International Journal of Innovative Technology and Interdisciplinary Sciences
https://journals.tultech.eu/index.php/ijitis
ISSN: 2613-7305
Volume 9, Issue 4
DOI: https://doi.org/10.15157/IJITIS.2026.9.4.2352-2378
Received: 07.08.2026; Revised: 30.09.2026; Accepted: 30.10.2026

Research Article

Artificial Intelligence Integration in STEM Education: An Evidence-Constrained Systems Readiness Assessment

Elda Spahiu[1], Suela Hoxhaj[2], Dafina Xhako[2], Niko Hyka[3], Altin Gjevori[2], Valbona Muda[2], Karrar Qasim Mohammed Ali Alfasili[4]

[1]Institute of Applied Nuclear Physics, Tirana, Albania
[2]Department of Physics Engineering, Polytechnic University of Tirana, Tirana, Albania
[3]Faculty of Technical Medical Sciences, University of Medicine, Tirana, Albania
[4]Department of Electrical and Computer Engineering, Babol Noshirvani University of Technology, Mazandaran, Iran
***d.xhako@fimif.edu.al**

Abstract

Artificial intelligence (AI) is entering university STEM education at a pace that may exceed the capacity of curricula, professional development, infrastructure, and governance to adapt. This study develops an evidence-constrained systems-readiness framework based on two parallel surveys conducted in Albania in January 2026, comprising 664 student responses (approximately 282 substantially complete entries) and 92 academic staff responses from public and private higher education institutions. As respondent-level data were unavailable for this secondary analysis, the investigation relied on statistics that could be defensibly reconstructed from aggregate reports, including chi-square tests, Holm-adjusted one-sample tests, Welch comparisons, effect sizes, Wilson confidence intervals, denominator and covariance sensitivity analyses, and report-level qualitative triangulation. The framework integrates reported curriculum maturity, responsible AI practices, academic capacity, infrastructure, governance, student preparedness, and perceived employability alignment, while distinguishing perceptions from demonstrated competence and independently verified institutional performance. Student and staff distributions across the four levels of AI integration showed no statistically detectable difference ($\chi^2(3) = 1.28$, $p = 0.734$, Cramér's $V = 0.069$), with reported curriculum maturity scores of 34.1/100 and 30.0/100, respectively. Students reported a substantial gap between the availability of project-based learning opportunities and the perceived effectiveness of practical learning (Hedges' $g \approx 0.80$). Academic staff rated existing infrastructure below the scale midpoint but projected institutional readiness over the following three years above the midpoint. The perceived demand–preparedness gap remained positive under assumed within-respondent correlations ranging from 0 to 0.6. The study contributes an implementation-agnostic and reproducible systems-readiness framework that identifies asynchronous institutional constraints, preserves uncertainty arising from aggregate data, and translates perception-based findings into testable pathways for AI integration in university STEM education.

Keywords: Artificial Intelligence; STEM Education; Systems Readiness; Curriculum Integration; Academic Staff; Student Perceptions.

©2026 Authors. This is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International License CC BY 4.0 (http://creativecommons.org/licenses/by/4.0).

INTRODUCTION

Artificial Intelligence (AI) has shifted from a specialized computing field into a general-purpose technology that influences research, engineering design and the analysis of scientific data, health technologies, and professional decision-making. Higher education now has a double challenge, learning to use AI wisely within the curriculum, while maintaining the human judgment, disciplinary logic, ethical considerations, and ability to interpret outputs that distinguish model from human generated reasoning. Increasingly global competencies for using digital and AI tools are defined as combinations of knowledge, application, critique and values, ethics, and human guidance but higher education evidence remains scattered [1-4]. Early reviews found that AI research focused most often on technical infrastructures over teachers, pedagogy, and institutional contexts [5]. Later syntheses described wider practices, although continued to report limitations in research methodology and ethics, comparable findings across disciplines, and institutional embedding [6, 7]. Reviews of generative AI similarly find that access and trials are proceeding faster than paradigms of learning, assessment, governance and staff awareness [8-10]. The problem is systemic, curriculum, professional knowledge, infrastructure, authentic practice, governance, and labour market assumptions do not necessarily keep in step.

AI literacy and competency frameworks detail the content knowledge learners and educators should be able to acquire, use, critically review and produce. Institutional AI literacy and competency frameworks have evolved across several complementary perspectives. Early definitions conceptualised AI literacy as the ability to understand, use, critically evaluate, and communicate about AI [11]. Authors in [12] extended this perspective by identifying technical, cognitive, behavioural, and ethical dimensions, while in [13] advocated integrating AI across the curriculum rather than treating it as a standalone subject. The study in [14] examined the pedagogical, governance, and operational dimensions of university AI policies. Subsequently, UNESCO established a student AI competency framework encompassing a human-centred mindset, AI ethics, AI techniques and applications, and AI system design, with progression from understanding to application and creation [15]. Its complementary teacher competency framework further emphasises AI-related pedagogical reasoning and professional learning [16]. These frameworks specify content of competence and progression, but do not alone reveal whether an institution has the pedagogic and professional development resources, curriculum infrastructure, community, based learning opportunities, and governance to put them into effect. Second set of boundaries arise from measurement research which shows that AI literacy instruments vary widely in structure and conduct validation across population samples [17, 18], and educator, competence frameworks still need empirical validation to determine usefulness as indicators of competence [19]. Accordingly, self-reports are treated solely as diagnostic indicators, rather than as objective measures of competence.

Hence, although international empirical comparisons provide valuable insights, they offer only a partial basis for assessment. Authors in [20] integrated a literature review, a survey of 29 UK universities and a pilot engineering AI module, and highlighted discipline, specific design and educational evaluation. The study in [21] comprised the attitudes toward intellectual growth, anxiety and self-efficacy. Authors in [22] explored the acceptance, perceived effects and educational use of large language models among engineering students. Institutional exemplars provide some illustration of embedded practice, MIT's Common Ground is blending computing and AI with disciplinary learning in all departments [23], and ETH Zurich is combining specialized AI instruction with team-based, real-world data science and research projects [24]. Although these studies provide insights into adoption dynamics and established design principles, their units of analysis are not directly suited to diagnosing system-level readiness across multiple institutions in resource-constrained contexts with limited AI adoption.

Albania offers a pertinent late, adopter context as the modernization of curricula, digital change, disparity in institutional assets and labour, market requirements are progressing simultaneously. A former objective curriculum audit by [25] over the entire country, entered accredited STEM program landscape provided coding for 35 programs in depth, and demonstrated AI-content to be focused on computer intensive disciplines. This audit created a curriculum level baseline, but did not assess the actual curriculum experienced by learners, staff/teacher endurance, practical experience, infrastructure, or perceived AI-readiness. The current research complements the latter dimensions while keeping the associated program audit data analytically distinct from the perceptions.

Integrated Theoretical Architecture

The revised framework explicitly incorporates theories that apply at different levels rather than treat them as interchangeable. At the individual adoption level, the TAM [26] and UTAUT [27] theories provide accounts of how perceived usefulness/performance expectancy, effort, social influence, and facilitating conditions influence technology adoption. These mechanisms are useful in explaining attitudes toward willingness, but it should be noted that adoption is different to curriculum maturity and capabilities. At the organizational change level, Weiner's readiness for change theory [28] differentiates commitments to change and change commitment from collective implementation capability and clearly links capabilities to task demands, resources, and situational factors. At the pedagogical level, constructive alignment [29] connects the learning outcomes, teaching and learning activities, and assessment, so that they are all mutually aligned. Accordingly, AI literacy and UNESCO competency frameworks [11, 16] specify the nature and progression of these outcomes, and Chan's ecological policy framework [14] supplies the governance and operational environment.

In combination, these dimensions suggest a multilevel model of systems readiness in which willingness to adopt an innovation may precede the development of the institutional conditions required for its effective integration. Consequently, organisational commitment may coexist with underdeveloped infrastructure, curriculum statements may not translate

into authentic learning practices or aligned assessment standards, and perceived readiness may remain distinct from objectively validated competence. The systems-readiness framework developed in this study operationalises this theoretical integration through a set of observable domains: institutional resources, faculty capacity, curriculum maturity, authentic practice, student preparedness, and labour-market needs. Governance, ethics, and quality assurance function as cross-cutting dimensions across all domains. At the current time, this model is intentionally diagnostic rather than causal in the present data set with the causal pathways posited for subsequent longitudinal testing.

