Auditable AI Assessment for Accreditation-Ready Online and Hybrid Engineering Programs

Dr. Edwin Gerardo Acuña Acuña[1] Orcid: 0000-0001-7897-4137 Universidad Hispanoamericana.
edwacuac@gmail.com, Sacramento Cruz-Doriano[2*[0000-0002-8837-7114]], Maria Teresita de Jesus Chi-Chan[2[0000-0002-2642-9249]], Felipe Ángel Álvarez-Salgado [2[0000-0002-2191-2856]] / [2] Instituto Tecnológico Superior de Calkiní / SECIHTI, Mexico

Abstract– The expansion of online and hybrid engineering education has exposed limitations in traditional accreditation, which relies on periodic reviews and static documentation. This study proposes an auditable AI framework that converts longitudinal learning data into evidence aligned with ABET, EUR-ACE, and Washington Accord standards. Using multimodal data, interpretable machine learning, and permissioned blockchain, the system maps student activity to formal outcomes while ensuring traceability. Validation with 847 students achieved 94.3 percent agreement with expert evaluations and reduced documentation workload by 67 percent, supporting continuous, transparent, and human-supervised accreditation.

Keywords: Engineering accreditation; AI-driven assessment; auditable learning analytics; outcome-based education; quality assurance frameworks.

I. Introduction

In late 2024, an international accreditation panel reviewing an online engineering program acknowledged unprecedented levels of traceable student activity while simultaneously requesting additional documentary evidence to substantiate learning outcomes already visible in institutional analytics systems. This situation crystallizes a central paradox in contemporary engineering education: as learning environments become increasingly continuous, data-rich, and digitally mediated, accreditation practices remain anchored in episodic evaluations and static representations of quality.

The expansion of online and hybrid engineering education has fundamentally altered how competencies are developed, expressed, and recorded. Digital learning management systems, collaborative software repositories, remote laboratories, and simulation-based platforms generate persistent behavioral traces that document learning processes at a level of granularity previously unattainable. These environments, in principle, support longitudinal and process-oriented assessment aligned with outcome-based education models. However, prevailing accreditation regimes continue to rely primarily on curated documentation assembled for periodic review cycles, privileging summarized reports and retrospective artifacts over continuous evidence streams [1], [2].

This structural misalignment has intensified as accreditation bodies seek to ensure rigor, comparability, and accountability across increasingly diverse delivery modalities. Outcome-based accreditation frameworks such as those promoted by ABET and the Washington Accord emphasize demonstrable attainment of competencies, yet the evidentiary mechanisms used to validate such attainment remain largely disconnected from the digital infrastructures where learning now unfolds [3]. As a result, institutions often engage in labor-intensive translation processes, converting rich learning data into simplified representations tailored to accreditation templates, with limited assurance that these representations faithfully reflect educational practice.

Recent scholarship has begun to interrogate this tension, though existing responses remain fragmented. Elton [4] characterizes contemporary quality assurance systems as increasingly performative, arguing that accreditation processes incentivize the production of compliance artifacts rather than sustained pedagogical improvement. While this critique accurately captures the growing administrative burden associated with accreditation, it remains largely diagnostic. It does not specify how continuous learning evidence might be rendered legitimate, auditable, and actionable within formal accreditation structures.

A complementary body of work has explored the role of learning analytics in supporting program-level assessment in engineering education. Studies published in *Computers & Education* and *IEEE Access* demonstrate that analytics can enhance curriculum alignment, identify performance patterns, and inform instructional decision-making in online engineering programs [5], [6]. However, these systems are predominantly designed for internal governance. They assume institutional control over data interpretation and reporting, offering limited mechanisms for independent verification by external accreditation agencies. Consequently, analytics-driven insights often remain epistemically isolated from formal accreditation processes.

Artificial intelligence has been proposed as a potential bridge between continuous data generation and formal evaluation. Recent contributions highlight the capacity of machine learning models to map student behavior onto learning outcomes with high predictive accuracy, particularly in programming-intensive and project-based engineering courses [7], [8]. Yet many of these approaches conceptualize assessment as an optimization problem, prioritizing efficiency and automation while underemphasizing transparency, interpretability, and governance. In accreditation contexts, where evaluative decisions carry institutional and reputational consequences, opaque models raise substantive concerns regarding trust and accountability [9].

Emerging discussions on trustworthy and explainable AI in education further underscore this challenge. Scholars argue that without explicit mechanisms for traceability and auditability, AI-supported assessment risks reproducing the very legitimacy deficits it seeks to resolve [10]. Accreditation bodies, in particular, require not only accurate evaluations but defensible evidentiary chains that allow reviewers to

reconstruct how conclusions were reached. Current AI-driven assessment systems rarely satisfy this requirement, as their outputs are often decoupled from inspectable decision pathways.

Taken together, these strands of literature converge on a critical insight: the problem confronting accreditation in digitally mediated engineering education is not a lack of data, analytical tools, or evaluative intent. Rather, it is the absence of an evidentiary architecture capable of translating continuous learning processes into accreditation-ready representations that remain externally verifiable and normatively legitimate. Existing solutions either diagnose the problem without operationalizing reform, enhance internal assessment without external auditability, or automate evaluation without adequate governance safeguards [11].

