I.J. Modern Education and Computer Science, **2026, 3, 121-134**
Published Online on June 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijmecs.2026.03.08

Evaluating LLM-Assisted CLO–PLO Alignment Decisions for AUN-QA–Aligned Curriculum Mapping

Suwut Tumthong*
Rajamangala University of Technology Suvarnabhumi, Ayutthaya, Thailand
E-mail: suwut.t@rmutsb.ac.th
ORCID iD: https://orcid.org/0009-0003-9610-6431
*Corresponding author

Nichanun Samakthai
Rajamangala University of Technology Suvarnabhumi, Ayutthaya, Thailand
E-mail: nichanun.s@rmutsb.ac.th
ORCID iD: https://orcid.org/0000-0003-2461-4077

Pinyaphat Tasatanattakool
Rajamangala University of Technology Suvarnabhumi, Ayutthaya, Thailand
E-mail: pinyaphat.t@rmutsb.ac.th
ORCID iD: https://orcid.org/0000-0002-3296-669X

Received: December 15, 2025; Revised: January 28, 2026; Accepted: February 24, 2026; Published: June 8, 2026

Abstract: Curriculum mapping for AUN-QA is often time-consuming and prone to inconsistency because learning-outcome evidence is scattered across multiple courses and program documents. This study examines the reliability of large language models in supporting course learning outcome–program learning outcome (CLO–PLO) alignment decisions for AUN-QA–aligned curriculum mapping. Using Mechatronics Engineering curriculum materials (2022–2024), AUN-QA v4.0 indicators were operationalized into a structured evidence schema to guide prompting and interpretation. GPT-4 produced 120 CLO–PLO alignment decisions, which were independently annotated by two domain experts and adjudicated by a third expert to establish reference labels. Model–reference agreement was evaluated using precision, recall, F1-score, and Cohen's kappa (κ), yielding 0.89, 0.85, 0.87, and 0.81, respectively. We also developed a dashboard that summarizes alignment coverage and supports PDCA-based curriculum improvement by flagging potential gaps and redundancies. The findings suggest that LLM-assisted alignment can reduce mapping workload and improve auditability while remaining consistent with expert judgment, enabling more scalable evidence-based AUN-QA evaluation.

Keywords: AUN-QA Framework, Curriculum Management, Large Language Models, Learning Outcome Alignment, Quality Assurance in Higher Education, Educational Data Analytics

1. Introduction

Quality assurance (QA) plays a critical role in certifying curriculum quality, ensuring that educational programs adhere to clearly defined standards and continuously improve to achieve intended learning objectives [1]. The QA process is essential for maintaining the accuracy, relevance, and effectiveness of academic programs, which underpin educational provision. It provides a systematic approach to evaluating and enhancing various aspects of teaching and learning, including learning outcomes, instructional methods, and assessment strategies [2]. Such a systematic approach is crucial for both traditional and online modes of education, ensuring that all students receive high-quality learning experiences regardless of the delivery format [3]. As universities increasingly transition toward digital transformation, global frameworks have emphasized the need for evidence-based curriculum management and transparent evaluation processes [4]. However, QA activities remain predominantly manual, time-consuming, and heavily reliant on expert interpretation of qualitative data, which creates challenges for achieving consistency, efficiency, and scalability in QA [5].

Within the ASEAN region, the ASEAN University Network Quality Assurance (AUN-QA) framework serves as a central mechanism for benchmarking curriculum quality across higher education institutions [6]. Despite its robustness, the evaluation process remains predominantly manual and document-intensive, relying heavily on the review of TQF reports and Self-Assessment Reports (SAR). This dependence on human interpretation frequently leads to inconsistencies among assessors and variability in judgment [7]. The emerging literature has highlighted the need to modernize this process. Tumthong et al. [8] underscored the need to integrate intelligent support systems to automate selected components of the AUN-QA evaluation and reduce subjective variation. Recent advancements in Natural Language Processing (NLP) and Large Language Models (LLMs), such as GPT-4 and PaLM-2, offer new opportunities to enhance quality assurance processes by enabling near-human-level interpretation of complex academic texts [9]. In particular, Jayalath et al. [10] demonstrated that BERT-based models can automatically classify Course Learning Outcomes (CLOs) and link them with corresponding Program Learning Outcomes (PLOs), a task traditionally requiring significant expert effort. These findings suggest strong potential to leverage AI-driven semantic analysis to enhance reliability, reduce evaluator workload, and support objective outcome alignment in AUN-QA–aligned curriculum mapping at the CLO–PLO decision level.

An initial review of TQF documents collected between 2022 and 2024 revealed inconsistencies across course reports, particularly in CLO articulation and the justification of assessment methods. This variability underscores the practical limitations of manual QA workflows in the Thai higher education context, where maintaining decision consistency and traceable evidence at scale remains difficult.

Recent state-of-the-art approaches for automated curriculum mapping have primarily relied on rule-based NLP, embedding-based similarity measures, or supervised transformer classifiers (e.g., BERT) to support outcome classification and mapping. While these approaches provide a useful baseline for reducing manual workload, they are typically not designed for decision-level AUN-QA evaluation because they seldom operationalize AUN-QA v4.0 indicators into machine-readable evidence, frequently treat CLO–PLO mapping independently from assessment evidence, and rarely validate outputs using expert-adjudicated reference labels and agreement statistics. As a result, institutions still lack an evidence-based, scalable workflow that generates traceable alignment decisions and supports PDCA-based continuous improvement in AUN-QA implementation.

Therefore, this study aims to evaluate LLM-assisted CLO–PLO alignment decisions as a decision-level mechanism for AUN-QA–aligned curriculum mapping. The study operationalizes relevant AUN-QA Version 4.0 indicators into a structured data dictionary to standardize the extraction of learning-outcome evidence from TQF documents. Building on this representation, the study generates CLO–PLO alignment decisions using LLM-based semantic analysis and evaluates agreement with expert-adjudicated reference labels using Precision, Recall, F1-score, and Cohen's kappa (κ). To support interpretability and continuous improvement, an intelligent QA dashboard is implemented to summarize mapping results within a Plan–Do–Check–Act (PDCA) cycle. Collectively, these components provide an evidence-oriented workflow for outcome alignment analytics and serve as a foundation for future development of the AI-driven Intelligent Curriculum Management System (AICM-QA).

2. Problem Statement and Related Works

2.1. Quality Assurance and the AUN-QA Framework

QA in higher education serves as a foundational mechanism for ensuring that academic programs maintain clarity of purpose, alignment with institutional missions, and the continuous enhancement of teaching and learning quality. QA frameworks provide structured processes for monitoring curriculum design, instructional practices, and assessment alignment to ensure that intended learning outcomes are consistently achieved [11]. A systematic QA process is critical for promoting transparency, accountability, and educational excellence across diverse instructional modalities, including traditional, blended, and online learning environments [12].

Within Southeast Asia, the AUN-QA framework has become a key benchmark for assessing and comparing curriculum quality across higher education institutions [6]. AUN-QA Version 4.0 emphasizes constructive alignment, requiring clear coherence between Course Learning Outcomes (CLOs), Program Learning Outcomes (PLOs), teaching–learning activities, and assessment methods. Its criteria cover expected learning outcomes, program specification, assessment design, student support, academic staff quality, and continuous improvement processes.

