ASCILITE 2026

The Future is Connection:

Collaboration, Creativity and Compassion in a Digital Society

Towards AI-assisted validation of university course profiles: A work-in-progress framework for policy compliance

Xintong Tan, Aneesha Bakharia, Lizzie Li, Hui Zhou

The University of Queensland

Reviewing university course profiles against institutional policies remains a largely manual process that is time-consuming and repetitive. This paper presents the design of an AI-assisted framework for course profile university policy compliance checking that combines Large Language Model (LLM)-based semantic analysis with deterministic rule-based validation, together with the initial implementation and evaluation of its LLM-based validation component. A prototype using Claude Sonnet 4.5 was developed and evaluated though an exploratory case study on a published course profile, using a baseline scenario and three modified versions with deliberately introduced policy violations. Although the LLM was able to recognise a stated policy exception occasionally, it produced false positives on the compliant baseline course profile, failed to detect a numerical error, and returned inconsistent validation outcomes for the same input. These findings show that LLM-only validation is not sufficiently reliable, motivating a combined approach that integrates deterministic rule-based validation with an LLM-based agentic workflow. This work represents the first stage of a broader study investigating hybrid AI-assisted approaches to course profile policy validation in higher education.

Keywords: course profile, policy compliance checking, Large Language Models, agentic workflow, educational technology, higher education, case study

Introduction

Universities increasingly rely on curriculum management systems to manage teaching information, support academic governance, and communicate essential course information to students. Course profiles provide students with important information including learning outcomes, assessment requirements (i.e. assessment due dates, the number of assessments, the weighting of assessments, whether the assessment is a sequence of activities, extension and late submission rules) and teaching activities. Ensuring the accuracy, quality and timeliness of these documents is important because they support institutional compliance, influence students' decision making, and shape the overall teaching and learning experience (Ho et al., 2021). A course profile must also adhere to university (e.g., an assessment policy mandating that no assessment item can have a weighting of over 60%), faculty and school policies.

At The University of Queensland (UQ), course coordinators prepare an internal course instance within the curriculum management system, which is subsequently reviewed by learning designers and school reviewers before being published as a student-facing course profile. During this review process, reviewers manually check each course instance against university policies and faculty/school-specific checklist. These checks include requirements such as secure assessments weightings and extension policies. The review process often involves multiple rounds of revisions between course coordinators and reviewers, making it repetitive, time-consuming and susceptible to human error, particularly during periods of high workload.

Recent advances in artificial intelligence (AI), particularly LLMs have created new opportunities for automating document compliance checking. However, course profile validation extends beyond interpreting natural language. It also requires structural, numerical, and semantic checks, and logical validations across university and faculty policy and school-specific requirements. Although automated compliance checking is well established in finance and legal services, its application to higher-education course documentation remains comparatively underexplored (Zaki et al., 2023).

This paper presents the design of a hybrid AI-assisted framework for validating university course profiles, together with the initial implementation of its LLM-based validation component. To investigate the feasibility of the LLM component before introducing a deterministic layer, a preliminary experiment was conducted on a published UQ course profile. The findings motivate the development of a hybrid validation framework incorporating an agentic workflow that can decompose validation into smaller tasks and support iterative review cycles. This provides the foundation for a broader research project that will progressively investigate and compare additional AI-assisted validation approaches.

Related work

Rule-based validation is one of the earliest approaches to automated compliance checking. In higher education, von der Heyde et al. (2023) reformulated study regulations into a Formal Specification Language and validated structured records using Answer Set Programming (ASP) with high accuracy, and Banbara et al. (2019) similarly applied ASP to course timetabling problems. Such methods excel in deterministic checks but struggle with interpreting ambiguous or contextual policies.