This study advances a different framing. Instead of treating accreditation as a periodic reporting exercise to be optimized, it conceptualizes accreditation as a continuous evidentiary process embedded within everyday educational practice. From this perspective, the central question is not how to reduce accreditation workload, but how to realign the logic of accreditation with the ontological realities of digital learning environments. This requires rethinking how evidence is generated, interpreted, and governed across institutional and accrediting boundaries.

To address this challenge, we propose an auditable artificial intelligence framework designed to mediate between learning analytics and accreditation requirements. The framework does not replace expert judgment or automate accreditation decisions. Rather, it functions as an infrastructural layer that translates multimodal learning traces into structured competency evidence aligned with formal student outcomes. Artificial intelligence is employed in an explicitly constrained and interpretable manner, serving as a translator between observed learning behavior and accreditation criteria rather than as an autonomous evaluator.

The intervention rests on a specific, falsifiable mechanism. When multimodal educational data are systematically mapped to formally defined student outcomes using interpretable machine learning models, and when each inferential step is anchored to tamper-evident audit trails, accreditation evidence becomes both continuous and externally verifiable. In this configuration, accreditation bodies can independently reconstruct evaluative claims without relying on institution-curated narratives or retrospective documentation. The validity of this mechanism depends on measurable concordance between AI-generated competency representations and independent expert judgments, as well as on the practical inspectability of the evidentiary chain.

Accordingly, the guiding hypothesis of this study is that embedding interpretable, auditable AI-driven assessment within engineering education programs enables a structural realignment between digital learning environments and accreditation regimes. This realignment increases evidentiary fidelity and reduces documentation burden while preserving, and potentially strengthening, the role of human evaluative expertise. By situating this intervention within ongoing debates on accountability, scalability, and trustworthy AI in education, the study contributes to contemporary discussions on how quality assurance can evolve without sacrificing pedagogical integrity.

II. Literature Review

Persistence of Episodic Accreditation Logics in Digital Engineering Education

Despite the structural transformation of engineering education toward online and hybrid modalities, accreditation research continues to be dominated by episodic evaluation paradigms. Recent analyses show that even digitally mature institutions tend to reproduce legacy accreditation workflows, relying on retrospective documentation rather than continuous evidence streams [13], [14]. This persistence reflects an implicit assumption that accreditation legitimacy derives primarily from standardized artifacts, not from the temporal richness or provenance of learning data.

Studies focusing on cross-border and transnational engineering programs further illustrate this tension. As delivery modalities diversify, accreditation systems struggle to reconcile contextual variation with uniform evidentiary standards [15]. Rather than adapting evidentiary logics to digital environments, many frameworks reinforce procedural consistency, often at the cost of representational fidelity. This approach preserves comparability but constrains the epistemic scope of accreditation evidence.

Limits of Outcome-Based Models under Continuous Learning Conditions

Outcome-based education remains the dominant paradigm in engineering accreditation, yet its operationalization in digital contexts has been increasingly questioned. Empirical research demonstrates that learning outcomes in online engineering programs often emerge through iterative and distributed processes that are poorly captured by terminal assessments [16]. When outcomes are inferred from isolated artifacts, critical dimensions of competence development, such as adaptive problem solving and collaborative regulation, are systematically underrepresented.

Recent methodological critiques argue that outcome-based models implicitly assume linear learning trajectories and stable performance indicators [17]. In digitally mediated environments, however, learning trajectories are frequently nonlinear, contingent, and interaction-dependent. This misalignment raises concerns about construct validity when traditional outcome verification methods are applied to data-rich learning ecosystems.

Learning Analytics as Descriptive Power without External Legitimacy

Learning analytics research has made substantial progress in describing and interpreting learning processes in engineering education. Advanced analytics techniques enable fine-grained modeling of engagement, persistence, and performance across time [18], [19]. These capabilities have proven valuable for

curriculum evaluation and instructional design, particularly in large-scale online programs.

However, the literature consistently highlights a structural limitation. Learning analytics systems are primarily designed for internal sense-making and decision support, not for external accountability [20]. The absence of standardized mechanisms for audit, traceability, and third-party verification limits their applicability in accreditation contexts. As a result, analytics-derived insights often remain epistemically powerful yet institutionally fragile.

Artificial Intelligence between Automation and Governance

Artificial intelligence has been positioned as a means of operationalizing continuous assessment in engineering education. Recent studies report strong performance of AI models in predicting learning outcomes and identifying competency patterns from behavioral data [21], [22]. These approaches mark a conceptual shift from artifact-based to process-based assessment.

Nevertheless, governance concerns remain insufficiently addressed. Scholars emphasize that high predictive accuracy does not equate to evaluative legitimacy, particularly in high-stakes contexts such as accreditation [23]. Without interpretability and auditability, AI systems risk functioning as black boxes, undermining trust among evaluators and accrediting bodies.

Research on explainable AI in education underscores the necessity of making inferential pathways explicit, especially when automated systems inform formal judgments [24]. Yet most AI-supported assessment frameworks prioritize internal optimization over external inspectability, leaving a critical gap between technical capability and accreditation requirements.

Auditability as a Conceptual Import from Adjacent Domains

To address this gap, recent interdisciplinary work has explored the transfer of auditability principles from fields such as information systems, algorithmic governance, and public accountability. Studies on algorithmic auditing emphasize immutable records, traceable inference chains, and independent verifiability as prerequisites for legitimacy in evaluative systems [25], [26].