Research Gaps and State-of-the-Art Positioning

The four analytically distinct gaps lead the study. Firstly, a research gap exists for multi-institution STEM systems in smaller Southeastern European contexts where resource circumstances and curriculum systems differ from dominant AI in higher education pathways. Secondly, a research gap again exists between adoption and institution wide implementation studies, whose TAM/UTAUT informed willingness and use variable and whose competency in learning frameworks specify the learning targeted, but do not establish the regulatory interaction of resources, staff capacity, curriculum, implementation, governance and readiness as a transition system. Third, a research gap exists between perception, based indicators and validated AI, literacy in demand indicators and objective curriculum audit measures, hence existing studies cannot legitimately infer one from the other. Fourth, an implementation gap is between diagnosis of gaps in provision and a specification of a learnable improvement sequence under constrained information and resources. The framework cannot be delivered otherwise, it preserves the distinctions, but links them in a diagnostic structure.

Table 1 thereby identifies the study as an integrative bridge, rather than as a more refined replacement for existing approaches (adoption, competency, organizational, readiness, curriculum, alignment, and institutional, governance).

Table 1. State-of-the-Art and Theoretical Positioning of the Systems-Readiness Diagnostic Framework

Reference / benchmark	Design	What it establishes	Remaining limitation	Extension in this paper
[5-7]	Systematic/meta reviews of AI in HE	Broad landscape, ethics and rigor gaps	Limited system-level integration model for late-adopter STEM systems	Links curriculum, capacity, infrastructure, practice, governance and readiness as one diagnostic chain
[21]	982 students + 76 faculty, UTAUT-based attitudes	Dual-stakeholder adoption evidence	Single institution, focus on GenAI attitudes rather than curriculum maturity	Adds multi-institution STEM-specific curriculum/infrastructure/readiness diagnosis

[20]	Review + scan of 29 UK universities + pilot module	Discipline-specific module design and evaluation	Different national/institutional capacity, not a parallel student-staff systems diagnostic	Uses engineering curriculum evidence as an external design benchmark, not a direct rate comparator
[15,16]	Normative student and teacher competency frameworks	Progression, ethics, pedagogy, human agency	No direct empirical institutional readiness estimation	Operationalizes progression and staff development in a measurable maturity pathway
[17, 18]	Psychometric review / factor analyses	Measurement validity of AI-literacy constructs	Shows risks of unvalidated self-report inference	Explicitly separates perception indicators from objective competence and proposes future validated scales
[23, 24]	Institutional curricular and project models	Embedded interdisciplinary AI and authentic projects	Case-based, not directly comparable national metrics	Provides design features for the staged roadmap
[25]	National frame + objective coding of 35 programs	Program-level AI presence/depth in Albania	No student/staff experience, practical exposure, or readiness measures	Triangulates curriculum audit evidence with stakeholder experience without treating samples as identical
[26, 27]	TAM / UTAUT technology-adoption theory	Usefulness/performance expectancy, effort, social influence, facilitating conditions, and adoption intention	Explains adoption behaviour but not curriculum progression, competence, or institutional bottlenecks	Uses adoption theory as the individual-use layer while keeping it distinct from capability and readiness
[28]	Organizational readiness-for-change theory	Change commitment and collective change efficacy depend partly on task demands, resources, and situational conditions	Not specific to AI education or curriculum design	Explains why high future-readiness expectations can coexist with weak current implementation capacity
[29]	Constructive alignment	Learning outcomes, learning activities,	Primarily pedagogical/cours	Connects curriculum maturity to authentic

	and assessment should be aligned	e-level, does not model institutional transition	practice and assessment rather than counting AI content alone

Research Questions, Hypotheses and Contributions

The initial project surveys were not preregistered as confirmatory hypothesis, testing research. Thus, H1-H4 are published here in a transparent manner as post hoc, theory, guided analytical expectations informing the secondary analysis, rather than as preregistered predictions. RQ1 asks whether the perception of students and staff participants converges regarding the harmonized L0, L3 profile of AI integration. RQ2 investigates a "neutral" scalar position benchmark, specifically that with respect to the average insertion of practical learning, infrastructure, international alignment, and readiness ratings. RQ3 examines whether the perceived labour-market demand–preparedness gap remains positive under plausible assumptions about the unobserved within-person covariance. RQ4 asks what linked system elements should be prioritized in a measurable pathway for transitioning from a dispersed exposure to an integrated, sustainable performance.

The main hypothesis is as follows:

- H1 (system level convergence). If the normative low levels of curriculum maturity reported by both staff and students are to be indicative of an institutional rather than a role construct, then student and staff L0, L3 distributions should not differ at a significant level and have a small positive association.
- H2 (constructive alignment gap). If actual practice is underprovided with reference to its importance in learning, then the provision of practical, projects should be below the average (midpoint) of 3 and 4 and the subjective importance of practice should be above the average. This should result in a positive provision, importance gap.
- H3 (readiness capability mismatch). Supporting the view that the context for intervention in the 2014 curriculum will be one of complementary readiness and capacity, the institutional indexes for future institutional readiness should be above the average and the indices for the current infrastructure and the comparator (CCEE) should be below their averages with another positive gap.
- H4 (external alignment gap). Perceived labour, market demand should be significantly higher than perceived employer readiness for both stakeholder groups and the external, alignment difference should remain significant at identical coefficients of correlation.

Two systems propositions raise the observable tests into a near term future explanatory architecture. P1 posits institutional supports as indirectly shaping adoption readiness via academic capacity, curriculum maturity, and authentic practice. P2 assumes governance, ethics, quality assurance, assessment validity, and graduate/ employer feedback as comprehensive moderators and feedback monitors, paths not derived from the existing

aggregate data. The contribution of this study is threefold. Empirically, it provides parallel, multi-institutional stakeholder diagnostics using consistent denominators and explicit accounting for conceptual and measurement limitations. Methodologically, it integrates disparate aggregate indicators into an open, non-prescriptive diagnostic framework without claiming psychometric validation. Theoretically, it proposes an empirically falsifiable architecture linking innovation adoption and systems readiness through organisational capacity, curriculum implementation, ethical AI, and institutional governance.

MATERIALS AND METHODS

Study Design, Data Sources and Recruitment

This is a secondary cross, sectional survey of two project questionnaires using LimeSurvey electronically on the IQE survey site in January 2026. The first questionnaire was of students taking programs with STEM or related content, the second questionnaire was of academic staff involved in STEM teaching and/or research. The questionnaire links were circulated through the participating institutions and participation was voluntary. The student questionnaire was accessible free of charge, the staff questionnaire was circulated widely at Polytechnic University of Tirana (PUT) and University of Tirana (UT) and remained accessible to other institutions and was thus a nonprobability sample that was self-selected.

The source project-maintained counts of registered cases, and item, level aggregate statistics, but not a fully enumerated invitation frame from which a conventional response rate could be reconstructed. The staff survey was left open for longer than originally intended because response was low. Response rates and nonresponse weights could not therefore be calculated defensibly. Instead of assuming that nominal sample size indicated representativeness, the analysis reports institutional composition, item, specific denominators, Wilson confidence intervals for critical proportions, and denominator, sensitivity analyses where partial completion impacts interpretation.

Student survey, 664 entries in total were made in the nine public and private HEIs, of which some 282 were essentially complete. Valid number of items decreased in the later sections. Academic, staff survey, 92 entries in total. There were no further attempts to recruit additional participants for these secondary analyses. All reports of analysis (discussed below) are reconstructed from the source of project reports and their recorded aggregate data.

Participants and Survey Domains

The student sample was mainly from the Polytechnic University of Tirana (38.40%), the University of Medicine, Tirana (27.41%), and the University of Tirana (10.54%), with students from 6 more institutions. Bachelor students constituted the major part of the study, cycle sample. The staff sample was composed of the Polytechnic University of Tirana (27.40%), “Aleksandër Moisiu” University (27.40%), University of Tirana (21.92%),

University of New York Tirana (6.85%), and other institutions. STEM fields in the staff sample covered the areas of computing science and information technology, mathematics, physics and physical engineering, electrical/electronic engineering, mechanical engineering, biomedical, and others.

Table 2 depict the sample sizes for studies, design limitations and denominators for analysis.

Table 2. Sample Sizes for Studies, Design Limitations and Denominators for Analysis

Characteristic	Students	Academic staff
Registered survey entries	664	92
Substantially completed questionnaires	≈282	Not separately reported
Institutions represented	9	Multiple public/private HEIs
Main institutions	UPT 38.40%, UMT 27.41%, UT 10.54%	UPT 27.40%, UAMD 27.40%, UT 21.92%
Valid n, L0-L3 integration	220	50
Valid n, selected Likert tests	195-241	43-48
Sampling	Open voluntary / self-selection	Voluntary / self-selection
Conventional response rate	Not constructible	Not constructible

Instrument Structure, Measurement Evidence and Harmonization

The student and staff questionnaires covered aligned, although not identical, domains, including participant demographics, AI awareness and experience, AI use in courses, modules, projects, laboratories, placements, and research, curriculum content and skills, teaching and learning strategies, staff development, infrastructure, labour-market perceptions, constraints and priorities, and open-ended comments. The staff questionnaire contained 8 analytic subsections along with 5-point Likert items, single, and multiple answer items, and free text. The student questionnaire was similar in thematics, but not the same. The project reports did not describe a formal cognitive, interview pretest, an a priori psychometric validation study, or respondent, level item covariance matrices. As a result, Cronbach's α, McDonald's ω, EFA/CFA, measurement invariance, and structural equation modelling cannot be replicated with the available aggregate data.