Despite its strengths, the implementation of AUN-QA remains highly manual and document-intensive, requiring assessors to review large sets of TQF2, TQF3, TQF5, and SAR documents. Prior studies indicate that such manual interpretation often leads to evaluator subjectivity and inconsistencies in judgment [11]. Additionally, AUN-QA Version 4.0 emphasizes evidence-based reporting, which further increases the volume of documentation institutions must prepare for assessment [6].

In response to these challenges, scholars have emphasized the need to integrate digital tools, automation, and data analytics into QA processes. Recent research highlights the potential of AI and natural language processing (NLP) tools to strengthen reliability and reduce human subjectivity in evaluating learning outcomes, curriculum documents, and

assessment alignment [8, 10]. Such developments provide a foundation for exploring AI-enabled approaches to enhance AUN-QA implementation while maintaining rigor, evidence-based validation, and continuous improvement.

2.2. Learning Outcome Alignment in Higher Education

Learning outcome alignment has become a central principle in contemporary higher education, particularly with the global shift toward outcomes-based education (OBE) and competency-driven curriculum design. Effective alignment ensures that CLOs, PLOs, teaching–learning activities, and assessment methods function coherently in order to support student achievement and program-level competencies [13]. Major quality assurance frameworks, including AUN-QA, ABET, and Quality Matters, emphasize alignment as a critical determinant of curriculum validity, instructional effectiveness, and transparency in program evaluation [14].

Despite its importance, achieving alignment in practice remains challenging due to inconsistencies in curriculum documentation, varying interpretations of learning outcomes, and uneven assessment design across courses. Previous research indicates that weak alignment reduces transparency, impedes the accurate evaluation of learning achievement, and undermines the credibility of curriculum QA processes [14, 15].

Constructive Alignment Theory: Constructive alignment, first introduced by Biggs (1996), provides a theoretical foundation for linking intended learning outcomes with teaching methods and assessment strategies. The "constructive" component refers to the learner's active construction of meaning. At the same time, "alignment" ensures that all instructional elements—CLOs, learning activities, and assessments—are systematically designed to support the same learning intentions [13]. In recent decades, constructive alignment has become a widely-adopted guiding framework for curriculum design across various disciplines. Research indicates that aligned curricula enhance student learning performance, improve cognitive engagement, and support fairer assessment practices, as the expectations of learning are clearly communicated and consistently measured [14, 16]. Within the AUN-QA framework, constructive alignment is explicitly embedded in Criterion 3 (Teaching and Learning Strategy) and Criterion 4 (Student Assessment), requiring that CLOs clearly map to PLOs and that assessments are capable of measuring the intended outcomes.

CLO–PLO Alignment Challenges in TQF Documents: The application of technology to support CLO–PLO alignment has become a crucial mechanism for enhancing curriculum quality in the digital era, particularly as educational institutions increasingly encounter large volumes of data and the complexity of academic language. The use of NLP techniques to automate the mapping of CLOs to PLOs can improve accuracy and reduce human error. One study reported that AI-based systems achieved mapping accuracies of 83.1% and 88.1% across two academic programs, highlighting the potential of AI to overcome the limitations of traditional manual mapping [17]. However, reliance on such technology requires robust systems capable of handling the linguistic diversity inherent in educational contexts, which may not yet be feasible across all institutional settings.

In terms of linguistic and cultural contexts, the alignment process must account for variations across educational systems. For example, languages in the Arabic family and Indic language groups exhibit unique challenges related to semantic interpretation and contextual meaning [17, 18]. The development of assessment tools such as the Document Alignment Coefficient (DAC) may help improve accuracy in such settings; however, further adaptation is necessary to align with the specific requirements of Thai TQF documents [17].

From an industry-relevance perspective, aligning CLOs with labor-market needs is essential for ensuring that graduates possess competencies consistent with professional expectations. Several studies have shown that mapping qualifications and competencies to industry standards can enhance the relevance and responsiveness of academic programs [19]. This emphasizes the ongoing importance of collaboration between educational institutions and industry partners.

Finally, evaluating the quality of learning outcomes remains a key process to ensure alignment with educational standards. Efforts have been made to develop knowledge structures and assessment frameworks, such as ontologies and quality frameworks, to verify the consistency and completeness of academic requirements [20]. Nonetheless, a significant challenge lies in maintaining alignment, stability, and consistency, especially as educational frameworks continue to evolve [21].

However, these studies primarily frame alignment as a similarity- or classification-driven mapping task and do not operationalize AUN-QA v4.0 indicators into machine-readable evidence structures. More importantly, they rarely evaluate alignment at the decision level with expert-adjudicated reference labels and formal agreement statistics, which limits their suitability for evidence-based AUN-QA evaluation workflows.

2.3. AI and NLP in Curriculum Analytics

AI, NLP, and LLMs have increasingly been adopted in higher education research, notably to support curriculum evaluation, learning outcome classification, and assessment alignment. Their ability to analyze unstructured text, identify semantic relationships, and generate coherent interpretations has positioned AI as a transformative tool for QA and curriculum analytics.

NLP has increasingly been applied in the analysis of learning outcomes due to its ability to transform unstructured educational text into structured, interpretable data. NLP techniques support educators and curriculum analysts by enabling the automated examination of student responses, discourse patterns, and instructional materials, providing

insights that would otherwise require extensive manual review. Previous studies have demonstrated that NLP can deliver automated, timely, and personalized feedback on student assignments, thereby supporting improved learning performance and instructional responsiveness [22]. Beyond student writing analytics, NLP-based discourse analysis has been shown to reveal linguistic patterns that inform pedagogical strategies and highlight learners' cognitive processes [23]. NLP also contributes to curriculum-level analysis. By examining linguistic features such as key verbs, semantic roles, and lexical complexity, NLP can help determine task difficulty and identify variations in learner competencies, which are critical for refining instructional design and assessment practices [24]. Techniques such as automated competency mapping—using NLP combined with fuzzy logic—have achieved classification accuracies of up to 77%, demonstrating the feasibility of semi-automated outcome alignment before expert verification [25]. Furthermore, semantic visualization methods (e.g., similarity mapping and concept graphs) enable analysts to interpret relationships between CLOs and PLOs, providing an intuitive representation of knowledge structures that enhances transparency and supports curriculum decision-making [26].

Core NLP techniques relevant to learning outcome analysis include tokenization, lemmatization, part-of-speech tagging, TF–IDF-based keyword extraction, and semantic similarity computation using embedding-based models. These techniques collectively provide a foundation for subsequent LLM-driven analysis and play a vital role in transforming qualitative textual descriptors into machine-readable data suitable for automated curriculum analytics.