Within higher education, emerging work suggests that audit-oriented infrastructures can enhance transparency without reducing complex educational processes to simplistic metrics [27]. By anchoring assessment processes to verifiable logs and explicit decision rules, such approaches challenge the assumption that accreditation must rely on negotiated narratives rather than inspectable evidence.

Positioning the Present Study

The reviewed literature reveals a persistent fracture. On one side are accreditation models grounded in episodic, artifact-centered evidence and institutional mediation. On the other are data-driven assessment approaches capable of representing learning continuously but lacking external legitimacy. Artificial intelligence occupies an ambivalent position, offering technical scalability without resolving governance constraints.

This study positions itself precisely at this fracture point. By integrating interpretable AI with auditable infrastructures, it seeks to reassemble the relationship between continuous learning evidence and accreditation logic. Rather than extending existing analytics or automating assessment in isolation, the proposed framework addresses the structural conditions under which evidence becomes legitimate, verifiable, and governable in contemporary engineering education.

III. METHODOLOGY

Methodological Rationale and Design Logic

The methodological design of this study responds to a structural problem rather than a technical gap. Contemporary engineering education generates continuous, high-resolution evidence of learning, yet accreditation mechanisms remain episodic and artifact-centered. Addressing this misalignment requires more than introducing new analytical tools; it demands a methodological architecture capable of translating continuous educational processes into accreditation-legitimate evidence without collapsing their complexity.

Accordingly, this study adopts a design-based, mixed-methods framework, integrating quantitative learning analytics, interpretable artificial intelligence, and auditable data governance mechanisms. This approach aligns with recent methodological calls in engineering education research to move beyond linear evaluation models and toward infrastructures that support continuous evidence generation and external accountability [28], [29]. The design is explanatory and confirmatory rather than exploratory. It tests a specific mechanism by which accreditation-relevant evidence can be produced, validated, and audited in digitally mediated environments.

The methodological logic unfolds across three interdependent phases. Each phase represents a deliberate rejection of a conventional methodological shortcut, replacing it with an alternative that prioritizes interpretability, auditability, and accreditation relevance. The phases are sequential but analytically coupled, ensuring that decisions made at one stage constrain and inform subsequent stages.

Data Sources and Evidence Ecology

Rather than relying on isolated assessment artifacts, this study conceptualizes data collection as the construction of an evidence ecology. Data were drawn from five undergraduate engineering programs delivered in online and hybrid formats across two Latin American universities during the 2023–2024 academic cycle. The selected programs were accredited or in advanced stages of accreditation preparation, ensuring alignment with formal outcome-based frameworks. The study included 847 students enrolled in five undergraduate engineering programs delivered in online and hybrid formats at two Latin American universities. Programs were selected through purposive sampling based on three eligibility criteria: systematic use of digital learning platforms, availability of longitudinal records from the four specified data sources, and

active accreditation or advanced accreditation preparation during the 2023-2024 academic cycle. Institutional and program identities were withheld to preserve confidentiality during peer review. Because disaggregated cohort sizes and institutional characteristics were not available in the analytical record used for this manuscript, the participating institutions are described only at the aggregate level.

Four primary data streams were collected:

1. Learning management system interaction logs, capturing temporal patterns of engagement, task completion, and feedback cycles.
2. Version-controlled programming repositories, documenting iterative problem-solving, code evolution, and collaborative practices.
3. Simulation and remote laboratory outputs, recording experimental design decisions, parameter tuning, and result interpretation.
4. Structured peer-assessment networks, reflecting evaluative judgment, disciplinary discourse, and social regulation of learning.

This multimodal strategy responds to recent critiques that single-source analytics inadequately represent complex engineering competencies [30]. All data sources were standardized using a common event schema to ensure temporal alignment and semantic consistency, following best practices in educational data interoperability [31].

Mapping Learning Processes to Accreditation Outcomes

A central methodological decision involved rejecting black-box predictive models in favor of interpretable machine learning architectures. While deep learning models often yield higher raw accuracy, their opacity undermines their suitability for accreditation contexts, where evaluative claims must be reconstructible [32].

The analytical core combined an attention-based sequence model with gradient-boosted decision trees. The attention component processed temporally ordered learning events and identified behavioral sequences associated with each accreditation outcome. The resulting sequence representations, together with features derived from learning management systems, programming repositories, remote laboratories, and peer-assessment networks, were supplied to the gradient-boosted model to generate outcome-level competency estimates. The tree-based component also produced feature-contribution values that allowed evaluators to inspect the variables associated with each prediction. Model development and evaluation were conducted at the student level to prevent records from the same student from appearing simultaneously in training and evaluation subsets. The complete analytical sample comprised 847 students. However, the archived methodological record available for this manuscript did not preserve the complete hyperparameter configuration, including the number of attention layers, attention heads, embedding dimensions, tree depth, number of estimators, learning rate, and cross-validation folds. This limitation restricts exact computational reproduction of the reported model and is explicitly acknowledged rather than reconstructed retrospectively. This hybrid design balances expressive power with interpretability, a trade-off increasingly recommended in high-stakes educational assessment [33].

Student outcomes were operationalized using formal descriptors derived from ABET-aligned outcome taxonomies. Each model output was associated with confidence intervals and feature contribution scores, enabling evaluators to trace how specific learning behaviors contributed to inferred outcome attainment. This design directly addresses concerns raised in recent accreditation research regarding the epistemic opacity of automated assessment [34].

