

# AI SYSTEM FOR SYLLABUS COMPLIANCE

Mrs. A. JANANI PRASANNA

Anupama R, Dhanabalan M, Dhana Varshini S, Kowsalya P

BACHELOR OF TECHNOLOGY – 4th YEAR

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

SRI SHAKTHI OF ENGINEERING AND TECHNOLOGY(AUTONOMOUS)

COIMBATORE-641062

## ABSTRACT

Manual syllabus verification is inherently time-consuming, inconsistent, and dependent on subjective judgment. This paper presents an AI-powered Syllabus Compliance Analyzer that automates curriculum coverage evaluation through semantic similarity analysis. The system extracts structured topics from uploaded documents, generates contextual embeddings using Sentence-BERT, and employs FAISS-based vector search with cosine similarity to classify each topic as fully covered, partially covered, or missing. An overall compliance percentage is computed and presented through interactive dashboards with AI-driven improvement recommendations. Experimental results demonstrate that the semantic embedding approach significantly outperforms traditional keyword-matching methods in detecting contextual alignment between syllabus content and instructional materials, offering a scalable solution for institutional quality assurance and accreditation support.

**Keywords:** *Syllabus Compliance Analysis, Sentence-BERT, Semantic Similarity, FAISS, Curriculum Evaluation, Natural Language Processing, Academic Quality Assurance.*

## INTRODUCTION

Ensuring alignment between prescribed syllabi and actual instructional delivery is a critical component of academic quality assurance. Traditional verification methods rely on manual comparison by faculty or administrators, making the process labor-intensive, inconsistent, and susceptible to human oversight. Existing digital tools primarily employ keyword-based matching, which fails to capture the contextual and semantic relationships inherent in academic content. This limitation results in inaccurate coverage assessment and an inability to quantify curriculum compliance in a measurable, objective manner.

To address these shortcomings, this paper proposes an AI-driven Syllabus Compliance

Analyzer that leverages transformer-based Natural Language Processing to perform semantic evaluation of curriculum coverage. The system utilizes Sentence-BERT for contextual embedding generation and FAISS for efficient high-dimensional vector similarity search, enabling accurate classification of syllabus topics as fully covered, partially covered, or missing. Built on a FastAPI backend and React.js frontend, the platform delivers real-time compliance scoring, visual dashboards, and AI-generated recommendations, providing institutions with an automated, scalable, and data-driven mechanism for curriculum monitoring and accreditation support.

## LITERATURE REVIEW

Automated syllabus compliance and curriculum evaluation have gained increasing attention as institutions seek objective, data-driven approaches to academic quality assurance. Researchers have explored the integration of artificial intelligence, semantic analysis, and structured curriculum mapping to replace manual verification methods. These studies collectively establish the conceptual and technical foundation for intelligent systems capable of evaluating alignment between prescribed syllabi and instructional delivery.

Albert (2025) examined the integration of artificial intelligence into engineering curricula, emphasizing structured syllabus design and AI-based academic frameworks. The research highlighted the growing importance of intelligent systems in curriculum planning and quality enhancement, providing a conceptual foundation for automated syllabus compliance evaluation systems.

Ljubić, Tomić, and Volarić (2025) discussed structured approaches for simplifying syllabus development using digital tools. Their work focused on improving clarity, organization, and consistency in syllabus creation, supporting the need for automated systems that verify alignment between planned curriculum and instructional delivery.

Adorni, Grosso, and Ponzini (2024) examined challenges in syllabus alignment and curriculum evaluation within educational systems. The authors analyzed structured curriculum mapping approaches and highlighted gaps in topic coverage and consistency, emphasizing the importance of automated, data-driven evaluation methods to improve syllabus alignment accuracy across academic programs.

Mohamed Mahtar (2015) presented a structured framework for evaluating curriculum content using analytical methods, examining topic

coverage, learning objectives, and assessment alignment to ensure consistency with prescribed syllabi. The study demonstrated how intelligent analysis techniques can enhance academic quality assurance, forming a strong foundation for developing automated syllabus compliance system.

## METHODOLOGY

### 1. Data Collection and Preprocessing

Each syllabus and teaching content record in the AI Syllabus Compliance Analyzer is processed through structured and validated input mechanisms to ensure accuracy and consistency of academic information.

