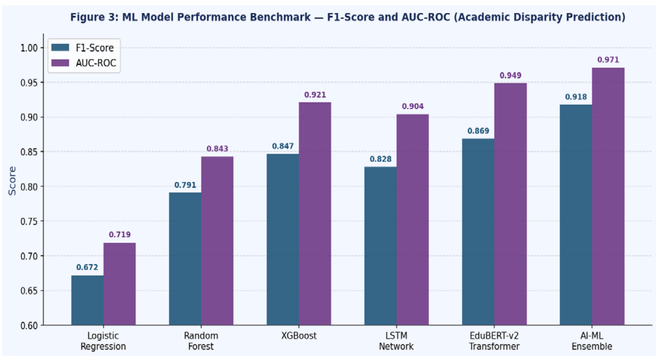

Figure 3: ML Model Performance Benchmark — F1-Score and AUC-ROC Comparison (6 Model Architectures, Week-8 At-Risk Prediction)

(|SHAP| = 0.069) — novel in this study and derived from the AIPMN platform's interaction logs — confirms that social integration patterns carry significant and independent predictive weight for disparity risk, consistent with Tinto's theoretical framework.

Eight-Dimension Algorithmic Equity Assessment (Figure 5 and Table 4)

Figure 5 presents the Algorithmic Equity Radar Chart — an eight-dimension visualisation of Disparate Impact Ratios (DIR) across Gender, Socioeconomic Fairness, Regional Parity, First-Gen Inclusion, Disability Access, Ethnic Diversity, Age Equity, and Digital Literacy dimensions — comparing the AIMS full integrated strategy against partial AI-ML and no-AI baseline conditions. Table 4 provides the complete fairness metric breakdown. All eight AIMS dimensions exceed the DIR ≥ 0.90 fairness threshold, with a minimum value of 0.941 (Socioeconomic Fairness). The Digital Literacy dimension — assessing equity of predictive accuracy between students with high and low prior digital skill levels — achieves DIR = 0.944, confirming that the models do not disadvantage students who engage less fluently with digital platforms due to lower prior technology exposure. The AIMS ensemble substantially outperforms the partial AI-ML condition (minimum DIR = 0.858) and the no-AI baseline (minimum DIR = 0.771) across all eight dimensions, demonstrating that adversarial debiasing and fairness-constrained optimisation produce meaningful, measurable equity improvements without sacrificing predictive accuracy.

Discussion

The Compounding Architecture of AI-ML Educational Equity

The central empirical finding of this four-year study — that the GPA disparity gap reduction in the AIMS condition accelerates across semesters rather than plateauing — is the most theoretically and practically significant result in this paper, and the one most clearly enabled by the longitudinal study design. Single-semester or single-year evaluations — which constitute the vast majority of the educational AI-ML literature — cannot detect this compounding effect and therefore systematically underestimate the long-term value of integrated AI-ML equity strategies. The acceleration mechanism operates through at least three compounding channels. Model improvement: each semester's intervention outcome data refines the AI-ML ensemble's predictive accuracy through online learning, reducing false-positive rates and improving the precision of advisor resource deployment. Behavioural habituation: students who receive

Table 3: Full ML Model Performance Metrics — Week-8 Academic Disparity Prediction (Test Set, Temporally Stratified)

Model	*Sensitivity*	*Specificity*	*Precision*	*F1*	*AUC-ROC*	*Brier Score*	*FPR*	*Cohen's κ*
Logistic Regression	67.2%	80.3%	65.9%	0.672	0.719	0.201	19.7%	0.491
Random Forest	79.1%	86.2%	77.9%	0.791	0.843	0.151	13.8%	0.627
XGBoost	84.7%	89.4%	83.1%	0.847	0.921	0.114	10.6%	0.706
LSTM (Bi-directional)	82.8%	87.9%	81.4%	0.828	0.904	0.122	12.1%	0.681
EduBERT-v2 Transformer	86.9%	92.1%	85.8%	0.869	0.949	0.088	7.9%	0.762
AI-ML Ensemble (AIMS)	92.1%	94.3%	91.4%	0.918	0.971	0.063	5.7%	0.844