To address these limitations, recent research has explored natural language processing (NLP) and LLMs for compliance checking. Wen et al. (2025) found that transformer-based embeddings, including BERT and Word2Vec, outperformed earlier methods for compliance document matching, and Datla et al. (2025) explored using LLMs to translate policy documents into executable rules. Although LLMs demonstrate strong semantic understanding and generate human-readable explanations, their limitations include hallucinations, prompt sensitivity and structural and logical reasoning. These complementary strengths and limitations suggest that combining LLMs with deterministic rule-based validation may provide a more reliable approach to policy compliance checking.

A further development in this area is the use of agentic workflows, in which AI agents make decisions, take actions and coordinate tasks that are decomposed from an LLM's work. An example is Dasgupta and Shankar's (2025) Agents-as-Judge, which adopted orchestration tools such as LangChain, Crew AI, Guidance and TruLens to assign document-splitting, structural check, consistency check and feedback generation tasks to separate different agents. It achieved a 50% reduction in error on compliance checking of enterprise documents, suggesting that decomposing a compliance checking task across different specialised agents can improve reliability and accuracy, compared to solely relying on a single LLM prompt. It was also found that LangGraph is a powerful graph-based framework that mixes deterministic and agentic steps in the same workflow, achieving the highest task success rate and lowest hallucination rate across different workflows for AutoGPT and ReAct (Bhaskara Reddy Udaru, 2026).

Overall, existing research shows that rule-based systems, LLM-based systems, and agentic workflows offer complementary strengths for addressing different components of the course profile validation task providing a clear motivation for the proposed hybrid framework.

Proposed framework

Based on the limitations identified in existing compliance checking approaches, this study proposes a hybrid AI-assisted framework that combines LLM-based semantic analysis with deterministic rule-based validation and an agentic workflow for course profiles validation. Within the proposed framework, LLMs are intended to interpret policy statements and identify semantic inconsistencies within course profiles, while deterministic rules are intended to validate requirements involving numerical calculations, structural constraints and predefined institutional policies. The framework also proposed to coordinate them as separate agents within a graph-based workflow, for example, LangGraph, aiming to improve validation task's reliability and accuracy through decomposing the task across specialised agents. Particularly, each agent is responsible for validating one single rule and their results will be aggregated to generate a result. Overall, this framework adopts a modular architecture to support the integration and evaluation of additional validation approaches as the project progresses.

An initial prototype was developed to investigate the feasibility of applying LLMs to course profile validation. The web-based prototype allows reviewers to upload course profile information as free texts and receive a validation report highlighting potential policy compliance issues with actional and explanatory feedback. Relevant university policies are collected and loaded for prompting or retrieval, while faculty and school-specific rules are represented separately to support the planned deterministic validation component. Rules can also be enabled, disabled or added as university regulations evolve without redesigning the system.

The current prototype implements only the LLM-based validation component of the proposed framework. Both the deterministic rule-based component and the agentic workflow have been designed but have not yet been implemented. Consequently, the preliminary evaluation reported in this paper solely focuses on investigating the feasibility and limitations of the LLM-based validation to inform the subsequent development of the hybrid framework.

Methodology

A preliminary evaluation was conducted as an exploratory case study using a published UQ course profile to investigate the feasibility of using an LLM for university course profile policy validation. To maximise reproducibility, all evaluations were performed using Claude Sonnet 4.5 with the same prompt, policy documents, and rules.

The published course profile was first analysed using the prototype to establish a baseline and assess whether the system could correctly identify existing policy compliance issues. To investigate the capability of LLM-based validation under different compliance scenarios, modified versions of the course profile were created by deliberately introducing policy violations. As shown in Table 1, these modifications included removing mandatory policy statements, altering assessment weightings, and introducing inconsistencies between related assessment items. Each test case was evaluated across three repeated runs using identical input.

The generated validation reports were manually compared with the introduced policy violations to determine whether the LLM correctly identified the compliance issues. Therefore, the analysis of this experiment focused on qualitative observations of the model's ability to detect compliance issues, interpret policy documents and give actional feedback.