Rather, this restriction is not taken as an authorization to produce psychometric results. The analysis explicitly constrained inference. Recent reviews demonstrate that AI literacy scales are inconsistent in construct definition and validation quality [17, 18]. Accordingly, we do not present the questionnaires as validated AI, literacy scales, and do not interpret self, reported preparedness as observed competence. Instead, only clearly defined individual indicators and harmonised categories are reported. Reported program integration is aligned to L0=no integration, L1=introductory module/course, L2=integration across several courses, and L3=dedicated AI course/module/program. Valid counts are 96, 46, 55, and 23 for students (N=220) and 26, 8, 11, and 5 for staff (N=50).

Systems Diagnostic Formulation

To add systems, science content without producing latent variables inconsistent with the data transparent rescaling diagnostics are used. Observed aggregate indicators are rescaled to a 0,100 scale only to accommodate nonuniform domains in a single visual domain for analysis. These variations are descriptive indicators, not validated psychometric measures. Reported curriculum maturity of stakeholder group g, as determined from the ordinal L0, L3, is as follows:

$$CAMI_g = \frac{100}{3N_g}\sum_{l=0}^{3} l\, n_{g,l} \tag{1}$$

where $n(g,l)$ is the number of valid responses in integration level l for stakeholder group g and Ng is the total number of valid L0-L3 responses in that group. A five-point Likert mean M is rescaled using,

$$S(M) = 25(M - 1) \tag{2}$$

The scores were linearly rescaled to a 0–100 range, such that a mean score of 1 corresponded to 0, the neutral midpoint of 3 corresponded to 50, and a mean score of 5 corresponded to 100. The normalized perceived demand-preparedness gap is as follows:

$$G_{DP} = 25(M_{demand} - M_{preparedness}) \tag{3}$$

These transformations represent a multidomain diagnostic profile rather than a compensatory composite index. While an overall Institutional AI Integration and Readiness Index (IAIRI) is presented for presentation, it is not estimated. This is because the component constructs and component weightings, as well as their measurement invariance, have not yet been validated at the respondent level. This suggests a non-compulsory decision, a high score in future institutional intention should not mathematically offset a low score in infrastructure or authentic practice.

Statistical Analysis, Robustness and Precision

Descriptive statistics are counts, percentages, means, standard deviations, and 95% confidence intervals. One-sample tests were used to compare the overall distribution across the four levels of integration among different stakeholder groups, while Cramér's V was used to quantify the effect size of the association. For a five-point Likert scale, the midpoint is 3, and Cohen's d should be calculated using the difference between the sample mean and the midpoint, divided by the sample standard deviation. The Holm adjustment is used separately within students and staff test families. Summary, level contrasts compare with Welch tests whenever sample sizes and variances are different, Hedges' g is the standardized effect. These procedures are inferential diagnostics of the aggregate indicators reported herein, they are not alternatives to multilevel or latent variable models for which respondent level data are needed.

For all key proportions, 95%Wilson score intervals [30] are given as they give more conservative interval estimates than the more popular Wald method which can be invalid with modest sample sizes or proportions away from 0.50. A post hoc power analysis indicated that the sample size (N = 270) provided approximately 80% statistical power at a significance level of $\alpha = 0.05$ to detect an association corresponding to Cramér's V = 0.20 (df = 3). The observed V=0.069 is thus interpretable as a small observed association, rather than as evidence of statistically equivalence. No prior power calculation was done, as the project surveys were not preregistered hypothesis, testing experiments. As the provided source reports do not preserve within, person covariance between perceived demand and preparedness, this revision performs a covariance, sensitivity analysis across assumed correlation r=0.0, 0.2, 0.4, 0.6, and reconstructs the standard deviation of the paired difference as,

$$SD_{diff}(r) = \sqrt{SD_{demand}^2 + SD_{prep}^2 - 2rSD_{demand}SD_{prep}} \quad (4)$$

with $SE(r) = SD_{diff}(r)/\sqrt{n}$. This procedure may not recover the true pair covariance but tests whether the substantive outcome is robust to a plausible range of correlation. Missing data sensitivity analysis is conducted separately by recomputing any selected student priority percentages with the entire registration denominator or with the item valid denominator demonstrating how prevalence interpretation varies by these denominators without imputing unasked questions. Open ended (or other non-closed category) responses were available for our present analysis solely by thematic summaries recorded in the source project reports, not as individual respondent text. This qualitative analysis was therefore used for triangulation of report, level data only and mapped only themes described in the reports to quantitative indicators, no inference about coding reliability, thematic saturation, grounded theory, or differences in inter-rater reliability is claimed.

Ethical Considerations

The original surveys were voluntary and anonymous, with no personally identifiable information, no sensitive personal data collected, and results reporting in aggregate form. Therefore, this research was considered minimal risk and did not require formal institutional ethics review using the appropriate institutional procedures. Electronic informed consent was presented prior to providing access to a questionnaire, a webpage was provided stating the purpose of the study, the voluntary participation of the individual, anonymity/confidentiality, use of the aggregated data for research purposes, and a statement informing the participant that they could stop participating at any time without penalty. Clicking to submit the questionnaire indicated the participant's informed consent in this study.

RESULTS

H1, Cross-Stakeholder Convergence in Reported Integration

L0 was identified for 43.64% of valid student integrations and 52.00% of staff integrations, and L3 was identified for 10.45% of student and 10.00% of staff valid integrations, see Table 3 and Figure 1. Four, category distributions failed to discriminate in any detectable way, χ^2 (3, N=270) =1.28, p=0.734, Cramér's V=0.069. H1 is thus supported only for the prespecified diagnostic. No stakeholder divergence was detected at the available precision and the observed association was small. This is not an equivalence result. Analytically, similar low CAMI values (34.1/100 students, 30.0/100 staff) are more compelling than the p-value alone because they show that both roles locate the bottleneck at the curriculum, maturity stage rather than recording contradictory institutional states.

Table 3. Harmonized AI-integration Levels with Wilson 95% Confidence Intervals

Integration level	Students n (%)	95% CI	Staff n (%)	95% CI
L0, No AI integration	96 (43.64%)	37.25-50.24%	26 (52.00%)	38.51-65.20%
L1, Introductory module/course	46 (20.91%)	16.06-26.76%	8 (16.00%)	8.34-28.51%
L2, Integrated in several courses	55 (25.00%)	19.74-31.12%	11 (22.00%)	12.75-35.24%
L3, Dedicated AI course/module	23 (10.45%)	7.07-15.20%	5 (10.00%)	4.35-21.36%
Total valid	220 (100%)	-	50 (100%)	-
Normalized CAMI	34.1/100	95% CI 29.4-38.8	30.0/100	95% CI 19.8-40.2

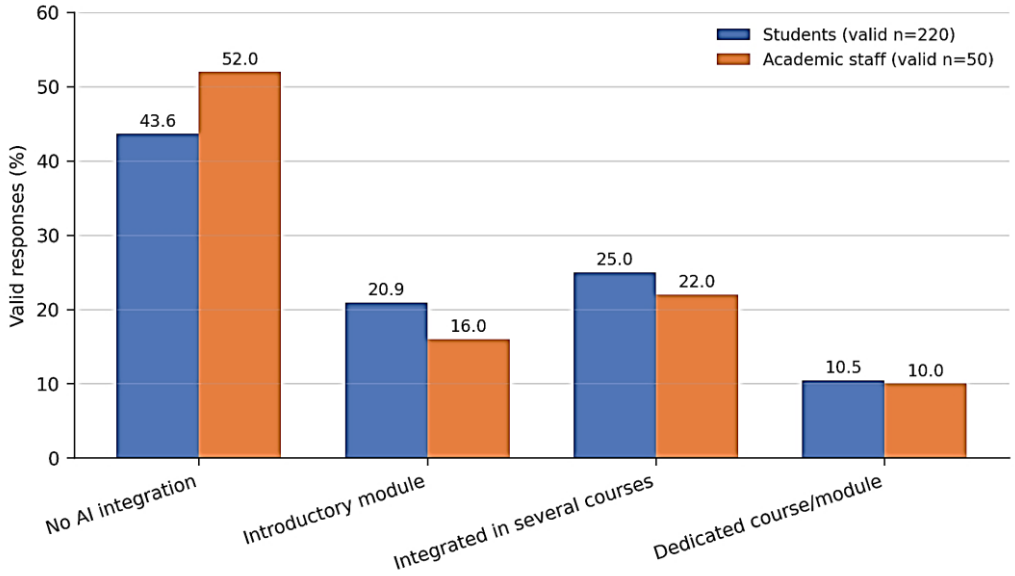


Figure 1. Harmonized Levels of Reported AI Integration Among Students and Academic Staff, Percentages Are Based on Valid Responses Within Each Group

H2, Practical Provision-Value Gap in the Student Experience

Students rated practical AI projects (M=2.41, SD=1.27, n=197), computer laboratories (M=2.50, SD=1.33, n=195), case studies (M=2.56, SD=1.27), and digital environments (M=2.78, SD=1.23, n=202) below the neutral midpoint after Holm adjustment. By contrast, practice as a learning mode was rated above neutral (M=3.39, SD=1.18, n=206, adjusted $p<0.001$, d=0.33). The direct summary-level contrast between perceived practice effectiveness and practical-project provision was 0.98 points (Welch t (395.5) =8.02, $p<0.001$, 95% CI 0.74-1.22, Hedges' g=0.80), equivalent to a 24.5-point difference on the normalized scale (59.8 vs 35.3). H2 is therefore supported. The analytical importance lies in the size and direction of the mismatch, the result is consistent with a constructive-alignment failure in which students value applied learning more strongly than the curriculum is reported to provide it, see Figure 2.