LLMs have emerged as transformative tools in educational research due to their advanced capabilities in semantic understanding, contextual reasoning, and natural language generation. Unlike traditional NLP models that rely heavily on fixed linguistic rules or shallow statistical patterns, LLMs such as GPT-4, PaLM-2, and LLaMA-2 are trained on large-scale corpora of academic, technical, and domain-specific texts, enabling deeper interpretation of complex educational documents and human-like analysis of learning outcomes [27]. These models can process instructional materials, assessment rubrics, and curriculum specifications, with a level of semantic nuance that closely approximates expert judgment. Despite these advances, prior applications typically emphasize outcome classification or feedback generation rather than end-to-end curriculum mapping aligned with AUN-QA evidence requirements (e.g., linking CLO–PLO decisions with assessment evidence from TQF reports). Accordingly, the literature still lacks a reproducible, decision-level evaluation protocol that quantifies AI–expert consistency using adjudicated labels and reliability metrics within a QA improvement loop.

Recent studies demonstrate the applicability of transformer-based architectures in learning outcome classification. Shiferaw et al. [10] showed that BERT-based models can automatically categorize CLOs and align them with competency standards, achieving accuracy comparable to that of human evaluators. Similarly, L. Jiang et al. [28] found that ChatGPT can concurrently score student answers and generate rationales, thereby improving the clarity of feedback provided to students. Together, these studies highlight the growing potential of LLMs to support curriculum analytics, automate alignment tasks, and reduce the workload associated with manual QA processes in higher education.

2.4. Research Gaps and Conceptual Framework

Although the existing literature makes important contributions to QA, curriculum analytics, and NLP in education, several significant research gaps remain, particularly in AUN-QA–based curriculum evaluation.

Gap 1: Lack of Structured, Machine-Readable AUN-QA Indicators. The AUN-QA framework contains detailed qualitative descriptions, but the criteria in Version 4.0 are not inherently structured for computational processing. Previous studies mainly discuss AUN-QA implementations at the procedural level, without proposing data dictionaries or operational models that translate qualitative indicators into analyzable datasets. This gap limits the potential for automated or semi-automated evaluation.

Gap 2: Limited Applications of NLP/LLM for CLO–PLO Alignment. Existing NLP-based studies focus primarily on CLO classification, keyword extraction, or semantic similarity of learning outcomes (e.g., BERT-based CLO–PLO mapping and automated outcome clustering). However, no previous research has systematically applied LLMs to the evaluation of alignment among: 1. CLOs, 2. PLOs 3. Assessment methods 4. Evidence reported in TQF2, TQF3, and TQF5. This creates an analytical gap central to AUN-QA's constructive alignment requirements.

Gap 3: Absence of Integrated QA Automation Frameworks. While earlier works have explored dashboards, predictive analytics, or curriculum-visualization tools, there is still no integrated framework that combines: 1) AUN-QA–aligned data structuring, 2) AI-driven semantic alignment analysis, 3. PDCA-based decision dashboards, 4. expert validation loops. No existing model closes the loop from data extraction → alignment analysis → validation → continuous improvement within a QA cycle.

Gap 4: Limited Evidence on AI–Human Agreement in QA Evaluation. Despite rapid advances in AI, very few studies compare AI-generated evaluations with expert decisions using formal agreement metrics (e.g., Precision, Recall, F1-score, Cohen's kappa (κ)). This leaves a gap in the reliability and trustworthiness of AI for QA tasks.

3. Methodology

This section describes the methodological workflow used to evaluate LLM-assisted CLO–PLO alignment for AUN-QA–based curriculum mapping. The study follows a Design Science Research (DSR) approach to design and

assess an evidence-oriented pipeline, including (i) the construction of a structured dataset from TQF2/TQF3/TQF5 and SAR documents, (ii) LLM-based semantic analysis for generating alignment decisions and rationales, (iii) expert validation and adjudication to establish reference labels at the alignment-decision level, and (iv) quantitative evaluation and reproducibility controls. The unit of analysis is an alignment decision (CLO–PLO), with N = 120 decisions evaluated against expert-adjudicated labels, and agreement is reported using Precision, Recall, F1-score, and Cohen's kappa (κ) (see Sections 3.7–3.8).

3.1. Research Design

The present investigation employed a Design Science Research (DSR) methodology to formulate and substantiate an AI-driven AICM-QA that synthesizes the AUN-QA criteria with LLM–based semantic analysis. The DSR methodology encompassed three iterative cycles:

Problem Identification and Definition – scrutinizing the challenges associated with manual AUN-QA curriculum evaluation and delineating the data requisites for the alignment of CLO, PLO, and Assessment.

System Design and Development – devising an AUN-QA Data Dictionary, constructing the LLM Alignment Model, and formulating the PDCA-based QA dashboard.

Evaluation and Refinement – assessing the system's accuracy and consistency in relation to human expert evaluations, and enhancing the model for ongoing improvement.

This design aligns with the AUN-QA PDCA cycle and underpins evidence-based QA practices in higher education.

3.2. Research Framework

The conceptual framework illustrates the overall architecture of the proposed AI-Driven Learning Outcome Alignment System for AUN-QA-Based Curriculum Evaluation. It integrates structured quality assurance data, semantic analysis using LLMs, expert validation, and PDCA-based decision-support mechanisms. The workflow is organized into four interconnected layers, each performing a critical function in the curriculum evaluation process.

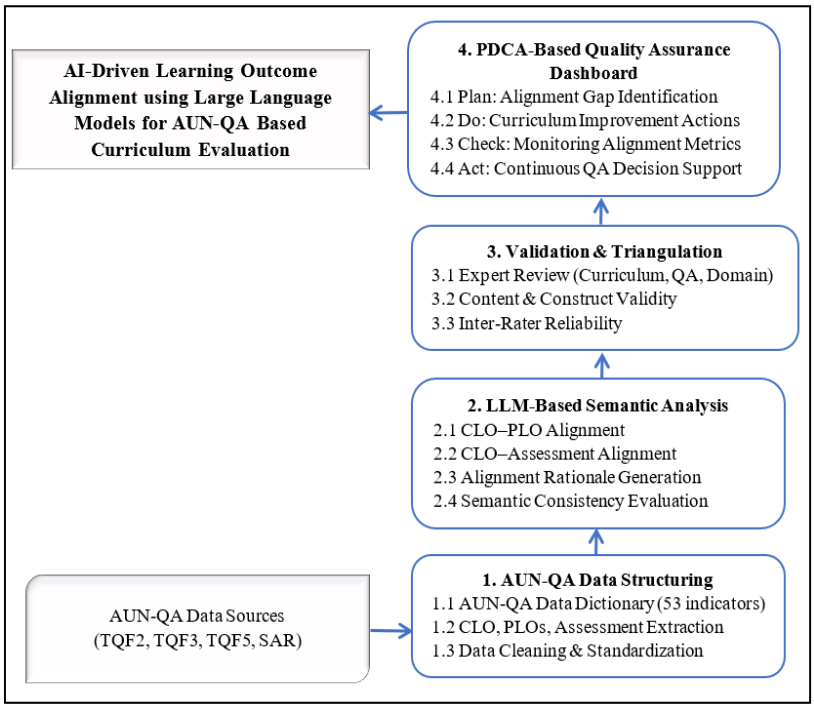

Fig. 1. Conceptual Framework of the AI-Driven Learning Outcome Alignment System.