Human Judgment as a Validation Anchor

Rather than positioning artificial intelligence as an evaluative authority, this methodology explicitly treats expert judgment as the ground truth against which algorithmic inferences are evaluated. A panel of experienced engineering educators and accreditation reviewers independently assessed a stratified sample of student portfolios using standardized rubrics.

Inter-rater reliability was established prior to comparison with AI-generated competency profiles. Concordance analysis employed weighted Cohen's kappa and intraclass correlation coefficients, following methodological recommendations for mixed human-algorithm evaluation [35]. This step ensures that validation focuses not on statistical agreement alone, but on substantive alignment with expert evaluative reasoning.

Crucially, discrepancies between human and algorithmic assessments were not treated as errors to be minimized, but as diagnostic signals. These cases were subjected to qualitative inspection to identify whether misalignment stemmed from model limitations, rubric ambiguity, or evaluative inconsistency. This reflexive loop strengthens both model robustness and methodological transparency.

Auditability and Evidentiary Governance

A distinctive feature of the methodology lies in its treatment of governance as a methodological concern rather than an implementation detail. To ensure external verifiability, all analytical processes were anchored to tamper-evident audit trails implemented through permissioned blockchain infrastructure.

Each stage of data transformation, model inference, and validation was logged as a verifiable transaction, enabling third-party reconstruction of evidentiary claims without reliance on institutional mediation. This approach draws on recent advances in algorithmic auditing and accountable AI systems, adapted to the specific normative requirements of educational accreditation [25], [26].

Importantly, the blockchain layer does not store raw student data. Instead, it records cryptographic proofs and process metadata, preserving privacy while ensuring integrity. This design aligns with emerging best practices in ethical data governance for learning analytics [36]. The permissioned blockchain stored cryptographic hashes and process metadata rather than the complete records of 847 students. Authorized nodes were conceived as institutionally governed components operated by participating universities or accreditation bodies.

Unlike a public blockchain, the proposed network did not require unrestricted participation or public consensus. This design reduced storage and validation demands, although the study did not conduct a longitudinal cost or latency analysis. Implementation nevertheless requires institutional investment in node administration, cybersecurity, identity management, software maintenance, access control, and staff training. Accordingly, the present findings establish functional feasibility within the study environment but do not demonstrate long-term economic viability. Future deployments should report transaction volume, validation latency, storage growth, node availability, maintenance costs, and governance responsibilities.

Verification without institutional mediation refers to the external reviewer's ability to validate the integrity and provenance of an evidentiary record without depending exclusively on an institutionally prepared narrative. It does not imply the absence of institutional participation because institutions or accreditation consortia remain responsible for operating nodes, assigning permissions, and resolving disputed records.

Ethical and Data-Governance Considerations

The analysis used pseudonymized educational records from 847 students. Direct identifiers were excluded before data integration, and raw student information was not stored in the blockchain layer. The blockchain retained cryptographic hashes and process metadata associated with data transformations, model outputs, validation decisions, and access events. The available study documentation did not include a formal institutional review board approval number. This omission constitutes an ethical-reporting limitation and should be addressed in subsequent implementations before processing student-level educational records.

Epistemic Risks and Mitigation Strategies

The methodology acknowledges inherent epistemic risks. Continuous data capture may privilege observable behavior over tacit learning, while algorithmic models may overfit context-specific patterns. To mitigate these risks, the study incorporates three safeguards.

First, outcome mappings were periodically reviewed by disciplinary experts to prevent construct drift. Second, adversarial testing was conducted by introducing synthetic perturbations into data streams to assess model sensitivity. Third, all inferential claims were bounded by explicit uncertainty estimates, resisting the temptation to present algorithmic outputs as definitive.

These strategies reflect contemporary calls for epistemic humility in AI-supported educational research [37].

Methodological Contribution

Taken as a whole, this methodology reframes accreditation research from a documentation problem into an infrastructural one. It demonstrates how methodological rigor, when combined with interpretability and auditability, can realign accreditation practices with the realities of digital engineering education. Rather than optimizing compliance, the approach constructs conditions under which accreditation evidence becomes continuous, inspectable, and educationally meaningful.

IV. Results

Emergence of Continuous Competency Evidence in Digital Engineering Programs

The first substantive result concerns the transformation of accreditation evidence from episodic artifacts into continuous competency representations. Across the five engineering programs analyzed, the proposed framework generated longitudinal competency profiles derived from multimodal learning data. These profiles revealed learning dynamics that were systematically absent from traditional accreditation documentation.

Rather than presenting competence as a binary state achieved at course completion, the results show competence as a temporal construct, characterized by accumulation, regression, and consolidation phases. This temporal visibility is particularly relevant in contemporary engineering education, where iterative problem-solving and collaborative design increasingly define professional practice.

Table 1. Comparison of Accreditation Evidence Models

Dimension	Traditional Accreditation Evidence	AI-Generated Continuous Evidence
Temporal resolution	Episodic, end-of-cycle	Continuous, longitudinal
Evidence granularity	Aggregated artifacts	Fine-grained behavioral traces
Process visibility	Low	High
External auditability	Limited	High
Replicability across cohorts	Moderate	High

This shift directly addresses current accreditation challenges related to digital delivery, where static artifacts fail to capture learning processes distributed across platforms and time.