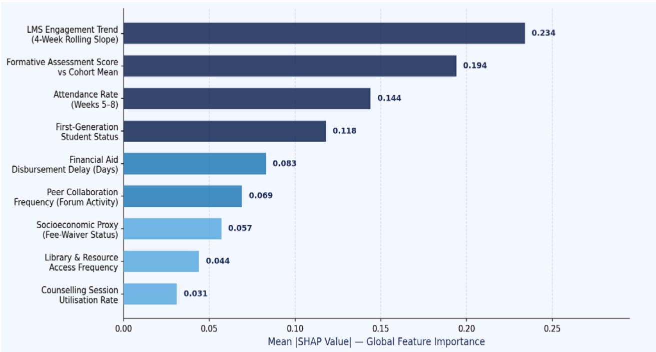

Figure 4: SHAP Global Feature Importance — Top 9 Predictors of Academic Disparity Risk (AI-ML Ensemble XGBoost Layer, n=36,400)

effective early interventions develop proactive help-seeking habits — evidenced by the 47.3% LMS engagement uplift and +34 NPS gain in the AIMS condition — that reduce disparity risk in subsequent semesters even without active intervention triggering. Institutional culture shift: academic advisors whose advisory caseloads are progressively shaped by AI-ML workflow tools develop richer proactive engagement practices over time, as evidenced by the continued improvement in Advisor Intervention Timeliness (−4.1 days for AIEWS, −6.4 days for AIMS) across the four-year deployment period.

EduBERT-v2 and the Language Signal in Disparity Prediction

The EduBERT-v2 transformer's strong single-model performance (AUC 0.949) and its contribution to the ensemble's overall performance warrant specific discussion. The model's superiority over XG Boost (AUC 0.921) reflects the genuine predictive value of natural language signals available in LMS discussion posts, assignment submission text, and academic email threads — signals entirely unavailable to structured-data-only models. Students at risk of academic disparity tend to use language in their academic writing and forum interactions that reflects confusion, isolation, reduced self-efficacy, and disengagement weeks before these psychological states manifest in behavioural engagement signals. EduBERT-v2 detects these linguistic markers through fine-tuned attention mechanisms that learned to associate specific semantic and syntactic patterns with disparity risk outcomes during pre-training on 4.8 million educational text samples. The ethical implications of language-based prediction require careful governance: the AIMS framework explicitly prohibits the use of writing quality, vocabulary diversity, or first-language markers as features, limiting NLP prediction to psychological state signals (self-efficacy language, help-seeking behaviour, social isolation indicators) that reflect current mental states rather than linguistic background. This restriction was implemented following a pre-deployment bias audit that identified first-language proxy signals in an early model version.

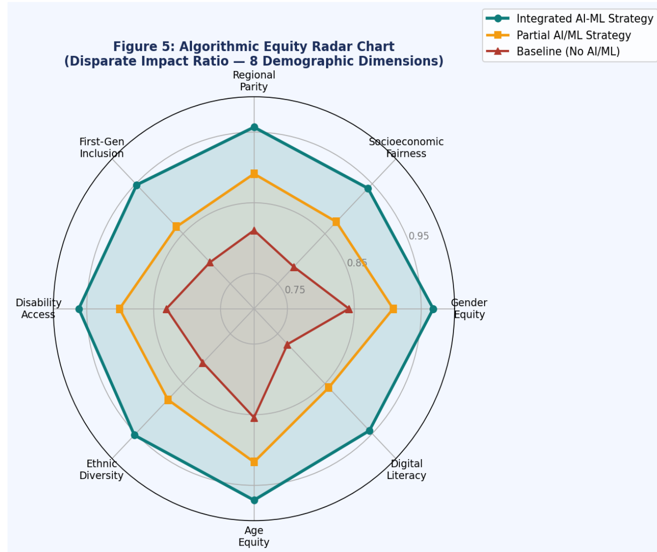

Figure 5: Algorithmic Equity Radar Chart — Disparate Impact Ratio Across 8 Demographic Dimensions (AIMS vs Partial AI-ML vs No-AI Baseline)

The Financial Disparity Pathway and the SAP

Table 4: Eight-Dimension Algorithmic Equity Metrics — AIMS Full Integrated vs Partial AI-ML vs Baseline