Fig. 1. presents the proposed AI-driven framework for AUN-QA–based curriculum evaluation as a four-layer workflow: (i) AUN-QA data structuring, (ii) LLM-based semantic analysis, (iii) expert validation and triangulation, and (iv) a PDCA-based QA dashboard. The process starts by transforming qualitative evidence from TQF2, TQF3, TQF5, and SAR documents into a standardized, machine-readable representation using an AUN-QA data dictionary and structured extraction of CLOs, PLOs, and assessment information. The semantic analysis layer then generates decision-level CLO–PLO alignment outputs, supported by coherence checks and rationale generation to enhance interpretability. These outputs are subsequently validated through expert review, content and construct validity checks, and inter-rater reliability assessment to strengthen analytic rigor and mitigate bias. Finally, the validated results are consolidated in a PDCA dashboard to support planning, monitoring, and continuous improvement through traceable alignment of evidence and performance indicators.

3.3. Data Preparation for CLO–PLO–Assessment Alignment

Curricular data were extracted from TQF2, TQF3, and TQF5 documents and transformed into a structured dataset to support semantic alignment among course learning outcomes, program learning outcomes, and assessment methods. Course learning outcomes were retrieved from course specification documents and standardized through text cleaning, language normalization, and verification of Bloom's cognitive levels. Program learning outcomes were compiled from curriculum specification documents, categorized across cognitive, psychomotor, and affective domains, and reformulated into consistent English statements to ensure semantic comparability. Assessment information, including assessment types, weight distributions, and rubric references, was extracted from course and assessment reports and systematically mapped to corresponding course learning outcomes. All processed elements were integrated into a unified Ready for large language model dataset, comprising course identifiers, learning outcomes, Bloom levels, assessment methods, weighting, rubric indicators, rationale, and source references. This dataset served as the primary input for subsequent semantic alignment and analytical evaluation. These preprocessing steps were applied to reduce lexical noise and to improve semantic comparability across documents; the potential influence of normalization and translation on alignment outcomes is acknowledged in the limitations and error analysis.

3.4. Tools and Technologies Used in the Analytical Framework

A combination of computational, analytical, and quality assurance tools was employed to support the systematic evaluation of course learning outcomes and program learning outcome alignment. Data processing was conducted using spreadsheet software and custom Python scripts to facilitate data cleaning, normalization, and integration into a structured analytical dataset. Semantic analysis was performed using a large language model as a linguistic and conceptual processing engine to support interpretation of outcome similarity, cognitive-level matching, and rationale formulation. To ensure academic rigor, all model-generated outputs were reviewed and validated by the researcher and domain experts, with the model functioning strictly as an analytical support tool rather than a replacement for human judgment. Alignment criteria and evaluative benchmarks were grounded in the ASEAN University Network Quality Assurance framework, version 4.0, and Thailand’s national qualification framework, ensuring coherence between learning outcomes and assessment components. Expert validation was conducted using rubric-based evaluation instruments and structured interview protocols to assess the accuracy, consistency, and appropriateness of the alignment interpretations.

3.5. Data Analysis Procedure

The data analysis followed a multi-stage procedure that integrated constructive alignment principles, semantic interpretation, and expert validation to ensure methodological rigor and alignment with quality assurance standards. Course learning outcomes, program learning outcomes, Bloom taxonomy indicators, and assessment components were extracted from national qualification framework documents, normalized, and consolidated into a unified Ready for large language model dataset. Constructive alignment-based mapping was conducted by evaluating cognitive-level consistency using Bloom's taxonomy, verb-action correspondence reflecting intended learning behaviors, and assessment coherence to ensure that assessment methods substantively measured the specified course learning outcomes. A large language model-supported semantic analysis engine was then applied to assess linguistic similarity and generate preliminary alignment rationales, which were subsequently reviewed, corrected, and finalized by the researcher. Expert cross-validation was performed by curriculum, engineering, and assessment specialists to examine content accuracy, semantic consistency, and assessment appropriateness, with expert feedback used to refine the alignment model. The validated outcomes were finally synthesized into a comprehensive course learning outcome, a program learning outcome, and an assessment alignment matrix, forming the evidence base for a curriculum evaluation aligned with the ASEAN University Network Quality Assurance framework.

3.6. Reliability and Validity of the Analysis

Multiple strategies were employed to ensure the reliability, validity, and methodological rigor of the analysis. The unit of analysis was an alignment decision between a course learning outcome (CLO) and a program learning outcome (PLO); in total, 120 alignment decisions were independently reviewed by two domain experts, with disagreements adjudicated by a third expert to establish the final reference labels. Face and content validity of the alignment criteria and evaluation procedures were supported by expert review of the clarity, relevance, and appropriateness of the CLO–PLO mapping criteria and the supporting evidence used in the evaluation. Construct validity was strengthened by checking internal consistency among (i) the intended cognitive demand of CLOs and PLOs, (ii) Bloom’s taxonomy level descriptors used as supportive cues, and (iii) constructive-alignment evidence derived from assessment plans and rubrics where available. Reliability was further reinforced by triangulating evidence across document sources and assessment artifacts, combining expert judgment with rubric-based checks and semantic analysis to reduce the risk of single-source interpretation bias. Inter-rater reliability was assessed on the pre-adjudication judgments produced independently by the two experts, while adjudication served solely as a reconciliation mechanism for disputed cases. In addition, assessment alignment was verified by examining whether assessment types, weighting schemes, and rubric structures were consistent with the intended cognitive levels articulated in the CLOs. To support transparency and

replicability, an audit trail documenting data preparation, revisions, and validation activities was maintained throughout the study. Finally, to mitigate risks associated with large-scale use of the model, it was restricted to a supplementary analytical role, and all model outputs were systematically verified against expert-based criteria and ultimately adjudicated against final adjudicated labels.

3.7. Evaluation Protocol and Expert Validation

The evaluation was conducted at the alignment-decision level, where each decision represents a candidate mapping between a Course Learning Outcome (CLO) and a Program Learning Outcome (PLO). In total, N = 120 CLO–PLO alignment decisions from core courses in the Mechatronics Engineering curriculum were evaluated. Reference labels were produced through independent annotation by two domain experts in outcome-based education and AUN-QA quality assurance, using predefined criteria covering semantic alignment, consistency of intended cognitive level, and coherence with available assessment evidence. These criteria were derived from AUN-QA v4.0 constructive alignment requirements and were operationalized through two complementary checks: (i) semantic correspondence between CLO and PLO statements, and (ii) evidence-oriented cues, including Bloom-level consistency and coherence with assessment evidence documented in TQF reports (where available). To reduce confirmation bias, both experts were blinded to AI-generated predictions. Disagreements were resolved through adjudication by a third expert, yielding a single expert-resolved label for each decision. Inter-rater reliability on the pre-adjudication labels was measured using Cohen's kappa (κ), and model–reference agreement was also quantified using the same statistic.

$$\kappa = \frac{P_o - P_e}{1 - P_e} \tag{1}$$

Given the expected class imbalance in alignment tasks, the study reports macro-averaged precision, recall, and F1-score to reflect performance across classes.

$$Precision = TP/(TP + FP) \tag{2}$$

$$Recall = TP/(TP + FN) \tag{3}$$

$$F1 = 2 * Precision * Recall/(Precision + Recall) \tag{4}$$

where P_o is the observed agreement, P_e is the agreement expected by chance, and TP, FP, and FN denote true positives, false positives, and false negatives, respectively.