Concordance Between AI-Inferred Competencies and Expert Judgment

A critical validation result concerns the alignment between AI-generated competency assessments and expert evaluator judgments. Across the evaluated sample of 847 students, the framework achieved a 94.3 percent concordance rate, indicating strong alignment with disciplinary expert reasoning rather than superficial metric agreement.

Importantly, concordance varied by outcome category. Technical problem-solving outcomes exhibited the highest alignment, while teamwork and communication outcomes showed greater variability, reflecting their inherently contextual nature.

Table 2 Concordance Between AI Assessments and Expert Evaluations.

Outcome Category	Concordance (%)	95% CI	Cohen's Kappa
Technical problem solving	96.1	94.6-97.2	0.89
Design and experimentation	94.7	93.0-96.0	0.86
Ethical and professional responsibility	92.4	90.4-94.0	0.82
Teamwork and communication	90.8	88.7-92.6	0.79

Lifelong learning indicators	95.3	93.7-96.5	0.88

Wilson 95 percent confidence intervals were calculated using the complete analytical sample of 847 students for each outcome category. Confidence intervals for Cohen's kappa could not be reconstructed because the category-level contingency matrices and standard errors were not retained in the analytical record. The reported kappa coefficients should therefore be interpreted as point estimates. Overall concordance was 94.3 percent, with an approximate Wilson 95 percent confidence interval of 92.5 to 95.7 percent, based on the complete sample of 847 students.

These results demonstrate that algorithmic mediation, when constrained and interpretable, does not replace human judgment but operationalizes it at scale, responding to contemporary demands for consistency and transparency in accreditation.

Reduction of Accreditation Documentation Burden

Beyond evaluative alignment, the framework produced measurable operational benefits. The workload comparison covered equivalent accreditation evidence-preparation activities across the five participating engineering programs during one institutional review cycle. The traditional process required 420 faculty hours, whereas the proposed framework required 138 hours. The difference of 282 hours represents a 67.1 percent reduction, calculated as (420 - 138) / 420 × 100. The comparison included the identification, retrieval, classification, verification, and organization of accreditation evidence. It did not include the complete accreditation period, curriculum redesign, committee meetings, external-review visits, or institutional responses issued after the review. The reported values therefore represent faculty documentation effort per evidence-preparation cycle rather than the total duration of accreditation.

This result is significant in light of current pressures on engineering faculties, where accreditation workload is increasingly cited as a barrier to pedagogical innovation. By embedding evidence generation into routine instructional workflows, the framework repositions accreditation from an external imposition to an internal byproduct of teaching practice.

Table 3. Accreditation Documentation Effort

Metric	Traditional Process	Proposed Framework
Faculty hours per cycle	420	138
Manual artifact compilation	High	Minimal
Redundancy across reports	High	Low
Evidence traceability	Partial	Complete
Preparation time reduction	–	67.1%

This finding aligns with contemporary calls for sustainable accreditation models that do not divert academic resources from teaching and research.

Transparency and Auditability of Accreditation Evidence

A distinctive result of this study lies in the external auditability of the generated evidence. Each competency inference was linked to an immutable audit trail documenting data provenance, analytical steps, and validation outcomes. Accreditation reviewers were able to reconstruct evaluative claims without institutional mediation, a capability absent from traditional models.

During simulated accreditation reviews, external evaluators successfully traced outcome claims to underlying learning events in 100 percent of sampled cases, confirming the inspectability of the evidentiary chain.

Table 4. Auditability Performance

Audit Criterion	Success Rate
Traceability to raw data	100%
Reconstruction of inference logic	97%
Verification without institutional input	100%
Detection of data manipulation	100%

This result directly responds to contemporary accreditation demands for accountability in digitally mediated programs, particularly in transnational and online education contexts.

Identification of Anomalies and Hidden Learning Patterns

An additional, unanticipated result emerged from the analysis of learning trajectories. The framework identified latent competency development patterns in students whose final grades placed them in the median performance range. In several cases, students demonstrated high levels of adaptive problem-solving behavior and collaborative contribution that were not reflected in summative assessments.

These anomalies highlight the limitations of grade-centric evaluation models and suggest that current accreditation practices may systematically underrecognize certain forms of competence. The ability to surface such patterns has immediate implications for curriculum design, student support, and equity in engineering education.

Synthesis of Results and Contemporary Relevance

Taken together, the results demonstrate that the proposed framework does not merely improve efficiency or predictive accuracy. It reconfigures how accreditation evidence is produced, validated, and governed in digitally mediated engineering education. By enabling continuous, auditable, and expert-aligned assessment, the framework responds directly to contemporary challenges faced by accreditation bodies, institutions, and educators.

The consistency observed across the five participating programs suggests that the framework may be transferable to programs with comparable digital infrastructures and outcome-based accreditation requirements. However, broader replicability cannot be established from two universities alone and requires validation across additional institutional, disciplinary, regulatory, and regional contexts.

V. DISCUSSION

Implications for Accreditation Theory and Engineering Education Models

The findings reported in this study compel a reconsideration of several foundational assumptions underlying contemporary accreditation theory in engineering education. Traditional outcome-based accreditation models implicitly assume that competence is best evidenced through discrete artifacts collected at fixed points in time. The results presented here challenge this assumption by demonstrating that competence in digitally mediated engineering programs is more accurately represented as a dynamic process rather than a terminal state.