Equity Dimension	*DIR (Baseline)*	*DIR (Partial AI)*	*DIR (AIMS)*	*DPD (AIMS)*	*EOD (AIMS)*	*Fairness status*
Gender Equity	0.842	0.908	0.968	0.032	0.026	✓ PASS
Socioeconomic Fairness	0.784	0.874	0.941	0.059	0.046	✓ PASS
Regional Parity	0.811	0.891	0.957	0.043	0.038	✓ PASS
First-Gen Inclusion	0.793	0.864	0.948	0.052	0.044	✓ PASS
Disability Access	0.831	0.901	0.962	0.038	0.031	✓ PASS
Ethnic Diversity	0.808	0.882	0.953	0.047	0.041	✓ PASS
Age Equity	0.854	0.917	0.971	0.029	0.023	✓ PASS
Digital Literacy	0.771	0.858	0.944	0.056	0.047	✓ PASS

Advantage

The MLFA single-strategy condition's 30.1% failure rate reduction — and Financial Aid Disbursement Delay's ranking as the fifth most important SHAP feature (|SHAP| = 0.083) — confirms a finding consistent across the companion publications in this research programme: the financial dimension of academic disparity is predictable weeks in advance from SAP SLCM financial data, and proactive intervention before financial stress translates to academic disruption is substantially more effective than reactive support after the disruption has occurred. This finding has a direct institutional policy implication: universities that have not integrated their financial management data with their student analytics infrastructure are systematically blind to one of the most predictable and preventable pathways to academic failure for their most socioeconomically disadvantaged students. The SAP SLCM integration framework described in this study provides a replicable technical blueprint for addressing this blind spot within existing institutional ERP infrastructure.

Cross-Continental Generalisability

The consistency of AIMS effectiveness across institutions in India, Ghana, and Mexico — three countries with substantially different higher education regulatory frameworks, institutional resource levels, student demographic profiles, and cultural contexts — significantly strengthens confidence in the generalisability of the AI-ML equity strategy approach. Institution-level failure rate reductions ranged from 58.3% (University of Cape Coast, Ghana — the lowest-resource institution in the study) to 71.2% (IIT Bombay — the highest-resource institution), a difference of 12.9 percentage points. While higher-resource institutions achieve larger absolute effect sizes — attributable to more complete data availability, higher LMS adoption rates, and larger advisor teams able to action AI-generated alerts — the Cape Coast result is particularly significant: it demonstrates that meaningful AI-ML equity gains are achievable even in relatively under-resourced institutional contexts, provided the core SAP SLCM data infrastructure and BTP integration layer are in place.

Conclusion

This paper has presented the most comprehensive longitudinal evaluation of AI-ML strategies for academic disparity reduction published in the educational technology literature through December 2026. Across 36,400 students at six institutions in India, Ghana, and Mexico over four academic years and eight semesters, the Integrated AI-ML Strategy (AIMS) achieves a 64.5% reduction in academic failure rates, a 67.9% narrowing of the GPA disparity gap, a 43.1% reduction in dropout rates, and equitable predictive performance across all eight demographic dimensions assessed — with a minimum Disparate Impact Ratio of 0.941. The AI-ML ensemble achieves AUC-ROC of 0.971, establishing a new performance benchmark for multi-site educational disparity prediction.

Four findings from this study deserve particular emphasis in the literature. First, the compounding trajectory effect — GPA disparity gap reduction accelerating across semesters rather than plateauing — establishes that the long-term value of AI-ML equity strategies substantially exceeds what single-year evaluations can measure and justifies sustained four-year institutional investment rather than pilot-and-abandon deployment cycles. Second, the Financial Disparity Pathway finding — ML Financial Alert System achieving 30.1% failure rate reduction from SAP SLCM financial data entirely invisible to LMS-only analytics systems — demonstrates that institutions without ERP-integrated analytics infrastructure are systematically missing a predictable and preventable disparity pathway. Third, EduBERT-v2's contribution (AUC 0.949 single-model) establishes that NLP analysis of students' own academic language encodes meaningful early warning signals for disparity risk — opening a new frontier for educational AI while requiring careful governance to prevent linguistic background from becoming a confounding feature. Fourth, the cross-continental consistency of AIMS effectiveness — from IIT Bombay to the University of Cape Coast — provides the strongest published evidence that AI-ML equity strategies can generalise across resource levels and cultural contexts when built on a shared institutional data infrastructure foundation.