Table 1
Experimental test cases for course profile validation

Test case	Description	Purpose
Baseline	Published UQ course profile	Test original course profile
Case 1	Missing course contact email	Test semantic detection
Case 2	Changed assessment weighting	Test numerical reasoning
Case 3	Introduced exceptions	Test exception handling

Preliminary evaluation

Four scenarios were evaluated using a published UQ course profile, including the original profile and three modified versions. The results of different test cases are presented in Table 2.

Table 2
Evaluation outcomes

Test case	Description	Outcome
Baseline	Published Anon Uni course profile	False positives
Case 1	Missing course contact email	Detected once
Case 2	Changed assessment weighting such that the total does not sum up to 100%	Not detected
Case 3	Introduced extension limit violation but explicitly approved by the Associate Dean (Academic) (ADA)	Flagged but noted as acceptable due to ADA approval

Baseline

Although the original published course profile was expected to be largely policy compliant, the prototype reported 13 potential compliance issues. Manual inspection showed that several of these were false positives instead of policy incompliances. One common example was the assessment check, where the LLM incorrectly interpreted the Web Project Demonstration and Code Review as an additional assessment item rather than an allowable sequence component of the existing Web Project, resulting in an incorrect assessment count. The model also did not catch that ADA approval had been given for a 100% late penalty, still flagging it as an error and suggesting obtaining ADA approval. Overall, the baseline evaluation suggests that while the LLM can identify potential policy issues, it also produces false positives that require manual verification.

Case 1: Semantic policy validation

Case 1 evaluated the ability of the LLM to identify a missing mandatory course contact email. The missing email was detected in an earlier execution of the experiment but was not identified when the same test case was repeated. This inconsistency suggests that LLM-based validation may produce different validation outcomes across repeated evaluations.

Case 2: Numerical reasoning

The second scenario evaluated numerical reasoning by increasing one assessment weighting, resulting in a total assessment weighting of 110%. According to university policy, the total assessment weighting must equal 100%, therefore the expected outcome was that the validator would report the error. However, the prototype did not successfully identify the injected arithmetic error. Instead, the model generated 11 compliance issues, none of which corresponded to the incorrect assessment weighting. This means that numerical validation is a key limitation of LLM-based compliance checking and that deterministic checks are more appropriate for rules requiring explicit numerical calculations.

Case 3: Policy exception handling

The third scenario investigated whether the LLM could correctly interpret policy exceptions. An assessment extension exceeding the standard policy limit was accompanied by an explicit statement indicating that the variation had been approved by the ADA. The model recognised the stated approval and downgraded the issue severity from an error to a warning, indicating an ability to handle contextual policy exceptions. However, this reasoning was not applied consistently across the course profile. Similar to the baseline evaluation, the model continued to flag the ADA approved 100% late penalty as an error. This inconsistency suggests that while the LLM can recognise policy exceptions, it does not apply exception reasoning consistently.

Limitations

The evaluation was conducted using a single published course profile with a small number of manually constructed test cases. These test cases do not represent the full scope of course profile policy requirements, and the evaluation only focused on qualitative capabilities rather than quantitative performance. Additionally, the evaluation demonstrated that although the model can identify many policy issues, it also produced several false positives, miscalculations and inconsistencies. False positives would negatively impact the productivity benefits of the review process since they suggest reviewers to investigate issues that did not violate any policy rules. As some of these false positives arise from semantic misunderstandings, for instance, the model flagging the ADA-approved variation as a policy violation, instead of numerical errors, they might not be resolved simply by adding another deterministic rule layer. This suggests that having an agentic workflow that cross-checks the semantic and deterministic components with consistency is highly important. Furthermore, reliance on an external, API-based LLM introduces inherent vulnerabilities to model drift, meaning the model's non-deterministic behaviour may yield varying validation outcomes over time, even with identical prompts. These findings also align with the weaknesses reported in the literature regarding compliance checking. Model hallucination, high false positive rates and inconsistent logical reasoning have been identified as the main limitations of LLM-based document compliance checking (Dasgupta & Shankar, 2025).