- **Diverse Data Sources:** Collect syllabus topics and instructional content directly from uploaded PDF and DOCX files, including unit headings, module topics, subtopics, and teaching material segments to ensure complete curriculum coverage.
- **Data Formats:** Process documents using structured text extraction pipelines with defined parsing rules for headings, paragraphs, and topic boundaries to maintain consistency across varied academic document formats.
- **Input Validation:** Apply validation checks within the React frontend interface to ensure uploaded files are properly formatted and non-empty before initiating the compliance analysis pipeline.
- **Standardization:** Normalize extracted text by removing unnecessary characters, correcting formatting inconsistencies, and standardizing case to maintain uniformity across all documents for semantic analysis.
- **Categorization:** Classify extracted syllabus components into structured units such as topics, subtopics, and modules, enabling precise mapping and comparative analysis against instructional content.

- **Data Storage:** Store all processed syllabus data and compliance reports securely within the backend system, ensuring persistence, easy retrieval, and integration with the analytical and visualization modules.

## 2. Feature Extraction and Topic Detection

Feature extraction identifies meaningful academic content patterns for compliance evaluation

- **Topic Extraction:** Apply pattern-based parsing and rule-driven segmentation to identify units, modules, headings, and subtopics from syllabus documents, converting each into an independent analytical entity.
- **Content Segmentation:** Divide teaching materials into smaller, meaningful textual chunks based on sentence boundaries and paragraph divisions to preserve semantic coherence and improve analytical precision.
- **Structural Analysis:** Analyze document layout and heading hierarchy to distinguish primary topics from subtopics, ensuring accurate representation of the syllabus structure.
- **Comparative Evaluation:** Map extracted syllabus topics against segmented instructional content to assess coverage depth and detect fully covered, partially covered, or missing topics.

## 3. Semantic Embedding Model and Architecture

The embedding architecture enables accurate semantic comparison between syllabus topics and instructional content.

- **Transformer-Based Design:** Utilize Sentence-BERT, a pre-trained transformer model, to generate high-quality contextual embeddings that capture semantic meaning beyond simple keyword matching.

- **Embedding Generation:** Convert each extracted syllabus topic and instructional content chunk into dense fixed-length vector representations that preserve conceptual relationships between academic terms and phrases.
- **Vector Indexing:** Integrate FAISS to build an efficient high-dimensional vector index, enabling fast storage and retrieval of embeddings across large academic content datasets for scalable compliance evaluation
- **Similarity Computation:** Compute cosine similarity scores between syllabus topic embeddings and teaching content embeddings to measure the degree of semantic alignment between prescribed and delivered curriculum content.

## 4. Compliance Classification and Scoring

The classification and scoring workflow ensures precise, reliable evaluation of syllabus coverage across diverse academic content.

- **Threshold-Based Classification:** Apply predefined similarity thresholds to categorize each syllabus topic as fully covered, partially covered, or missing based on computed cosine similarity scores.
- **Coverage Mapping:** Record the highest similarity match for each topic to ensure accurate alignment detection, even when different terminology or phrasing is used across documents.
- **Compliance Score Calculation:** Aggregate individual topic classification results to compute an overall compliance percentage that quantitatively represents the degree of curriculum adherence.
- **Weighted Scoring:** Apply weighted scoring mechanisms that assign higher importance to fully covered topics while proportionally accounting for partial coverage in the final compliance metric.

## 5. Performance Evaluation

The following metrics and subjective evaluations are used:

- **Similarity Metrics:** Evaluate the system using cosine similarity scores per topic, overall compliance percentage, and classification accuracy across fully covered, partially covered, and missing categories.
- **Threshold Tuning:** Periodically review and adjust similarity thresholds to reduce false positives and false negatives, improving classification reliability across diverse academic subjects.
- **Validation Testing:** Test the system on varied academic content including different subjects, teaching styles, and document formats to ensure robustness and generalizability.
- **Cross-Validation:** Employ diverse academic datasets during evaluation to assess model stability and semantic precision under varying content distributions.
- **Error Analysis:** Analyze misclassified topics caused by specialized terminology, ambiguous phrasing, or inconsistent writing styles to guide future refinements and threshold adjustments.