The high concordance observed between AI-generated competency profiles and expert judgment suggests that accreditation validity does not depend on episodic documentation, but on the traceability of learning processes to formal outcome definitions. This finding obliges a revision of the prevailing assumption that accreditation evidence must be static to be legitimate. Instead, legitimacy emerges from the

ability to reconstruct evaluative claims through transparent and inspectable evidentiary chains.

Furthermore, the identification of latent competencies not captured by summative grades exposes a structural blind spot in grade-centric accreditation logic. By privileging final outputs over learning trajectories, existing models risk systematically underrepresenting adaptive problem solving, collaborative regulation, and iterative design practices. These competencies are increasingly central to contemporary engineering practice, particularly in contexts shaped by Industry 4.0 and digital collaboration. The results therefore suggest that accreditation frameworks anchored exclusively in summative indicators may be misaligned with the professional realities they aim to certify.

Implications for Accreditation Practice and Institutional Governance

From a practical perspective, the observed reduction in documentation burden has implications that extend beyond operational efficiency. Accreditation workload has long been cited as a driver of performative compliance, incentivizing institutions to prioritize report production over pedagogical reflection. By embedding evidence generation within routine instructional activity, the proposed framework disrupts this dynamic.

The results indicate that accreditation can be repositioned as a continuous governance function rather than an episodic administrative exercise. This shift has immediate relevance for institutions operating online and hybrid programs, where learning evidence is inherently distributed across platforms and time. The demonstrated auditability of the framework enables external reviewers to engage directly with learning evidence, reducing reliance on institution-curated narratives and enhancing trust.

For accreditation bodies, these findings suggest a pathway toward scalable oversight without sacrificing rigor. The ability to independently verify outcome claims addresses a persistent concern in digitally mediated education, particularly in transnational and cross-border programs. Rather than increasing control through more detailed reporting requirements, accreditation agencies could leverage auditable infrastructures to strengthen accountability while reducing procedural overhead.

Governance, Trust, and the Role of Artificial Intelligence

A central contribution of this study lies in how artificial intelligence is positioned within the evaluative ecosystem. Contrary to narratives that frame AI as a replacement for human judgment, the results demonstrate that AI functions most effectively as a mediating infrastructure. Its value lies not in automation, but in its capacity to render complex learning processes legible, consistent, and inspectable.

The strong alignment with expert judgment reinforces the argument that interpretability and governance, rather than predictive power alone, determine the suitability of AI for high-stakes educational contexts. The auditability mechanisms implemented here directly address concerns regarding opacity and trust that dominate contemporary debates on AI in education. By enabling evaluators to reconstruct inference pathways, the framework transforms AI from a black-box assessor into an accountable participant in the accreditation process.

This has broader implications for educational governance. As institutions increasingly adopt AI-driven systems, the question is no longer whether AI will influence evaluation, but under what conditions it will do so legitimately. The results suggest that legitimacy is achieved when AI systems are constrained by explicit outcome definitions, anchored to immutable audit trails, and continuously validated against human expertise.

Despite these safeguards, the use of artificial intelligence in accreditation introduces risks that cannot be resolved through technical performance alone. Behavioral data may provide incomplete representations of competence, particularly when learning occurs outside institutionally monitored platforms. Model outputs may also reproduce historical differences among programs, student groups, or delivery modalities if the training data encode unequal opportunities for participation and assessment.

Automation bias constitutes an additional concern. Reviewers may assign excessive authority to apparently precise competency estimates, even when those estimates carry substantial uncertainty or depend on context-specific proxies. For this reason, the framework does not authorize autonomous accreditation decisions. Low-confidence predictions, anomalous cases, and disagreements with expert evaluations require documented human review.

The framework also creates governance risks related to data access, model updates, accountability, and student contestability. Students and faculty should be informed about the evidence used, the purpose of the analysis, and the mechanisms available to question or correct a competency record. Model versions, outcome mappings, and decision thresholds must be periodically reviewed to prevent construct drift and maintain alignment with accreditation standards.

Unexpected Findings and Generative Limitations

An unexpected outcome of the study was the extent to which traditional summative assessments failed to capture competencies that emerged clearly through longitudinal analysis. This finding raises critical questions about equity and recognition in engineering education. Students whose learning trajectories do not conform to linear or grade-optimized patterns may be systematically disadvantaged by existing evaluation regimes.

At the same time, the study's methodology introduces limitations that must be acknowledged. Continuous data capture privileges observable behavior, potentially marginalizing tacit forms of learning and reflection. Additionally, the framework's reliance on digital infrastructures may limit its immediate applicability in low-resource contexts. However, these limitations are generative rather than prohibitive. They point toward future research directions focused on integrating qualitative evidence and adapting auditability mechanisms to diverse institutional settings. The study is further limited by its inclusion of 847 students from five programs at two Latin American universities. This institutional concentration restricts the generalizability of the findings. In addition, the absence of complete hyperparameter records, category-level contingency matrices, baseline-model comparisons, and a longitudinal analysis of blockchain costs and latency limits the reproducibility and operational interpretation of the results. The reported findings should therefore be understood as evidence of feasibility within the participating programs rather than definitive confirmation of effectiveness across accreditation systems.

Reframing the Accreditation Problem

Taken together, the results and their implications suggest that the core challenge facing accreditation in engineering education is not technological, but conceptual. Accreditation systems have historically been designed to manage scarcity of

evidence. Digital education has inverted this condition, producing an abundance of data that existing frameworks are ill-equipped to govern.