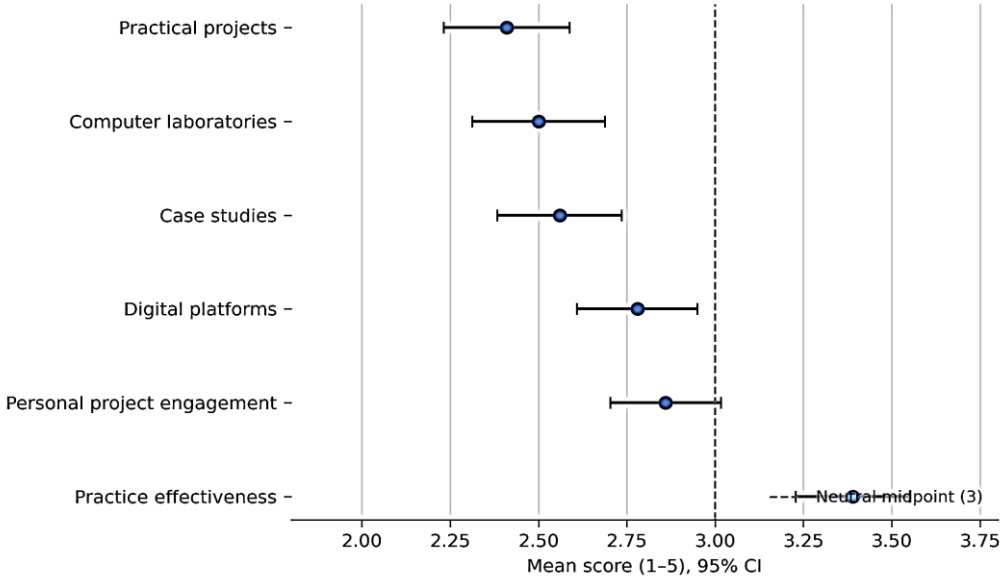


Figure 2. Satisfaction with Practical Learning Methods and Perceived Effectiveness. Error Bars Indicate 95% Confidence Intervals; The Dashed Line Marks the Neutral Midpoint (3).

Curriculum Depth, Informal Learning and Missing-Data Sensitivity

Self, directed learning was most commonly cited as a source of AI learning (26.05% expressed against all 664 registered entries), with formal program courses accounting for 6.78%. These full registration percentages are a lower bound descriptive statistic of prevalence in the population, as nonresponse increased in later sections, as shown by the denominator, sensitivity analysis. For the student, priority item for example, 109 students selected 'more practice', (versus a 16.42% of all 664 registered entries), but this was 54.23% of the 201 valid responses to that item (8.41% of the total registered entries). The same denominator effect holds for the other priorities in Table 4. The finding therefore supports the direction of priority expressed among respondents who answered the item, but not a national prevalence claim.

Table 4. Denominator-Sensitivity Analysis for Student Priorities

Priority	Count	% of 664 registered	% of 201 valid	Wilson 95% CI (valid)
More practice	109	16.42%	54.23%	47.33-60.97%
Better infrastructure	33	4.97%	16.42%	11.94-22.16%
Industry collaboration	26	3.92%	12.94%	8.98-18.28%
More AI courses	19	2.86%	9.45%	6.14-14.29%
Additional training	14	2.11%	6.97%	4.19-11.35%

The interpretation varies substantially with the denominator. Reporting only 16.42% would underestimate the tendency of the students who saw the item, while reporting only 54.23% may suggest unwarranted high occurrence in the full registered sample. Both figures are thus among the missing data robustness controls rather than correction for the sample selection effect.

H3, Staff Capacity, Infrastructure and Institutional Readiness

Staff ratings indicate a mismatch between current institutional capacity and future readiness, see Table 5. Infrastructure is rated as inadequate (mean 2.56, SD=1.03, n=48, adjusted p=0.015, effect size d=,0.43), as is current AI competence/content according to internationally accepted expectations (mean 2.59, SD=1.06, n=44, adjusted p=0.028, effect size d=,0.39), and three, year institutional readiness for change (mean 3.70, SD=0.94, n=43, adjusted p<0.001, effect size d=0.74), while actual institutional AI training-certification has been limited to 14.55% and only 34.55% reports an interest in current training opportunities. H3 is thus confirmed. From an organisational-readiness perspective [28], this pattern indicates that expectations of change exceed the perceived capacity to implement it and should not be interpreted as evidence that institutions possess the capabilities required for AI integration at scale.

Table 5. One-Sample Tests of Selected Likert Indicators Against the Neutral Midpoint (3)

Group	Indicator	n	Mean ± SD	t	Holm p	d
Students	Practical projects with AI	197	2.41 ± 1.27	-6.52	<0.001	-0.46
Students	Computer laboratories	195	2.50 ± 1.33	-5.25	<0.001	-0.38
Students	Digital platforms	202	2.78 ± 1.23	-2.54	0.047	-0.18
Students	Practice more effective than theory	206	3.39 ± 1.18	4.74	<0.001	0.33
Students	Personal labor-market preparedness	206	2.78 ± 1.30	-2.43	0.048	-0.17
Students	Perceived labor-market demand	196	3.36 ± 1.09	4.62	<0.001	0.33
Staff	Alignment with international practice	44	2.59 ± 1.06	-2.57	0.028	-0.39
Staff	Infrastructure adequacy	48	2.56 ± 1.03	-2.96	0.015	-0.43
Staff	Perceived student preparedness	45	2.58 ± 0.92	-3.06	0.015	-0.46
Staff	Perceived labor-market demand	44	3.75 ± 1.01	4.93	<0.001	0.74
Staff	Institutional readiness (3 years)	43	3.70 ± 0.94	4.88	<0.001	0.74

H4, Labor-Market Demand-Preparedness Gap and Covariance Sensitivity

On the diagnostic scale (0,100), students indicated a slightly below, neutral sense of preparedness (M=2.78, SD=1.30) and a above, neutral perception of demand (M=3.36, SD=1.09). Staff indicators of student demand revealed a lower sense of preparedness (M=2.58, SD=0.92) and a higher perception of demand (M=3.75, SD=1.01). Raw demand (demand preparedness) gaps (−0.58, SD=1.30 and 1.17 SD=1.01) corresponded to 14.5 and 29.3 points on the diagnostic scale. These were perceived external levels of alignment, vis-à-vis employer perceptions of demand or job applicants' skills, not actual market levels, see Figure 3.

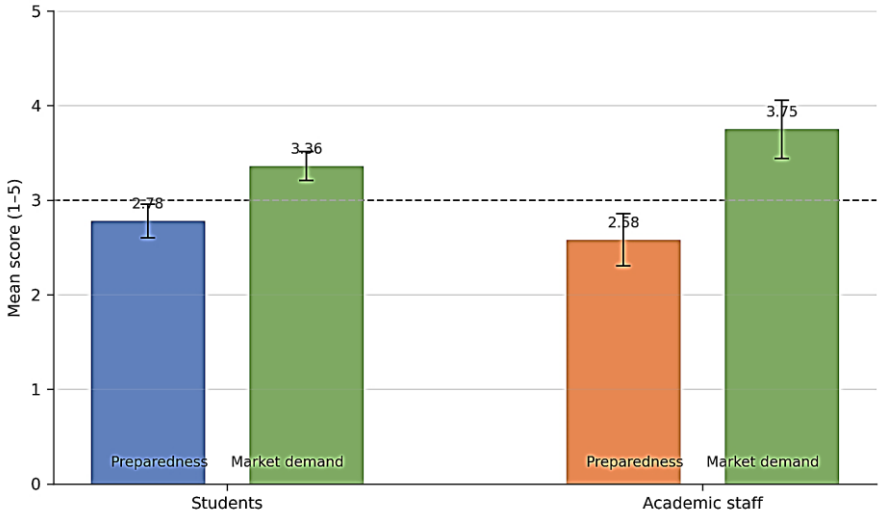


Figure 3. Perceived Labour-Market Demand and Preparedness Among Students and Academic Staff. Error Bars Indicate 95% Confidence Intervals.