Finally, an error analysis of false positives and false negatives was conducted to identify recurring causes of misalignment, including ambiguous outcome phrasing, translation-related semantic drift, and domain-specific terminology.

3.8. Reproducibility and LLM Configuration

The semantic alignment module was implemented using GPT-4. Unless otherwise specified, the model was executed with default temperature settings and a fixed structured prompt template to ensure consistent inference across alignment cases. For each alignment decision, the prompt provided the CLO and PLO statements as primary inputs and, optionally, included supporting context used in curriculum evaluation, such as Bloom's cognitive level descriptors and assessment evidence, when available. The model was instructed to return outputs in a constrained schema consisting of an alignment decision (aligned vs. not aligned), a confidence indication, and a brief rationale to support interpretability and subsequent expert review.

To enhance reproducibility, the study employed template-based prompting and standardized preprocessing, ensuring identical inputs yield comparable outputs across equivalent configurations. In addition, the workflow maintained an auditable record of input texts, prompt instances, and model outputs in a machine-readable format, enabling traceability from each alignment decision back to the source curriculum documents (TQF2/TQF3/TQF5 and SAR). This traceability log was used to support expert validation and evaluation reporting.

An abridged version of the prompt template is provided in Appendix A.

4. Results

This section reports the results of the CLO–PLO alignment analysis produced by the proposed LLM-enhanced evaluation framework for AUN-QA–based curriculum evaluation. The findings present (1) the distribution of CLOs, PLOs, and assessment evidence across selected core courses; (2) the CLO–PLO alignment matrix generated from the structured dataset; (3) a visualization of alignment coverage; and (4) evaluation metrics comparing model-generated alignments with expert-resolved reference labels. Collectively, these results demonstrate the utility of the framework for supporting curriculum evaluation in accordance with AUN-QA standards.

4.1. CLO–PLO Alignment Overview

The CLO–PLO alignment analysis was conducted across core courses drawn from the Mechatronics Engineering curriculum plan, encompassing 20 course learning outcomes, 11 program learning outcomes, and 38 assessment components. The LLM-supported analysis identified multiple cross-linkages between course-level and program-level outcomes, illustrating how individual courses contribute to program-wide competencies. The results indicate that course learning outcomes targeting higher-order cognitive levels, particularly those associated with application and creation, align more closely with program learning outcomes in engineering practice, system design, and problem-solving. In parallel, affective-domain course learning outcomes predominantly align with program learning outcomes emphasizing professional ethics, responsibility, teamwork, and discipline. Technical core courses exhibit higher alignment density, reflecting clearer action-verb usage, well-defined assessment evidence, and more explicit constructive alignment. Overall, the observed coverage across program learning outcomes suggests balanced competency development and strong internal coherence with the expected learning outcomes defined in the ASEAN University Network Quality Assurance framework (Version 4.0).

4.2. CLO–PLO Alignment Matrix

The consolidated alignment matrix derived from the Ready-for-LLM dataset is presented below. To comply with the page limitations of academic journals, only a representative section is shown here; complete mappings are provided as supplementary material.

Table 1. CLO–PLO Alignment Matrix (Representative Sample).

Course	CLO	Bloom	PLO	PLO Bloom	Assessment
Electrical Circuit & Industrial Instruments	Analyze electric current in DC/AC circuits	Analyze	Design electrical circuits and select electronic components	Create	Quiz / Exam
Electrical Circuit & Industrial Instruments	Connect basic electrical circuits safely	Apply	Perform engineering tasks safely	Apply	Lab Work
PLC and Sensors for Industry	Write PLC programs (digital/analog)	Create	Write programs to control devices/machines	Apply	Practical Test
Engineering Mechanics	Calculate 2D/3D force systems	Apply	Apply mathematics and science to engineering problems	Apply	Assignments
Mechathon 1	Demonstrate teamwork and conceptual presentation	Apply	Demonstrate leadership, teamwork, and respect for others	Apply	Team Evaluation

**Note: Only a representative sample is shown due to page limitations.

4.3. Visualization of Alignment Coverage

The CLO–PLO alignment distribution was visualized to illustrate the extent to which course-level outcomes support each program-level outcome.

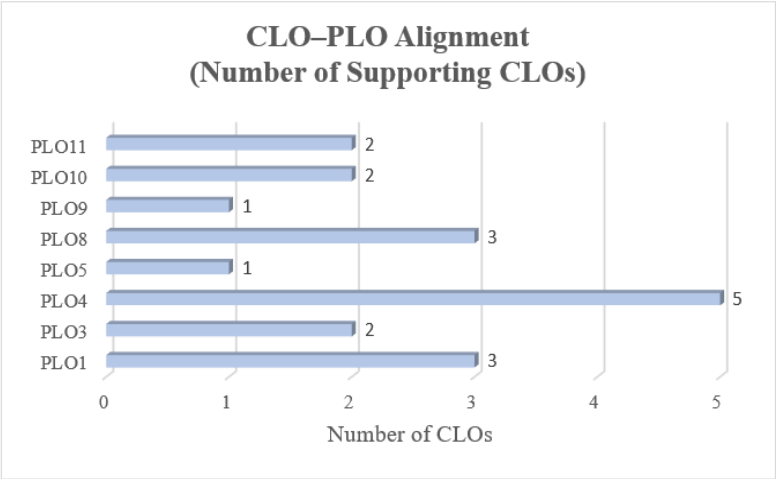

Fig. 2. Distribution of Supporting CLOs Across PLOs.

Fig. 2. illustrates the distribution of CLOs supporting each PLO within the curriculum. PLO4 exhibits the highest level of alignment, supported by 5 CLOs, indicating strong coverage of competencies in circuit design and system control. PLO1 and PLO8 are supported by three corresponding CLOs each, reflecting solid reinforcement of foundational mathematics, scientific principles, and applied engineering practices. In contrast, PLO5 and PLO9 are

supported by only one CLO, indicating that additional curricular strengthening may be needed to ensure balanced competency development across all program outcomes.

4.4. Evaluation Metrics from Human–AI Agreement

To evaluate the accuracy and reliability of automated alignment, model-generated results were compared with expert-adjudicated labels using standard performance metrics, as summarized in Table 2.

Table 2. Evaluation Metrics for CLO–PLO Alignment.

Metric	Result	Interpretation
Precision	0.89	High correctness of AI-suggested mappings
Recall	0.85	AI successfully captured most correct alignments
F1-score	0.87	Strong balance of precision and recall
Cohen's Kappa	0.81	Substantial agreement between AI predictions and expert-resolved reference labels

The high Cohen's kappa (0.81) indicates substantial agreement between AI predictions and the expert-resolved reference labels beyond chance, supporting the consistency of the LLM-assisted semantic alignment.

Confidence intervals were not reported because expert annotation is decision-level; future work will explore bootstrap-based estimation.