## 6. Report Generation and Deployment

This section highlights potential or actual integration of the models into real-world systems:

- **Report Generation:** Produce structured analytical reports containing topic-wise similarity scores, coverage classifications, and compliance summaries for faculty and administrative review.
- **Visual Dashboards:** Display compliance percentages, topic coverage distributions, and similarity score charts through interactive React.js dashboards for clear interpretation of results.

- **AI Recommendations:** Generate AI-based suggestions highlighting missing or weakly covered syllabus topics, providing actionable guidance to educators for improving curriculum alignment.
- **Continuous Optimization:** Continually refine similarity thresholds, embedding configurations, and classification boundaries through user feedback and performance monitoring to improve system accuracy and scalability.

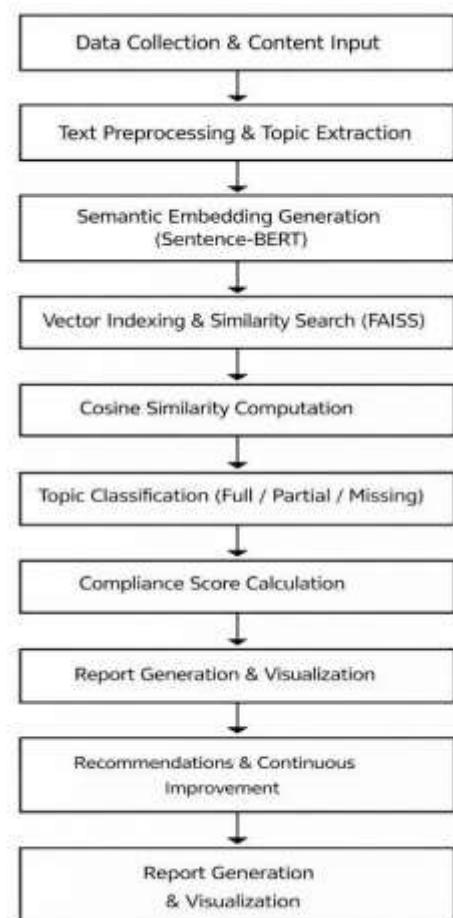


Fig 1: Proposed Methodology

## SYSTEM DESIGN

The system design for the AI Syllabus Compliance Analyzer integrates several detailed stages to ensure accurate and reliable curriculum evaluation across diverse academic

content. The process begins with comprehensive data collection and preprocessing, where syllabus documents and teaching materials are uploaded in PDF or DOCX formats and subjected to structured text extraction, normalization, and segmentation. Raw academic content is cleaned, standardized, and divided into meaningful topic units and instructional chunks, ensuring that all input text is well-structured and semantically coherent before analysis begins. Topic extraction is performed using pattern-based parsing and rule-driven segmentation techniques that identify units, modules, headings, and subtopics from syllabus documents, converting each component into an independent analytical entity for precise comparison against instructional materials.

The core architecture employs a transformer-based semantic embedding approach, utilizing Sentence-BERT to generate high-dimensional contextual vector representations of both syllabus topics and segmented teaching content. Unlike traditional keyword-matching systems, this model captures deep semantic relationships between academic concepts, enabling accurate similarity detection even when different terminology or phrasing is used across documents.

Classification and compliance scoring refine the analysis outputs by applying predefined similarity thresholds to categorize each syllabus topic as fully covered, partially covered, or missing. Individual topic scores are aggregated using weighted scoring mechanisms to produce an overall compliance percentage that objectively represents curriculum adherence. The React.js frontend delivers interactive dashboards with real-time compliance visualizations, topic-wise analysis tables, and AI-generated recommendations, enabling faculty and administrators to identify curriculum gaps and make informed decisions for improving academic quality assurance and accreditation readiness.