As respondent-level covariance between demand and preparedness cannot be estimated, Table 6 presents a sensitivity analysis under alternative assumed correlations rather than incorporating observed correlations.

Table 6. Sensitivity of Demand-Preparedness Gap to the Assumed Within-Person Correlation

Group	Assumed r	Gap	95% CI	t	p	dz
Students	0.0	0.58	0.34-0.82	4.79	<0.001	0.34
Students	0.2	0.58	0.37-0.79	5.34	<0.001	0.38
Students	0.4	0.58	0.39-0.77	6.15	<0.001	0.44
Students	0.6	0.58	0.43-0.73	7.48	<0.001	0.53
Staff	0.0	1.17	0.75-1.59	5.68	<0.001	0.86
Staff	0.2	1.17	0.80-1.54	6.35	<0.001	0.96
Staff	0.4	1.17	0.85-1.49	7.32	<0.001	1.10
Staff	0.6	1.17	0.91-1.43	8.95	<0.001	1.35

In both samples, the mean gap remains positive and statistically significant ($p < 0.001$) under both assumed correlation values ($r = 0.0$ and $r = 0.6$). The implied within, person standardized impact varies from dz=0.34, 0.53 for students and from dz=0.86, 1.35 for staff. H4 proves resilient to this range of implied covariances, despite the inability to estimate

the magnitude of the associated paired, effect precisely. The apparent effect size difference shows up far more strongly in perceived alignment than it does in objective labour demand.

Barriers, Priorities and Precision of Staff Estimates

The top five barriers selected were lack of qualified staff (66%), infrastructure deficits (64%), outdated curriculum design (46%), lack of funding (42%) and weak industry linkage between AI education programs and industry demands (40%), see Table 7. The top five priorities were staff training (78%), curriculum review (68%), setting, up AI labs (62%), international collaboration (54%) and developing new AI programs (44%). The ordering is analytically important: capacity-building measures precede and are linked to stand-alone provision, consistent with the proposed upstream-to-downstream sequence of the system. Because respondents could add more than one choice, and n=50 results in wide Wilson confidence intervals, the following rankings are diagnostic priorities, not statistically demarcated tiers of preference, see Figure 4.

Table 7. Academic-Staff Barriers and Priorities with Wilson 95% Confidence Intervals (n=50)

Type	Item	n (%)	Wilson 95% CI
Barrier	Lack of qualified staff	33 (66%)	52.15-77.56%
Barrier	Insufficient infrastructure	32 (64%)	50.14-75.86%
Barrier	Rigid curriculum	23 (46%)	32.97-59.60%
Barrier	Insufficient funding	21 (42%)	29.38-55.77%
Barrier	Weak industry collaboration	20 (40%)	27.61-53.82%
Priority	Staff training	39 (78%)	64.76-87.25%
Priority	Curriculum updating	34 (68%)	54.19-79.24%
Priority	AI laboratories	31 (62%)	48.15-74.14%
Priority	International collaboration	27 (54%)	40.40-67.03%
Priority	New AI programs	22 (44%)	31.16-57.69%

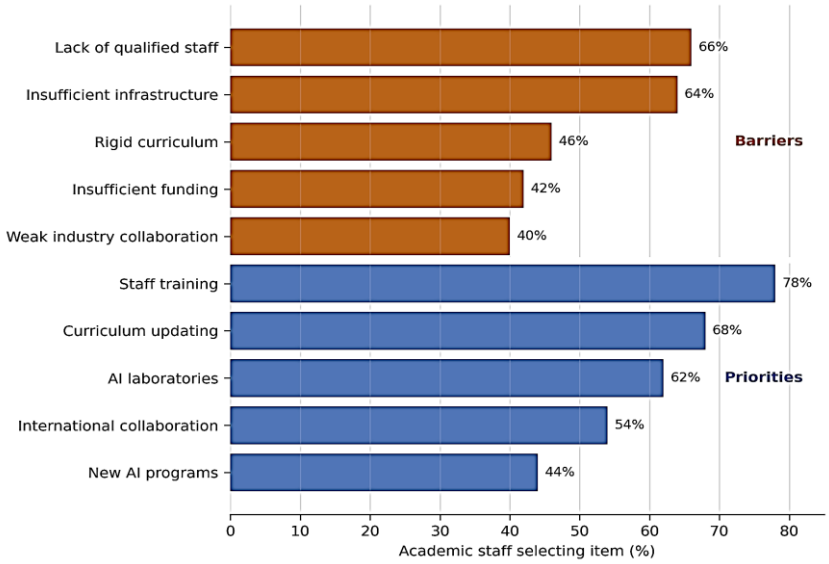


Figure 4. Barriers and Three-Year Priorities Identified by Academic Staff (Multiple-Response Items, N=50)

Report-Level Qualitative Triangulation

The source reports contain synthesized themes from open-ended responses. Because the original comments were not recoded in this revision, the qualitative material is used only for triangulation. Student responses emphasised the importance of authentic learning experiences, laboratory access, software availability, curriculum updates, real-world projects, industry engagement, and clear guidelines for responsible AI use. Key concerns included potential declines in critical thinking, factual inaccuracies, privacy risks, misuse, and over-reliance on AI. Staff-oriented themes emphasized training, infrastructure, curriculum flexibility, funding, and external partnership.

Table 8 maps those documented themes to quantitative indicators without assigning new frequencies or claiming an independent qualitative analysis.

Table 8. Report-Level Qualitative Triangulation of Documented Themes with Quantitative Indicators

Theme	Documented qualitative signal	Quantitative link	Triangulation
Authentic practice	More projects, laboratories, real datasets, applied tasks	Practical projects M=2.41, valid-priority item, 54.2% selected more practice	High convergence
Infrastructure access	Software, laboratories, accessible platforms	Students report limited access, staff infrastructure M=2.56, 64% staff barrier	High convergence
Curriculum flexibility	More visible and progressive AI integration	L0=43.6% students / 52.0% staff, student rigid-curriculum barrier documented	Convergent
Ethics and academic integrity	Clear rules, disclosure, verification, privacy, preserve critical thinking	Qualitative theme, not represented by a validated quantitative scale in current data	Qualitative-only support
Industry and employability	Real projects, internships, employer input	Demand exceeds preparedness in both groups, weak industry collaboration selected by 40% staff	Convergent
Staff development	Training linked to teaching practice	14.55% formally trained, training selected by 78% staff	High convergence

Systems Diagnostic Profile

As shown in figure 5, this is non-compulsory by design and it maintains domain imbalance (rather than averaging dissimilar constructs into a single readiness number). Reported curriculum maturity scores were approximately 30.3–34.1/100, while practical and project-based learning provision scored 35.3/100. Staff-rated infrastructure adequacy reached 39.0/100, and perceived international alignment scored 39.8/100. By contrast, projected institutional readiness over the following three years was substantially higher, at 67.5/100. The asymmetry on this signature is the key systems one. Expectations for future change lag current enabling conditions by a staggering amount. Put simply in a bottleneck

reading, more powerful intention cannot make up for weak upstream infrastructure, academic capability, curriculum alignment or authentic practice, these bottlenecks must be addressed sequentially.