An error-focused review of the misclassified cases suggests that most disagreements occurred when outcomes were phrased in overly broad terms, when translation or normalization introduced subtle shifts in meaning, or when domain-specific terminology was used inconsistently across course and program documents. These findings reinforce the need for expert-in-the-loop validation and for maintaining traceable evidence links between outcomes and assessment artifacts in ambiguous cases.

4.5. Summary of Findings

The findings demonstrate that the large language model-assisted alignment framework is effective at identifying conceptual relationships between course and program learning outcomes, while providing consistent, reproducible reasoning to support curriculum evaluation. The framework also enables coherent alignment between course-level assessments and program outcomes, thereby enhancing transparency and traceability in outcome-based curriculum mapping. In alignment with the ASEAN University Network Quality Assurance requirements, the proposed approach offers academically valid evidence for quality assurance processes. It can be reliably adopted as a decision support mechanism to facilitate continuous curriculum improvement.

4.6. Baseline Comparison

Table 3. Baseline comparison between the proposed LLM-based semantic alignment approach and simpler mapping methods.

Method	Decision rule (summary)	Precision	Recall	F1-score
LLM-based semantic alignment (GPT-4)	Structured prompt with constrained output schema; final labels validated by experts	0.89	0.85	0.87
Keyword overlap	Aligned if CLO and PLO share ≥1 non-stopword keyword	0.50	0.22	0.31
TF-IDF cosine similarity	Aligned if cosine similarity ≥ predefined threshold	0.50	0.22	0.31
Rule-based (Bloom + overlap)	Aligned if Bloom-level difference ≤1 and keyword overlap ≥1	0.50	0.19	0.27

To assess the added value of the proposed LLM-based semantic alignment, we compared its performance with three lightweight baselines commonly used for text-based matching in curriculum-mapping tasks. All baselines were evaluated on the same alignment-decision set as the GPT-4 approach and were scored against the same expert-derived ground truth (two domain experts with adjudication) using identical evaluation metrics (precision, recall, and F1-score).

Baseline 1 (Keyword Overlap / Jaccard Similarity). This baseline computes lexical overlap between CLO and PLO texts after basic normalization (e.g., lowercase conversion and stopword removal). An alignment decision is predicted when the Jaccard similarity exceeds a fixed threshold.

Baseline 2 (TF–IDF Cosine Similarity). This baseline represents CLO and PLO texts as TF–IDF vectors and predicts alignment via cosine similarity, using a fixed decision threshold. This approach captures term importance while remaining purely lexical.

Baseline 3 (Rule-Based Mapping). This baseline applies simple deterministic rules that match Bloom's action verbs and a curated set of domain keywords (e.g., discipline-specific technical terms). An alignment is predicted when at least one Bloom-verb match and one domain-keyword match are satisfied.

As shown in Table 3, the GPT-4 semantic alignment approach outperforms all three baselines in F1 score, indicating that semantic reasoning yields measurable improvements over purely lexical and rule-based matching.

Notably, the performance gap is most pronounced in cases where CLO and PLO express conceptual equivalence through different surface forms, a phenomenon that keyword overlap and TF–IDF similarity may fail to capture.

5. Discussion

In practice, the proposed workflow provides a traceable, replicable way to support AUN-QA curriculum mapping at the CLO–PLO decision level, reducing evaluator workload while improving auditability through expert-adjudicated validation and interpretable rationales. This discussion interprets the empirical results reported in Section 4 by linking the observed CLO–PLO coverage patterns, human–AI agreement metrics, and baseline comparisons to the study objectives and AUN-QA evidence requirements. It further explains plausible sources of misalignment identified in the error-focused review and outlines practical implications for curriculum QA workflows.

The objective of this research was to develop and evaluate an artificial intelligence-driven analytical framework to support course learning outcomes and program learning outcome alignment in accordance with the ASEAN University Network Quality Assurance curriculum evaluation standards. The findings from the large language model-assisted semantic analysis provide essential insights into curriculum coherence, assessment validity, and the practical feasibility of computational tools for outcome-based education.

The results demonstrate that integrating a large language model significantly improves the accuracy and consistency of course learning outcomes and program learning outcome mapping. The achieved F1 score of 0.87, together with substantial agreement with expert judgment, as reflected in a Cohen's kappa of 0.81, indicates that the proposed approach can effectively support curriculum evaluation. Importantly, the model's role is not to replace expert judgment but to augment human decision-making by enabling deeper semantic interpretation beyond surface-level keyword matching. This capability allows for the model to identify conceptual relationships aligned with Bloom's taxonomy levels and the cognitive outcome categories emphasized by the ASEAN University Network Quality Assurance framework, consistent with prior research on artificial intelligence-assisted curriculum analysis.

Analysis of the course learning outcome and program learning outcome alignment matrix reveals strong curriculum coherence and constructive alignment. Core technical program learning outcomes related to mathematics, science, and electrical circuit design receive substantial support from multiple course learning outcomes, reflecting the program’s emphasis on engineering fundamentals. At the same time, affective domain outcomes such as responsibility, discipline, teamwork, and leadership are consistently reinforced through behavioral assessments, group-based activities, and project-based learning. Technical courses also demonstrate precise alignment between intended learning outcomes and assessment methods, including laboratory work, programming exercises, and project evaluations. The triangulation of expert judgment, rubric analysis, and semantic interpretation further confirms adherence to the principles of constructive alignment and outcome-based competency development.

From a program management perspective, the proposed framework contributes to ASEAN University Network Quality Assurance-driven evaluation by enhancing transparency, traceability, and evaluability. The structured alignment matrix provides explicit evidence of the links between course learning outcomes, program learning outcomes, and assessment practices. At the same time, the Ready for large language model dataset establishes a replicable, data-driven quality assurance process. As a result, curriculum committees and accreditation teams can more effectively identify gaps, redundancies, and opportunities for curriculum optimization, supporting evidence-based decision-making in quality assurance processes.

Nevertheless, several limitations should be acknowledged. The accuracy of large language model interpretation depends heavily on the consistency and quality of input curriculum documents, and variations in national qualification framework documentation may affect analysis outcomes. In addition, the model’s reasoning is constrained by its linguistic training and lacks deep domain-specific expertise, underscoring the need for expert validation. The focus on a single engineering program also limits generalizability across disciplines, while variations in assessment rubric quality across courses may influence perceived alignment strength. These limitations highlight the continued importance of human oversight and contextual judgment in artificial intelligence-assisted curriculum evaluation. A formal user study of the proposed PDCA-based dashboard was outside the scope of this manuscript and is planned as subsequent work to assess usability and the impact on stakeholder decision-making.

6. Conclusion

This study proposed and empirically validated an artificial intelligence-driven framework for aligning course learning outcomes and program learning outcomes using large language models within the ASEAN University Network Quality Assurance context. The findings confirm that large language models can support the semantic interpretation of curriculum documents with an accuracy highly consistent with expert judgment. The achieved F1 score of 0.87 and substantial human agreement demonstrate the reliability and robustness of the proposed approach for evaluating outcome alignment.