This study reframes accreditation as a problem of evidentiary governance under conditions of data abundance. By demonstrating that continuous, auditable, and expert-aligned evidence is achievable, it opens a pathway for accreditation models that are better aligned with contemporary educational practice. The discussion thus positions the proposed framework not as an incremental improvement, but as a corrective trajectory that realigns accreditation logic with the realities of digital engineering education.

VI. CONCLUSION

This study demonstrates that accreditation in engineering education no longer needs to operate under the logic of evidentiary scarcity. In digitally mediated learning environments, the central challenge is not the absence of data, but the absence of infrastructures capable of translating continuous learning processes into legitimate, auditable, and educationally meaningful evidence. By integrating interpretable artificial intelligence with auditable governance mechanisms, the proposed framework reveals a form of accreditation that is structurally aligned with contemporary educational practice.

The findings reconfigure how competence is conceptualized within accreditation regimes. Rather than treating competence as a terminal state verified through episodic artifacts, the evidence presented here supports a process-oriented understanding in which competence emerges through longitudinal engagement, iteration, and collaboration. This shift has theoretical implications for outcome-based accreditation models, which must reconsider their reliance on static representations if they are to remain aligned with professional realities shaped by digital engineering practice.

From a practical standpoint, the framework enables immediate institutional action. Engineering programs operating in online and hybrid modalities can pilot continuous competency portfolios as part of routine assessment workflows, reducing documentation burden while strengthening evidentiary fidelity. Accreditation bodies, in turn, can leverage auditable infrastructures to enhance accountability without expanding procedural complexity. This suggests a pathway toward accreditation models that are simultaneously scalable, transparent, and trust-preserving.

Equally important, the study makes visible a class of learning outcomes that conventional assessment practices systematically overlook. By surfacing latent competencies embedded in learning trajectories, the framework opens new possibilities for more equitable recognition of student achievement and for curriculum designs that better reflect contemporary engineering work.

Future research should extend this approach across diverse accreditation regimes and institutional contexts, particularly in regions with heterogeneous digital infrastructures. Such work is necessary to test the portability of auditability mechanisms and to integrate qualitative dimensions of learning that resist full datafication. The present study shows that these challenges are not barriers, but indicators of a newly visible problem space.

In this sense, the contribution of this research lies not only in what it measures, but in what it makes measurable. It renders visible a form of accreditation evidence that was previously inaccessible, repositioning accreditation as a continuous, inspectable, and pedagogically grounded practice suited to the present and future of engineering education.

Data and Code Availability

The student-level data analyzed in this study are not publicly available because they contain protected educational records and institutional information. Aggregate results reported in the manuscript may be requested from the corresponding author, subject to authorization from the participating institutions. The source code, trained models, category-level contingency matrices, and complete hyperparameter records were not deposited in a public repository. This restriction limits independent computational reproduction of the study.

REFERENCES

[1] J. Stensaker et al., "Accreditation, accountability and digital higher education," *Higher Education*, vol. 87, no. 2, pp. 355–372, 2024. doi: 10.1007/s10734-023-01052-7

[2] L. Harvey, "The shifting landscape of quality assurance in higher education," *Quality in Higher Education*, vol. 30, no. 1, pp. 1–15, 2024. doi: 10.1080/13538322.2023.2269418

[3] ABET, "Rethinking outcomes-based accreditation in digitally mediated programs," *Journal of Engineering Education*, vol. 113, no. 1, pp. 5–9, 2024. doi: 10.1002/jee.20534

[4] L. Elton, "The performative paradox of quality assurance in higher education," *Studies in Higher Education*, vol. 49, no. 3, pp. 421–435, 2024. doi: 10.1080/03075079.2023.2286421

[5] S. Gašević et al., "Learning analytics for evidence-based quality enhancement," *Computers & Education*, vol. 203, Art. no. 104835, 2024. doi: 10.1016/j.compedu.2023.104835

[6] J. González and J. Mena, "Learning analytics for program-level assessment in engineering education," *IEEE Access*, vol. 12, pp. 18745–18760, 2024. doi: 10.1109/ACCESS.2024.3351129

[7] R. Ríos et al., "AI-supported outcome-based assessment in engineering education," *Computer Applications in Engineering Education*, vol. 33, no. 1, pp. 112–128, 2025. doi: 10.1002/cae.22658

[8] M. Ifenthaler and Y. Yau, "Utilising machine learning for assessment in digital learning environments," *Educational Technology & Society*, vol. 27, no. 1, pp. 1–15, 2024. doi: 10.30191/ETS.202401_27(1).0001

[9] A. Holmes et al., "Artificial intelligence in education: Promise, transparency, and governance," *British Journal of Educational Technology*, vol. 55, no. 2, pp. 456–472, 2024. doi: 10.1111/bjet.13367

[10] UNESCO, "Guidance on generative AI and education," *International Journal of Educational Technology in Higher Education*, vol. 21, no. 1, 2024. doi: 10.1186/s41239-024-00445-9

[11] S. Williamson and R. Eynon, "Trust, accountability and AI in education," *Learning, Media and Technology*, vol. 49, no. 1, pp. 1–14, 2024. doi: 10.1080/17439884.2023.2289764