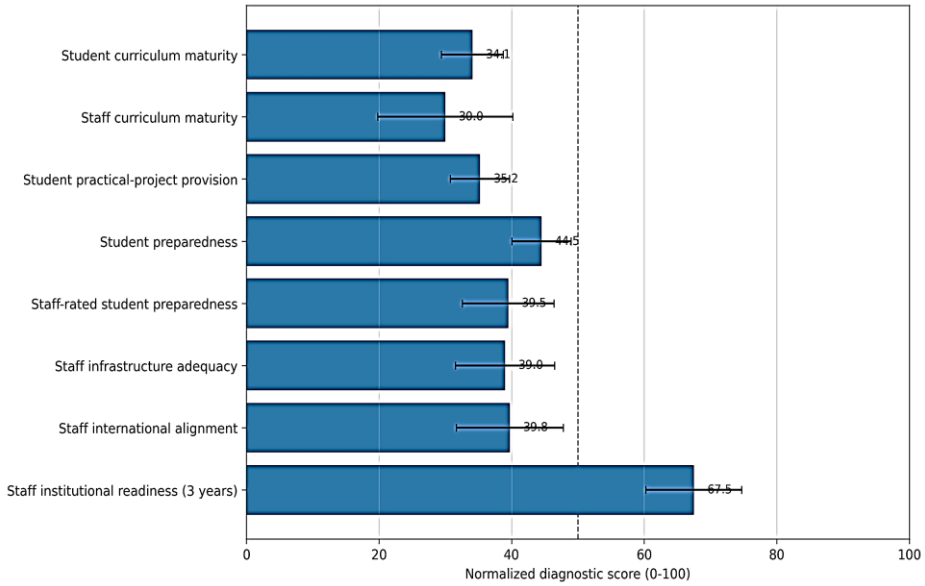


Figure 5. Normalized Systems-Readiness Diagnostic Profile for Selected Observed Indicators, Error Bars Show Transformed 95% Confidence Intervals and 50 Marks the Neutral Likert Midpoint Where Applicable

DISCUSSION

Principal Findings, Novelty and Contribution to Knowledge

None of the individual empirical findings—including limited reported integration, infrastructure constraints, the demand–preparedness gap, and the preference for more practical learning—are inherently novel. Rather, the contribution lies in their integration within a coherent analytical framework and the cautious interpretation of these signals within clearly defined empirical boundaries. The findings indicate an asynchronous readiness configuration in which student and staff perceptions converge around expectations of low curriculum maturity. Authentic practical learning is considered more important than its reported provision, while infrastructure and international calibration lag behind institutional readiness and demand–preparedness expectations, which exceed perceived labour-market strength. In sum, these discontinuities suggest a transition system where intent and external pressure are accelerating ahead of the mean, level educational system processes that are intended to convert them into capability.

Across stakeholder convergence strengthens the diagnosis but does not establish equivalence. The small observed Cramér's V and analogous CAMI values suggest that the two respondents are not actually characterizing diametrically opposite states of the curricula, which is important since stakeholder disagreement would raise a different set of measurement and communication concerns than convergence. The direction of the significant result is consistent with that of the authors' previous Albanian curriculum audit

[25], which also found uneven content focused around the computing activities. The two studies, however, use different units (page content vs. lessons) and cannot be combined statistically, and the audit provides objective program, level context while the current study depicts reported experience and reported readiness. Their contribution is triangulation among different sources of evidence, not testing one as a measure of the other.

The provision, value gap is the most educationally significant quantitative finding. The 0.98, point difference and g≈0.80 standards deviations tell us not merely that there is broad discontent among the students, but rather that there is disengagement between the learning mode capable of being effective in an AI, learning context and the opportunities students say they are receiving in the program. Constructive alignment [29] suggests that intended results, activities, and assessment should support each other. AI standards frameworks suggest that application, review, and creation [11, 12, 15] is a universal pedagogical imperative. And thus, the gap tells us that authentic practice is a bottleneck at the mechanism level, merely adding more AI, relevant performance objectives (more AI terminology, more individual modules) will not close the gap unless we expect students to have use for, capacity to verify, obligation to explain, and authority to make defences in discipline, relevant situations.

Theoretical Integration, From Adoption to Systems Readiness

TAM and UTAUT model technology acceptance at the level of the individual [26, 27], yet acceptance is only one aspect of educational implementation. A student or instructor could find AI to be useful and still lack the prerequisite infrastructure matching learning goal, authentic tasks, or assessment model. Weiner's college, based readiness theory [28] further distinguishes the institutional willingness to change from the communal on-site ability to implement change. The current pattern, relatively high long-term readiness, about three years, from the survey and a low level of current infrastructure, is indeed compatible with that distinction. Constructive alignment [29] then provides the pedagogical mechanism through which curriculum's maturity can pass through authentic practice, matching learning outcomes, activities, and assessment to reliably demonstrate competence.

The architecture is therefore purposely stratified. Adoption theory frames the use/intention layer, organization, readiness theory frames the resources and capabilities layer, constructive alignment and AI literacy frameworks shape how the curriculum transforms into incrementally advancing learning and assessment, and ecological policy/competency frameworks [14, 16] set the conditions for governance, ethics and professional development. An individual theory is not required to be responsible for all levels. This theoretical division of labour is what makes the system a system, readiness framework. Failure can happen at different nodes and succeeding at one stage does not

necessarily lead to downstream success unless neighbouring mechanisms are also improving.

Comparison with International State of the Art

The state-of-the-art comparison clarifies who and how. Authors in [21] explained AI, adoption within UTAUT among students and faculty, this paper builds on that work by norming the use, curriculum maturity, infrastructure, authentic practice, and readiness nodes as independent variables. The study in [20] measured the quality of engineering AI education through program scan and pilot module, here, the huge provision, value gap explains why authentic practice cannot be optional pedagogical content but a system bottleneck. The authors of [22] emphasised that engineering students' acceptance of AI and perceived efficacy extend beyond mere access, underscoring the need to distinguish technology adoption from demonstrated competence and pedagogical integration.

Measurement research is a necessary boundary condition. The studies in [17, 18] revealed that AI-literacy measurement has multiple dimensions as the seemingly similar instruments could have very different latent structures. The current questionnaires were not validated as latent AI literacy measures, as this revision does not manufacture Cronbach's α, CFA, SEM, mediation, or moderation coefficients from aggregate data summaries. Instead, the paper distinguishes three evidence classes, observed aggregate responses, transparent diagnostic transformation of that data, and untested causal propositions. That separation yields a methodological contribution because it recognizes a system rather than elevating imperfect data to psychometric idealism. MIT and ETH Zurich [23, 24] are incorporated solely as design benchmarks but not as numerical benchmarks. Their sophisticated and well, developed infrastructures exemplify interdisciplinary learning, faculty collaboration, research, based practice, and long, term computing access.

Albania differs considerably in scale of organization, resource availability, and program type therefore numerical benchmarks are irrelevant. The relevant comparison is structural, mature systems have interlocking curriculum, capability, infrastructure, authentic projects, and governance, whereas the Albanian data profile shows these constructs developing separately. When translated into local indicators, the framework enhances systems understanding without importing inapt external standards.

Systems Model, From Resources to Readiness

Figure 6 attempts to formalize the process suggested by both the theory and the diagnostic pattern. Institutional resources enable the development of academic capacity, which supports the design and delivery of increasingly aligned curricula. Greater curriculum maturity creates opportunities for authentic learning experiences, which, in turn, strengthen student preparedness. The effectiveness of this progression should ultimately be evaluated against graduate and employer outcomes. Governance, ethics quality assurance and assessment validity are factors spread throughout all nodes of the dynamic system, with graduate and employer evidence gathering where that cycle can

inform curriculum and resource decisions. For any particular resource/academics/gov/ethics/QA/assessment outcome dynamic, the process arrows are directional conjectures to be tested, not causal effects derived from the present aggregate data. The architecture builds upon these prior efforts in five respects. First, it disaggregates technology use and adoption from pedagogical and technological capabilities, positive attitudes and high use need not guarantee mature curricula. Second, it is multilevel, cascading down to constituents within the organization, the technology itself, the state, and the institutional sector.

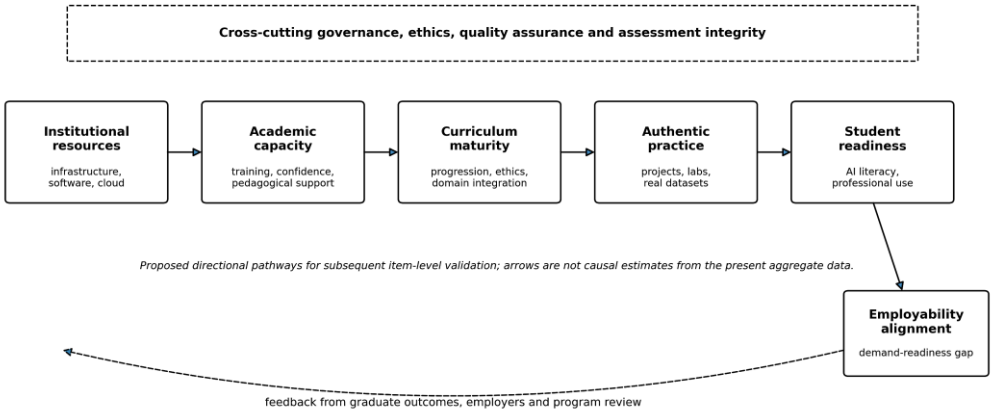


Figure 6. Proposed Systems-Readiness Architecture Linking Institutional Resources, Academic Capacity, Curriculum Maturity, Practical Learning, Student Readiness, and Employability. Arrows Indicate Hypotheses for Future Validation, Not Causal Estimates

Third, it is non-compulsory, high future intentions cannot compensate for weak infrastructure or practice in a composite indicator. Fourth, it is temporal and feedback sensitive, facilitating explicit transition pathways rather than static explanatory narratives. Fifth, it is evidence, constrained, observed causative indicators, identified but untested causal pathways, and diagnostic transformations are labelled independently. An overall IAIRI score is not calculated at this stage; its development requires prior validation of the component constructs, assessment of measurement invariance, establishment of theoretically and empirically justified weighting schemes, and modelling of institutional groupings.

$$X_{t+1} = F(X_t, U_t, G_t) + \varepsilon_t, \quad X_t = [R_t, A_t, C_t, P_t, S_t, E_t]^T \tag{5}$$

In the equation (5), X_t is the system-state vector for institutional resources (R), academic capacity (A), curriculum maturity (C), authentic practice (P), student readiness (S), and employability alignment (E), U_t represents intervention inputs and G_t represents cross-cutting governance/quality-assurance conditions. F is intentionally not estimated here. The formalization makes the model falsifiable, future longitudinal data can test whether

upstream improvements propagate downstream, whether bottleneck effects dominate, and whether governance or institutional context moderates those pathways.