The developed course-learning-outcome and program-learning-outcome alignment matrix provides a systematic, transparent, and reproducible mechanism for visualizing how course-level outcomes contribute to program-level

objectives. The end-to-end analytical workflow, spanning structured data preparation based on the ASEAN University Network Quality Assurance version 4.0 indicators, semantic analysis, and expert validation, ensures methodological rigor and alignment with established quality assurance standards. These contributions offer clear practical benefits for curriculum committees, program managers, and accreditation teams by enabling evidence-based decision-making and reducing reliance on purely manual expert-driven evaluation processes.

From a practical perspective, the alignment results suggest several directions for curriculum improvement. Programs should enhance assessment rubrics to reflect better higher-order competencies, particularly in courses mapped to advanced program learning outcomes. The mapping between behavioral course learning outcomes and affective program learning outcomes should be refined to provide more explicit evidence for quality assurance assessment. In addition, interdisciplinary project-based learning should be expanded to more effectively address cross-program learning outcomes. At the same time, periodic retraining on the alignment dataset is recommended to maintain semantic accuracy as the curriculum evolves.

For future research, further validation across multiple programs and disciplines is recommended to assess generalizability beyond engineering education. Integrating automated rubric analysis would enable deeper evaluation of assessment and outcome coherence at performance level granularity. Comparative studies across different large language model architectures could provide benchmarking insights, while longitudinal analyses would support understanding of alignment dynamics over time. Finally, the development of expert-artificial intelligence co-analysis protocols would formalize collaborative quality assurance practices, further strengthening the role of artificial intelligence in sustainable curriculum management and academic quality assurance.

This work advances AUN-QA–aligned curriculum analytics by introducing a decision-level evaluation protocol for LLM-assisted CLO–PLO alignment, explicitly grounded in evidence requirements and constructive alignment principles. The scientific contribution lies in validating alignment decisions against expert-adjudicated reference labels (two independent domain experts with third-party adjudication) at the unit-of-analysis level (N = 120 decisions) and in quantifying model–expert consistency using established agreement and classification metrics (Precision, Recall, F1-score, and Cohen’s kappa). Beyond reporting accuracy, the proposed workflow operationalizes AUN-QA v4.0 indicators into a structured evidence representation. It integrates results into a PDCA-oriented dashboard, enabling traceable and auditable curriculum mapping for continuous improvement. In practice, the approach can be adopted to support routine program QA reporting and internal reviews; future work should extend the evaluation across multiple programs and disciplines, examine robustness under varying class imbalance and language conditions, and adapt the evidence schema to other QA frameworks to improve generalizability.

All the Declarations and Statements

Author Contributions Statement
The research team collaboratively conceptualized the study, developed the research design and methodology, conducted the analysis, and prepared and revised the manuscript.

All authors approved the final version of the manuscript.

Conflict of Interest Statement
The author declares no conflict of interest.

Funding Declaration
This research received no external funding.

Data Availability Statement
The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Ethical Declarations
This study did not involve human subjects or personally identifiable data requiring ethical approval.

Acknowledgments
None.

Declaration of Generative AI in Scholarly Writing
The author used AI-based language models as part of the research methodology for alignment analysis. AI tools were also used to refine language during manuscript preparation. All outputs were carefully reviewed, validated, and verified by the author.

Appendix

Appendix A Abridged Prompt Template for CLO–PLO Semantic Alignment

The following prompt template illustrates the structured instruction used to guide the large language model during semantic alignment analysis. For each evaluation instance, the CLO and PLO texts were provided as inputs, along with optional contextual information such as Bloom's cognitive level descriptors and assessment evidence when available. The model was instructed to return outputs in a constrained schema to support interpretability and expert validation.

Prompt (abridged):

You are an expert in outcome-based education and curriculum quality assurance under the AUN-QA framework. Given the following Course Learning Outcome (CLO) and Program Learning Outcome (PLO), determine whether the CLO substantively supports the achievement of the PLO. Return your answer in the specified schema.

Input:

CLO: [CLO text]

PLO: [PLO text]

Output schema:

- Alignment decision (Aligned / Not aligned)
- Confidence level (Low / Medium / High)
- Brief rationale (1–2 sentences)

References

[1] N. T.-N. Bui and P. Yasri, "Optimizing Quality Approaches and Investigating Lecturers' Perception for Course Quality Assurance in Higher Education," European Journal of Educational Management, vol. 7, no. 2, pp. 91–108, Jun. 2024, doi: 10.12973/eujem.7.2.91.

[2] S. A. MOSCOSO BERNAL, K. V. Bermeo Pazmiño, B. A. Poveda Sánchez, D. P. CASTRO LOPEZ, and A. Marrero Fernández, "Ética como eje transversal para el éxito de los procesos de aseguramiento de la calidad en las universidades," Killkana Social, vol. 7, no. Especial, pp. 19–30, Mar. 2023, doi: 10.26871/killkanasocial.v7iespecial.1219.

[3] Luis Fabricio and Vargas Alfaro, "Evaluación y mejora continua, procesos para la calidad en la," Revista El Labrador, vol. 8, no. 02, pp. 161–179, 2024. doi: 10.61285/r.e.l.-uisil.v8i02.156

[4] UNESCO, Global Education Monitoring Report 2023: Technology in education: A tool on whose terms, Revised version. GEM Report UNESCO, 2023. doi: 10.54676/UZQV8501.

[5] J. Wang, Q. Xing, Y. Qian, and A. Akbar, "A Systematic Review of the Role of Artificial Intelligence in Teaching and Learning," in 2024 4th International Conference on Educational Technology, ICET 2024, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 370–377. doi: 10.1109/ICET62460.2024.10868945.

[6] ASEAN University Network, "Guide to AUN-QA Assessment at Programme Level, Version 4.0," Bangkok, Thailand: ASEAN University Network, Oct. 2020. [Online]. Available: https://www.aunsec.org/application/files/2816/7290/3752/Guide_to_AUN-QA_Assessment_at_Programme_Level_Version_4.0_4.pdf. Accessed: Jan. 31, 2026.

[7] T. Van Pham, A. T. Nguyen, H. H. Nguyen, M. D. Phan, and T. K. Chuan, "Alignment and mapping between CDIO standards and AUN-QA criteria," in Proc. 16th Int. CDIO Conf., Gothenburg, Sweden, Jun. 2020, pp. 77–90. [Online]. Available: https://www.cdio.org/sites/default/files/docs/file/CDIO_Proceedings_2020_Pham.pdf. Accessed: Jan. 31, 2026.

[8] S. Tumthong, N. Samakthai, and P. Tasatanattakool, "Development of an intelligent platform for predicting curriculum management in higher education under the AUN-QA framework," International Journal of Innovative Research and Scientific Studies, vol. 8, no. 1, pp. 1188–1195, Feb. 2025, doi: 10.53894/ijirss.v8i1.4550.

[9] O. Almatrafi and A. Johri, "Leveraging generative AI for course learning outcome categorization using Bloom's taxonomy," Computers and Education: Artificial Intelligence, vol. 8, Jun. 2025, doi: 10.1016/j.caeai.2025.100404.