Conclusion and future development

This paper presented a work-in-progress framework for AI-assisted validation of university course profile using LLMs. A prototype was developed to validate course profiles against university policies and faculty/school-specific requirements. Preliminary evaluation demonstrated that LLMs can effectively identify many policy issues and provide helpful explanatory feedback. However, the evaluation also revealed several limitations, including false positives, unreliable numerical checks and inconsistent detection across repeated trials. These findings suggest that course profile policy validation is inherently a hybrid task for which no single method is sufficient. Future research will develop a hybrid agentic workflow that decomposes course profile validation into separate policy checks, incorporates LLM-based reviewers, and uses deterministic validation for numerical requirements such as assessment weightings and assessment counts. This approach will enable different validation strategies to be applied where they are most appropriate, with the aim of improving overall accuracy and reliability. An accurate automated course profile policy checker could save valuable time for course coordinators, administrative reviewers, and learning designers, while providing course coordinators with real-time feedback on potential compliance issues before submission.

References

Banbara, M., Inoue, K., Kaufmann, B., Okimoto, T., Schaub, T., Soh, T., Tamura, N., & Wanko, P. (2019). Solving the curriculum-based course timetabling problems with answer set programming. *Annals of Operations Research, 275*(1), 3–37. https://doi.org/10.1007/s10479-018-2757-7

Bhaskara Reddy Udaru. (2026). From retrieval to cognitive orchestration: Standardizing context management in agentic AI systems. *International Journal of Computational and Experimental Science and Engineering*, *12*(1). https://doi.org/10.22399/ijcesen.4970

Dasgupta, S., & Shankar, H. (2025). AI agents-as-judge: Automated assessment of accuracy, consistency, completeness and clarity for enterprise documents. *arXiv*. https://doi.org/10.48550/arXiv.2506.22485

Datla, G. V., Vurity, A., Dash, T., Ahmad, T., Adnan, M., & Rafi, S. (2025). Executable governance for AI: Translating policies into rules using LLMs. *arXiv.* https://doi.org/10.48550/arXiv.2512.04408

Ho, D. V., Phan, T. H., & Pham, N. D. (2021). An efficient curriculum management ICT-based system to support AUN-QA. *AIP Conference Proceedings, 2406*(1), 020018. https://doi.org/10.1063/5.0066453

von der Heyde, M., Otunuya, C., Goebel, M., Zoerner, D., & Lucke, U. (2023). Automatic and interactive validation of study regulations in accreditation processes of higher education institutions. In *Proceedings of the European University Information Systems Congress 2023*. https://doi.org/10.29007/wc93

Wen, B., Wang, T., Xu, J., Liu, Y., Li, J., & Lin, S. (2025). File compliance detection using a Word2Vec-based semantic similarity framework. *Informatica, 49*(18). https://doi.org/10.31449/inf.v49i18.7421

Zaki, N., Turaev, S., Shuaib, K., Krishnan, A., & Mohamed, E. (2023). Automating the mapping of course learning outcomes to program learning outcomes using natural language processing for accurate educational program evaluation. *Education and Information Technologies, 28*(12), 16723–16742. https://doi.org/10.1007/s10639-023-11877-4

Tan, X., Bakharia, A., Li, L. & Zhou, H. (2026). Towards AI-assisted validation of university course profiles: A work-in-progress framework for policy compliance. In Cochrane, T., Narayan, V., Poole, T., Lai, P., Lake, P., Moore, O., & Li, L. (Eds.), *The Future is Connection: Collaboration, Creativity and Compassion in a Digital Society*. Proceedings ASCILITE 2026. Brisbane (pp. 837–841). https://doi.org/10.65106/apubs.2026.3317

Note: All published papers are refereed, having undergone a double-blind peer-review process.
The author(s) assign a Creative Commons by attribution license enabling others to distribute, remix, tweak, and build upon their work, even commercially, as long as credit is given to the author(s) for the original creation.

© Tan, X., Bakharia, A., Li, L. & Zhou, H. 2026

ASCILITE 2026

The Future is Connection:

Collaboration, Creativity and Compassion in a Digital Society