Contribution to Knowledge

There are four specific incremental advances in this study. (1) It develops a non-compulsory system, readiness diagnostic profile that reveals asynchronous development across curriculum, practice, infrastructure, and future readiness not averaging together into one score. (2) It brings together student and staff evidence, unpatterned, in order to distinguish common system condition variables from role content variables (development levels). (3) It offers demonstrably useful respondent, level methods (denominator sensitivity, covariance sensitivity, interval estimation, explicit evidence classes) for incorporating diagnosis without respondent, level data. (4) It makes the diagnosis into a measurable transition architecture with nodes, directional hypotheses, cross, cutting governance, and feedback. The advances herein are incremental. They are not claims of a new universal theory of institutional transition, but they do advance the utility of diagnosis beyond a bland survey of low IA across administrative areas and populations toward a reproducible diagnostic framework.

Policy, Equity, Assessment and Quality-Assurance Implications

From a systems perspective, implementation sequence is consequential. Launching new AI programs before strengthening staff capability, practical infrastructure, curriculum pathways, and assessment design risks symbolic adoption, visible curricular change without the enabling conditions required for competence development. The staff priority ordering - training, curriculum updating, and laboratories ahead of new AI programs is consistent with this capacity-first logic and with UNESCO competency progression [15, 16], AI-across-the-curriculum [13], organizational readiness [28], and constructive alignment [29].

Equity is also a systems property. When AI learning depends mainly on autonomous use students with stronger prior digital skills, better devices, paid tools, English-language confidence, or external mentoring can accumulate advantage. Shared virtual laboratories, open-source environments, institutional cloud access, program-level minimum outcomes, and supported practice can reduce access disparities. Equity also requires pedagogy, students need explicit instruction in verification, bias, privacy, intellectual-property issues, disclosure, and the limits of model outputs [2, 15, 16].

Assessment is a critical control point because generative AI can produce text, code, and solutions that weaken the evidential value of product-only assessment [8,10]. The source-report comments favour clear rules over blanket prohibition. Process evidence - annotated code, oral defence, version history, portfolios, error analysis, reproducibility checks, and explicit disclosure of AI assistance can preserve visibility of individual reasoning. In constructive-alignment terms [29], assessment must verify the same disciplinary and AI competencies that curriculum and authentic practice are intended to develop.

Quality assurance should evaluate pathways rather than the mere presence of an 'AI course'. Program review can map where students encounter foundational concepts, disciplinary data, model evaluation, ethics, authentic projects, and progressively more demanding assessment. The earlier Albanian curriculum audit [25] provides an objective program-level baseline, the present study adds information about reported experience and institutional conditions. Keeping that evidence types separate allows quality assurance to ask a stronger question, not only whether AI appears in curricula, but whether institutional conditions allow the intended curriculum to become an experienced and assessable competence pathway.

Prioritized and Resource-Sensitive Implementation Roadmap

Table 9 translates the systems architecture into a staged implementation roadmap. Sequencing follows the hypothesized dependency structure, low- to moderate-cost actions that establish governance, curriculum maps, staff development, and shared environments precede wider program expansion. The targets are management indicators for implementation monitoring, not empirically validated international benchmarks. Their purpose is to make the framework auditable; each stage produces observable evidence that can be reviewed before resources are committed to the next stage.

Table 9. Staged Systems-Readiness Roadmap with Measurable Indicators

Horizon	Resource intensity	Priority actions	Illustrative measurable indicators
0-12 months	Low to moderate	Map L0-L3 outcomes, publish AI-use/disclosure guidance, launch discipline-specific staff workshops, establish shared open-source virtual lab	100% STEM programs mapped, institutional AI-use guidance approved, ≥25% staff complete applied workshop, baseline access metrics recorded
12-36 months	Moderate	Embed assessed AI projects, expand shared labs/cloud, introduce portfolio/oral-defense assessment, formalize industry co-design and guest expertise	≥50% mapped programs include assessed authentic project, ≥50% staff complete applied development, annual infrastructure uptime/access report, employer advisory feedback documented
3-5 years	Moderate to high	Validate AI-literacy and readiness measures, integrate QA indicators into program review, establish longitudinal graduate/employer follow-up, selectively expand advanced AI tracks	Validated measurement model, program review includes AI progression and ethics, graduate/employer follow-up ≥60%, advanced pathways tied to demonstrated capacity

Strengths, Limitations and Reproducibility

From a systems standpoint, implementation ordering matters. Introducing new AI programs before laying the groundwork with staff capability, logistical infrastructure, curriculum pathways, and assessment design risks symbolic adoption visible curriculum change without the enabling conditions needed for the development of competence. The staff to task ordering capacity building through training, curriculum update, and laboratory capacity before new AI programs is consistent with this capacity first reasoning and with UNESCO competence progression [15, 16], AI across the curriculum [13], organization readiness [28], and constructive alignment [29]. Equity too is a system property. When AI learning relies primarily on individual initiative, students with better prior digital skills, higher, end devices, paid tools, command of English, or external mentors can cache advantage. Shared virtual laboratories, open-source environments, institutional cloud access, program minimum outcomes, and supported practice can mitigate access inequality. Equity also demands pedagogy explicit instruction in documentation, bias, verification, privacy, intellectual property issues, disclosure, and the limits of model output [2, 15, 16]. Assessment is a leverage point because generative AI can produce text, code, and solutions that erodes the evidential value of product, only assessment practices [8, 10]. Report comments favour clear rules rather than blanket bans.

Process evidence annotated code, oral defence, version history, portfolios, error analysis, reproducibility checks, and voluntary disclosure of AI assistance can preserve transparency of individual reasoning. In terms of constructively aligned education [29], assessment must measure the same AI and disciplinary competencies that curriculum and real, world work intends to develop. Quality assurance should focus on pathways rather than the simple existence of an AI course. Program review can identify where students engage with course material on foundational concepts, data, model testing, ethics, real projects, and increasingly advanced assessment. The Albanian curriculum audit [25] previously provided an objective baseline of the program. This current document adds insights about context and reported experiences. Keeping evidence types separate allows quality assurance to deliver a better guidance does a specific institution's curriculum/pathway structure enable the intended pathway of doing and assessing competence?

CONCLUSION

The study does not establish the national prevalence of AI integration, objectively measured labour-market demand, or demonstrated AI competencies. It does, however, identify a reproducible pattern of perceived system-level readiness imbalances. Students and academic staff report similarly low curriculum maturity on the harmonised L0–L3 scale. Students place significantly greater value on practical learning than they perceive to be provided. Academic staff report limited current infrastructure adequacy and international alignment, alongside more optimistic expectations of institutional readiness over the following three years. Furthermore, both stakeholder groups perceive labour-

market demand to exceed preparedness, with this gap remaining positive across the range of within-person covariance assumptions examined.

The principal contribution lies in interpreting these combined mismatches through a systems-readiness perspective. AI adoption, organisational capacity, curriculum alignment, authentic practice, governance, and external relevance are treated as distinct but interconnected dimensions. This framework helps explain why institutional AI strategies and stand-alone courses may have limited impact when the underlying resources, academic capacity, assessment approaches, and opportunities for authentic application remain constrained. The originality of the study therefore lies not in identifying isolated weaknesses in AI integration, but in developing an evidence-constrained diagnostic architecture that identifies potential implementation bottlenecks, distinguishes perceived readiness from objectively validated competence, and establishes operationalizable pathways for subsequent investigation.

The findings also suggest a staged approach to institutional implementation. Priority should be given to defining measurable curriculum outcomes, establishing governance and disclosure requirements, strengthening academic capacity through discipline-specific applied training, developing shared physical and virtual environments for practical learning, and aligning assessment with demonstrable AI competencies. The expansion of AI-related provision should be informed by evidence that the necessary institutional and academic capabilities are in place. Although the cross-sectional design does not establish causal relationships, the findings provide an empirical baseline and a falsifiable framework for subsequent longitudinal, multilevel, performance-based, and employer-linked validation.