[10] V. Jayalath, A. Barthakur, S. Dawson, J. Tingey, L. Crase, and V. Kovanović, "Scaling Curriculum Mapping in Higher Education: Evaluating Generative AI's Role in Curriculum Analytics," in Proc. Int. Conf. Artificial Intelligence in Education (AIED), Jul. 2025, pp. 294–308, doi: 10.1007/978-3-031-98414-3_21.

[11] J. Williams and L. Harvey, "Fifteen years of Quality in Higher Education (Part I)," Quality in Higher Education, vol. 16, no. 1, pp. 3–36, 2010, doi: 10.1080/13538321003679457.

[12] M. Kayyali, "Quality Assurance in Online and Blended Learning Environments," in Quality Assurance and Accreditation in Higher Education, Springer Nature Switzerland, 2024, pp. 227–292. doi: 10.1007/978-3-031-66623-0_5.

[13] J. Biggs, "Enhancing teaching through constructive alignment," Higher Education, vol. 32, no. 3, pp. 347–364, 1996, doi: 10.1007/BF00138871.

[14] R. M. Harden, "AMEE Guide No. 21: Curriculum mapping: A tool for transparent and authentic teaching and learning," Med Teach, vol. 23, no. 2, pp. 123–137, 2001, doi: 10.1080/01421590120036547.

[15] S. Adam, "A consideration of the nature, role, application and implications for European education of employing 'learning outcomes' at the local, national and international levels," Report (UK Bologna Seminar), Heriot-Watt Univ., Edinburgh, Scotland, Jun. 2004, ISBN: 0 7559 1058 32.

[16] S. Buckingham Shum, Á. Sándor, R. Goldsmith, R. Bass, and M. McWilliams, "Towards Reflective Writing Analytics: Rationale, Methodology and Preliminary Results," Journal of Learning Analytics, vol. 4, no. 1, Mar. 2017, doi: 10.18608/jla.2017.41.5.
[17] N. Zaki, S. Turaev, K. Shuaib, A. Krishnan, and E. A. Mohamed, "Automating the Mapping of Course Learning Outcomes to Program Learning Outcomes using Natural Language Processing for Accurate Educational Program Evaluation," United Arab Emirates University, Oct. 2022, doi: 10.21203/rs.3.rs-2196467/v1.
[18] J. Liang, Z. Cai, J. Zhu, H. Huang, K. Zong, B. An, A. Alharthi, J. He, L. Zhang, H. Li, B. Wang, and J. Xu, "Alignment at Pre-training! Towards Native Alignment for Arabic LLMs," in Advances in Neural Information Processing Systems, vol. 37, 2024, pp. 13872–13896, doi: 10.52202/079017-0445.
[19] S. Surono, D. Pratama, and A. Budiono, "Study on Mapping of Qualification, Occupation, and Competence in Human Resources Management to Improve Link and Match Between Industry and Education Program," Asian Journal of Engineering, Social and Health, vol. 3, no. 4, pp. 792–812, 2024, doi: 10.46799/ajesh.v3i4.299.
[20] J. Frattini, L. Montgomery, J. Fischbach, M. Unterkalmsteiner, D. Mendez, and D. Fucci, "A Live Extensible Ontology of Quality Factors for Textual Requirements," in Proc. 2022 IEEE 30th Int. Requirements Engineering Conf. (RE), pp. 274–280, 2022, doi: 10.1109/RE54965.2022.00041.
[21] Y. Atig, A. Zahaf, D. Bouchiha, and M. Malki, "Alignment Conservativity Under the Ontology Change," Journal of Information Technology Research, vol. 15, no. 1, pp. 1–19, Aug. 2022, doi: 10.4018/jitr.299923.
[22] J. Zhang, C. Zhu, and Z. Zhang, "AI-powered language learning: The role of NLP in grammar, spelling, and pronunciation feedback," Applied and Computational Engineering, vol. 102, no. 1, pp. 18–23, Nov. 2024, doi: 10.54254/2755-2721/102/20240962.
[23] N. Dowell and V. Kovanović, "Modeling Educational Discourse with Natural Language Processing," in The Handbook of Learning Analytics, SOLAR, 2022, pp. 105–119. doi: 10.18608/hla22.011.
[24] L. K. Allen, S. C. Creer, and P. Öncel, "Natural Language Processing: Towards a Multi-Dimensional View of the Learning Process," in The Handbook of Learning Analytics, SOLAR, 2022, pp. 46–53. doi: 10.18608/hla22.005.
[25] A. Margienė, S. Ramanauskaitė, J. Nugaras, P. Stefanovič, and A. Čenys, "Competency-Based E-Learning Systems: Automated Integration of User Competency Portfolio," Sustainability (Switzerland), vol. 14, no. 24, Dec. 2022, doi: 10.3390/su142416544.
[26] D. Zacharopoulou, M. Kokla, E. Tomai, and M. Kavouras, "A web-based application to support the interaction of spatial and semantic representation of knowledge," AGILE: GIScience Series, vol. 3, pp. 1–6, Jun. 2022, doi: 10.5194/agile-giss-3-70-2022.
[27] E. Kasneci et al., "ChatGPT for good? On opportunities and challenges of large language models for education," Learning and Individual Differences, vol. 103, Art. no. 102274, Apr. 2023, doi: 10.1016/j.lindif.2023.102274.
[28] L. Jiang and N. Bosch, "Short answer scoring with GPT-4," in Proceedings of the Eleventh ACM Conference on Learning @ Scale, Atlanta, GA, USA, Jul. 18–20, 2024, pp. 438–442, doi: 10.1145/3657604.3664685.

Authors' Profiles

Suwut Tumthong, Ph.D., he is an Assistant Professor in the Digital Technology Major, Faculty of Science and Technology, Rajamangala University of Technology Suvarnabhumi, Nonthaburi, Thailand. He holds a Doctor of Philosophy (Information and Communication Technology for Education) from King Mongkut's University of Technology North Bangkok. He currently works as a lecturer in Information Technology, Digital Media Technology, and Computer Science, and holds the position of Vice President.

Nichanun Samakthai holds a Master's degree in Information Technology from King Mongkut's University of Technology North Bangkok. She is currently a lecturer in the Department of Digital Technology at the Faculty of Science and Technology, Rajamangala University of Technology Suvarnabhumi.

Pinyaphat Tasatanattakool, Ph.D., she is an Associate Professor in the Digital Technology Major, Faculty of Science and Technology, Rajamangala University of Technology Suvarnabhumi, Nonthaburi, Thailand. She holds a Doctor of Philosophy (Information and Communication Technology for Education) from King Mongkut's University of Technology North Bangkok. She currently works as a lecturer in Information Technology, Digital Media Technology, and Computer Science, and holds the position of Head of the Digital Technology Department at the Nonthaburi Center.

How to cite this paper

Suwut Tumthong, Nichanun Samakthai, Pinyaphat Tasatanattakool, "Evaluating LLM-Assisted CLO–PLO Alignment Decisions for AUN-QA–Aligned Curriculum Mapping", International Journal of Modern Education and Computer Science(IJMECS), Vol.18, No.3, pp. 121-134, 2026. DOI:10.5815/ijmecs.2026.03.08