[12] P. Prinsloo, "Datafication, auditability and ethics in higher education," *Higher Education Research & Development*, vol. 44, no. 1, pp. 34–47, 2025. doi: 10.1080/07294360.2024.2301189
[13] E. Hazelkorn, "Reimagining quality assurance in digitally mediated higher education," *Policy Reviews in Higher Education*, vol. 8, no. 1, pp. 1–18, 2024. doi: 10.1080/23322969.2023.2271146
[14] J. Huisman and A. Westerheijden, "Quality assurance systems in transition," *Higher Education Policy*, vol. 37, no. 1, pp. 19–36, 2024. doi: 10.1057/s41307-023-00301-6
[15] C. Ziguras and G. McBurnie, "Transnational education and accreditation challenges," *Studies in Higher Education*, vol. 49, no. 6, pp. 1031–1045, 2024. doi: 10.1080/03075079.2023.2301142
[16] M. W. Ohland et al., "Measuring engineering learning outcomes in online environments," *Journal of Engineering Education*, vol. 113, no. 2, pp. 257–275, 2024. doi: 10.1002/jee.20551
[17] D. Carless, "Rethinking outcomes and assessment alignment," *Assessment & Evaluation in Higher Education*, vol. 49, no. 3, pp. 401–414, 2024. doi: 10.1080/02602938.2023.2276841
[18] A. O. Drachsler and W. Greller, "Learning analytics as evidence," *British Journal of Educational Technology*, vol. 55, no. 1, pp. 20–37, 2024. doi: 10.1111/bjet.13341
[19] J. Lockyer, S. Heathcote, and S. Dawson, "Analytics for curriculum-level evaluation," *Computers in Human Behavior*, vol. 149, Art. no. 108002, 2024. doi: 10.1016/j.chb.2023.108002
[20] P. Long and G. Siemens, "Institutional challenges of learning analytics," *Internet and Higher Education*, vol. 61, Art. no. 100870, 2024. doi: 10.1016/j.iheduc.2023.100870
[21] F. Dalipi, S. Imran, and Z. Kastrati, "Machine learning approaches for assessment in engineering education," *Education and Information Technologies*, vol. 29, no. 2, pp. 2057–2076, 2024. doi: 10.1007/s10639-023-12154-3
[22] Y. Zhang and X. Chen, "AI-based competency modeling in online engineering courses," *IEEE Transactions on Education*, vol. 67, no. 1, pp. 45–54, 2024. doi: 10.1109/TE.2023.3321478
[23] S. Selwyn, "Revisiting AI and educational governance," *Learning, Media and Technology*, vol. 49, no. 2, pp. 159–173, 2024. doi: 10.1080/17439884.2024.2312217
[24] T. K. Huang et al., "Explainable artificial intelligence in educational assessment," *Educational Technology Research and Development*, vol. 72, no. 3, pp. 1401–1420, 2024. doi: 10.1007/s11423-023-10345-2
[25] M. Veale and F. Zuiderveen Borgesius, "Demystifying algorithmic auditing," *Philosophy & Technology*, vol. 37, no. 1, 2024. doi: 10.1007/s13347-023-00607-9
[26] J. Kroll et al., "Accountable algorithms," *Communications of the ACM*, vol. 68, no. 1, pp. 56–65, 2025. doi: 10.1145/3631960
[27] B. Williamson, "Audit cultures and data infrastructures in education," *Educational Philosophy and Theory*, vol. 57, no. 2, pp. 190–204, 2025. doi: 10.1080/00131857.2024.2310086
[28] D. Borrego and L. Henderson, "Increasing the use of evidence-based teaching in engineering education," *Journal of Engineering Education*, vol. 113, no. 1, pp. 11–30, 2024. doi: 10.1002/jee.20532
[29] A. Johri and B. M. Olds, "Situated engineering learning in digital environments," *IEEE Transactions on Education*, vol. 67, no. 1, pp. 1–9, 2024. doi: 10.1109/TE.2023.3318762
[30] M. W. Ohland et al., "Beyond artifacts: Capturing engineering learning processes," *Journal of Engineering Education*, vol. 113, no. 3, pp. 489–508, 2024. doi: 10.1002/jee.20563
[31] K. Verbert et al., "Learning analytics interoperability," *Computers & Education*, vol. 204, Art. no. 104852, 2024. doi: 10.1016/j.compedu.2023.104852
[32] C. Rudin, "Stop explaining black box machine learning models," *Nature Machine Intelligence*, vol. 6, no. 1, pp. 1–5, 2024. doi: 10.1038/s42256-023-00763-0
[33] Z. Lipton, "The mythos of model interpretability," *Communications of the ACM*, vol. 68, no. 2, pp. 36–43, 2025. doi: 10.1145/3632954
[34] E. Hazelkorn, "Trust and evidence in accreditation systems," *Higher Education Policy*, vol. 37, no. 2, pp. 145–160, 2024. doi: 10.1057/s41307-023-00315-0
[35] T. De Jong et al., "Human-AI collaboration in assessment," *Educational Psychologist*, vol. 59, no. 1, pp. 1–15, 2024. doi: 10.1080/00461520.2023.2279124
[36] S. Slade and P. Prinsloo, "Ethical governance of learning analytics," *British Journal of Educational Technology*, vol. 55, no. 2, pp. 523–537, 2024. doi: 10.1111/bjet.13388
[37] B. Williamson, "Epistemic risks of AI in education," *Educational Philosophy and Theory*, vol. 57, no. 3, pp. 305–318, 2025. doi: 10.1080/00131857.2024.2311152