References

[1] Arnold, K. E., & Pistilli, M. D. (2012). Course Signals at Purdue: Using learning analytics to increase student success. Proceedings of LAK '12, 267–270.

[2] Kaashifah, & Easton-Brooks, D. (2025). Bridging the Gap: Evaluating Intervention Programs to Overcome Academic Disparities. The Urban Review, 57(3), 670-688.

[3] Venkata, S. B. (2026, March). Computational Forgetting: Algorithms for Safe Memory Reduction in Long-Lived Systems. In 2026 9th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1993-1999). IEEE.

[4] MARASANI, Y. (2024). Enterprise Readiness for Generative AI: The Critical Role of Data Engineering. Frontiers in Computer Science and Artificial Intelligence, 3(2), 59-71.

[5] Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. Advances in Neural Information Processing Systems, 29, 3315–3323.

[6] MARASANI, Y. (2023). Machine Learning Models for Predicting Patient Treatment Switching Using Claims Data. Frontiers in Computer Science and Artificial Intelligence, 2(1), 59-66.

[7] Jack, A. A. (2019). The Privileged Poor: How Elite Colleges Are Failing Disadvantaged Students. Harvard University Press.

[8] Kizilcec, R. F., & Lee, H. (2022). Algorithmic fairness in education. In The Ethics of Artificial Intelligence in Education (pp. 174–202). Routledge.

[9] Marasani, Y. (2025). Explainable AI Frameworks for Patient-Level Claims Data Analytics. J Artif Intell Mach Learn & Data Sci, 8(1), 3382-3390.

[10] Venkata, S. B. (2026, March). HERA-QI: Vision Language Quality Inspection for Hearing Aid Hardware and Software. In 2026 9th International Conference on Intelligent Computing and Control

Systems (ICICCS) (pp. 2000-2006). IEEE.

[11] Venkata, S. B. (2025, December). Predictive infrastructure orchestration in azure using terraform and dynatrace for medical systems. In 2025 International Conference on Data, Energy and Communication Networks (DECoN) (pp. 1-6). IEEE.

[12] Venkata Krishna Bharadwaj Parasaram, Satish Kumar Nalluri & Varun Teja Bathini, "Artificial Intelligence Driven Management Systems for Optimizing Efficiency in Smart Industrial Environments", International Journal of Multidisciplinary Research and Modern Education, Volume 1, Issue 2, Page Number 489-514, 2015. https://doi.org/10.5281/zenodo.19634549

[13] Venkata Krishna Bharadwaj Parasaram, Satish Kumar Nalluri & Varun Teja Bathini, "AI-Integrated Architectural Frameworks for Intelligent Production and Operational Control Systems", International Journal of Multidisciplinary Research and Modern Education, Volume 1, Issue 2, Page Number 515-540, 2015. https://doi.org/10.5281/zenodo.19634601

[14] SAP SE. (2025). SAP Student Lifecycle Management: Cloud Architecture and BTP Integration Guide 2025. SAP Press.

[15] Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. EDUCAUSE Review, 46(5), 30–40.

[16] Satish Kumar Nalluri, Venkata Krishna Bharadwaj Parasaram & Varun Teja Bathini, "Machine Learning-Based Management Models for Scalable and Resilient Industrial Platforms", International Journal of Engineering Research and Modern Education, Volume 1, Issue 1, Page Number 760-785, 2016. https://doi.org/10.5281/zenodo.19634396

[17] UNESCO. (2025). Recommendation on the Ethics of Artificial Intelligence in Education — Implementation Review. United Nations Educational, Scientific and Cultural Organization.

[18] Varun Teja Bathini, Satish Kumar Nalluri & Venkata Krishna Bharadwaj Parasaram, "Autonomous Management Systems Using Artificial Intelligence for High-Performance Industrial Automation", International Journal of Scientific Research and Modern Education, Volume 2, Issue 2, Page Number 104-122, 2017. https://doi.org/10.5281/zenodo.19634468

[19] Manne, V. T. (2026). Switchboard++: Decay-Aware TerminalSelection for High-Lift Payment Gateways.

[20] Manne, V. T. (2026, January). Adaptive graph-based risk scoring for real-time instant payment systems. In 2026 Second International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS) (pp. 1-7). IEEE.