NOMENCLATURE

AI	Artificial intelligence
ARI	Academic Readiness/Capacity domain (proposed systems model)
CAMI	Curriculum AI Maturity Index, normalized ordinal L0-L3 profile
CI	Confidence interval
d	Cohen's standardized mean difference
g	Hedges' standardized mean difference
HEI	Higher-education institution
IAIRI	Institutional AI Integration and Readiness Index, proposed future composite, not estimated here
IRI	Infrastructure Readiness Index, normalized staff infrastructure rating
IQE	Institute for Quality in Education
L0-L3	Harmonized levels of reported AI integration
M	Arithmetic mean
n / N	Valid analytic count / total analytic count
p	Probability value
r	Assumed within-person correlation in sensitivity analysis

RQ	Research question
SAIRI	Student AI Readiness Index, proposed future validated scale
SD	Standard deviation
STEM	Science, technology, engineering, and mathematics
UPT	Polytechnic University of Tirana
UT	University of Tirana
UMT	University of Medicine, Tirana
UAMD	University "Aleksandër Moisiu" Durres
V	Cramér's V effect size
χ^2	Pearson chi-square statistic

AUTHOR CONTRIBUTIONS

Conceptualization, E.S., S.H., D.X., N.H. and A.G., methodology, E.S. and D.X., validation, D.X. and N.H., formal analysis, A.G., investigation, S.H. and D.X., resources, A.G., data curation, D.X. and E.S., writing-original draft preparation, D.X., writing-review and editing, D.X., N.H., E.S., S.H., A.G., visualization, E.S., K.Q.M.A.A., and S.H., supervision and project administration, D.X. All authors reviewed the minor-revision manuscript and approved the submitted version.

DATA AVAILABILITY

All the data's underlying the results can be provided by the corresponding author upon reasonable request

ACKNOWLEDGMENT

The authors acknowledge the participating students and academic staff, the Polytechnic University of Tirana, the University of Tirana, other participating higher-education institutions, and the Institute for Quality in Education for supporting survey dissemination and project implementation.

FUNDING

This research was funded by the National Agency for Scientific Research and Innovation (NASRI), Albania, through the project "Integration of AI Applications into University STEM Programs", contract no. 951/1, dated 12 June 2025.

CONFLICT OF INTERESTS

The authors declare no conflicts of interest. The funder had no role in the study design, data analysis, interpretation of the findings, preparation of the manuscript, or the decision to submit the work for publication.

REFERENCES

1. OECD. *OECD Digital Education Outlook 2023: Towards an Effective Digital Education Ecosystem*; OECD Publishing: Paris, 2023. Available online: https://www.oecd.org/en/publications/oecd-digital-education-outlook-2023_c74f03de-en.html (Accessed on 14 March 2026).
2. Miao, F., Holmes, W. *Guidance for Generative AI in Education and Research*, UNESCO: Paris, 2023. Available online: https://unesdoc.unesco.org/ark:/48223/pf0000386693?locale=en (Accessed on 14 March 2026).
3. Vuorikari, R., Kluzer, S., Punie, Y. *DigComp 2.2: The Digital Competence Framework for Citizens—With New Examples of Knowledge, Skills and Attitudes*, Publications Office of the European Union: Luxembourg, **2022**.
4. World Economic Forum. *The Future of Jobs Report 2025*, World Economic Forum: Geneva, 2025. Available online: https://www.weforum.org/publications/the-future-of-jobs-report-2025 (Accessed 21 June 2026).
5. Zawacki-Richter, O., Marín, V. I., Bond, M., Gouverneur, F. Systematic Review of Research on Artificial Intelligence Applications in Higher Education—Where Are the Educators? *Int. J. Educ. Technol. High. Educ.* **2019**, *16*, 39.
6. Crompton, H., Burke, D. Artificial Intelligence in Higher Education: The State of the Field. *Int. J. Educ. Technol. High. Educ.* **2023**, *20*, 22.
7. Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., et al. A Meta Systematic Review of Artificial Intelligence in Higher Education: A Call for Increased Ethics, Collaboration, and Rigour. *Int. J. Educ. Technol. High. Educ.* **2024**, *21*, 4.
8. McGrath, C., Farazouli, A., Cerratto-Pargman, T. Generative AI Chatbots in Higher Education: A Review of an Emerging Research Area. *High. Educ.* **2025**, *89*, 1533–1549.
9. Dwivedi, Y.K., Kshetri, N., Hughes, L., Slade, E.L., Jeyaraj, A., Kar, A.K., et al. "So What If ChatGPT Wrote It?" Multidisciplinary Perspectives on Opportunities, Challenges and Implications of Generative Conversational AI for Research, Practice and Policy. *Int. J. Inf. Manage.* **2023**, *71*, 102642.
10. Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., et al. ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education. *Learn. Individ. Differ.* **2023**, *103*, 102274.
11. Long, D., Magerko, B. What Is AI Literacy? Competencies and Design Considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ACM: New York, 2020, pp. 1–16.
12. Ng, D.T.K., Leung, J.K.L., Chu, S.K.W., Qiao, M. S. Conceptualizing AI Literacy: An Exploratory Review. *Comput. Educ.: Artif. Intell.* **2021**, *2*, 100041.
13. Southworth, J., Migliaccio, K., Glover, J., et al. Developing a Model for AI across the Curriculum: Transforming the Higher Education Landscape via Innovation in AI Literacy. *Comput. Educ.: Artif. Intell.* **2023**, *4*, 100127.
14. Chan, C. K. Y. A Comprehensive AI Policy Education Framework for University Teaching and Learning. *Int. J. Educ. Technol. High. Educ.* **2023**, *20*, 38.
15. Miao, F., Shiohira, K., Lao, N. *AI Competency Framework for Students*, UNESCO: Paris, **2024**.
16. Ioannou, M., Spyropoulou, N., Kameas, A. AI Competence Frameworks for STEAM Educators: A Knowledge Codification Perspective. *Encyclopedia* **2026**, *6*, 168.

17. Lintner, T. A Systematic Review of AI Literacy Scales. *npj Sci. Learn.* **2024**, *9*, 50.
18. Koch, M.J., Wienrich, C., Straka, S., Latoschik, M.E., Carolus, A. Overview and Confirmatory and Exploratory Factor Analysis of AI Literacy Scale. *Comput. Educ.: Artif. Intell.* **2024**, *7*, 100310.
19. Mikeladze, T., Meijer, P.C., Verhoeff, R.P. A Comprehensive Exploration of Artificial Intelligence Competence Frameworks for Educators: A Critical Review. *Eur. J. Educ.* **2024**, *59*, e12663.
20. Hao, Y., Liu, Y., Liu, B., Amarantidis, G., Ghannam, R. Integrating AI in Engineering Education: A Comprehensive Review and Student-Informed Module Design for U.K. Students. *IEEE Trans. Educ.* **2025**, *68*(2), 173–185.
21. Kim, J., Klopfer, M., Grohs, J.R., Eldardiry, H., Weichert, J., Cox, L.A., II, et al. Examining Faculty and Student Perceptions of Generative AI in University Courses. *Innov. High. Educ.* **2025**, *50*, 1281–1313.
22. Bernabei, M., Colabianchi, S., Falegnami, A., Costantino, F. Students' Use of Large Language Models in Engineering Education: A Case Study on Technology Acceptance, Perceptions, Efficacy, and Detection Chances. *Comput. Educ.: Artif. Intell.* **2023**, *5*, 100172.
23. Kumar, A.N., Raj, R.K., Aly, S.G., Anderson, M.D., Becker, B.A., Blumenthal, R.L., Eaton, E., Epstein, S. L., Goldweber, M., Jalote, P., et al. Computer Science Curricula 2023, ACM Press, IEEE Computer Society Press, AAAI Press, **2023**.
24. ETH Zurich AI Center. *Courses*. Available online: https://ai.ethz.ch/education/courses.html (Accessed on 22 June 2026).
25. Xhako, D., Hyka, N., Gjevori, A., Muda, V., Duro, Ç., Demirneli, M., Spahiu, E., Hoxhaj, S. The Level of AI Application in University STEM Study Programs: A Comprehensive Review. *Int. J. Innov. Technol. Interdiscip. Sci.* **2025**, *8* (4), 1244–1283.
26. Davis, F.D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Q.* **1989**, *13* (3), 319–340.
27. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D. User Acceptance of Information Technology: Toward a Unified View. *MIS Q.* **2003**, *27*(3), 425–478.
28. Weiner, B.J. A Theory of Organizational Readiness for Change. *Implement. Sci.* **2009**, *4*, 67.
29. Biggs, J. Enhancing Teaching through Constructive Alignment. *High. Educ.* **1996**, *32*, 347–364.
30. Wilson, E.B. Probable Inference, the Law of Succession, and Statistical Inference. *J. Am. Stat. Assoc.* **1927**, *22*(158), 209–212.