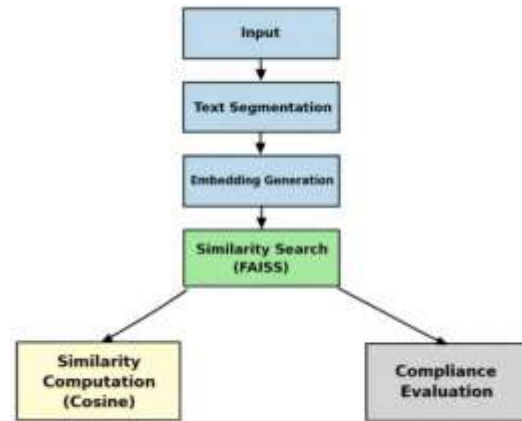


Fig 2: Semantic Embedding Model Workflow

## IMPLEMENTATION

**Step 1:** Collect syllabus documents and teaching materials in PDF and DOCX formats from diverse academic sources. Ensure the dataset includes variations in document structure, subject domains, writing styles, and curriculum formats to support robust compliance evaluation across different institutional contexts.

**Step 2:** Upload syllabus and instructional content files through the React.js frontend interface. Parse uploaded documents using PyMuPDF and python-docx libraries to extract raw text content. Store extracted text in structured format within the FastAPI backend for further preprocessing and analysis.

**Step 3:** Apply text preprocessing techniques to clean and standardize extracted academic content.

- Remove unnecessary characters, symbols, and formatting inconsistencies from raw extracted text to ensure input clarity.
- Standardize text case and correct spacing errors to maintain uniformity across all documents before semantic analysis.
- Apply structured chunking techniques based on sentence boundaries and

paragraph divisions to segment teaching materials into meaningful, computationally efficient units.

- Eliminate duplicate or irrelevant content segments to prevent semantic noise during embedding generation and similarity computation.

**Step 4:** Apply topic extraction using pattern-based parsing and rule-driven segmentation to identify units, modules, headings, and subtopics from syllabus documents. Convert each extracted topic into an independent analytical entity to enable precise one-to-one mapping against instructional content segments during compliance evaluation.

**Step 5:** Design the semantic embedding pipeline using Sentence-BERT, a pre-trained transformer model optimized for generating meaningful sentence-level embeddings. Pass each extracted syllabus topic and segmented teaching content chunk through the transformer encoder, where multiple self-attention layers compute contextual representations by assigning weights to semantically important words within each sequence.

**Step 6:** Generate fixed-length dense vector embeddings for all syllabus topics and instructional content segments using Sentence-BERT. Ensure embeddings capture contextual meaning and preserve semantic relationships between academic concepts, enabling accurate similarity detection even when different terminology or sentence structures are used across documents.

**Step 7:** Integrate FAISS to index all generated high-dimensional embedding vectors from instructional content segments. Build an efficient vector index structure that enables fast nearest-neighbor similarity retrieval across large academic datasets, ensuring scalable and real-time compliance analysis performance.

**Step 8:** Compute cosine similarity scores between each syllabus topic embedding and all indexed instructional content embeddings using

FAISS-based vector search. Retrieve the highest similarity match for each topic to determine the strongest conceptual alignment between the prescribed syllabus requirement and the available teaching material.

**Step 9:** Apply predefined similarity thresholds to classify each syllabus topic based on computed cosine similarity scores.

- Topics with similarity scores at or above 0.75 are classified as fully covered, indicating strong semantic alignment with instructional content.
- Topics with scores between 0.50 and 0.75 are classified as partially covered, indicating the concept is mentioned but insufficiently detailed.
- Topics with scores below 0.50 are classified as missing, highlighting areas where instructional content does not adequately address the syllabus requirement.

**Step 10:** Aggregate individual topic classification results to compute an overall compliance percentage representing the degree of curriculum adherence. Apply weighted scoring mechanisms that proportionally account for fully covered and partially covered topics in the final compliance metric to produce an objective quantitative measure of syllabus alignment.

**Step 11:** Generate AI-based recommendations by analyzing missing and partially covered topics. Prioritize recommendations based on similarity score severity, flagging topics with scores below 0.2 as high priority and moderately partial topics as medium priority, providing actionable guidance to educators for improving curriculum delivery.

**Step 12:** Develop structured analytical reports and interactive visual dashboards using React.js displaying compliance percentages, topic-wise similarity scores, coverage classification tables, and AI-generated suggestions. Evaluate system performance by monitoring classification

accuracy, similarity score distributions, and threshold effectiveness across diverse academic subjects and document formats to support continuous optimization and refinement.

## OUTPUT

The system provides an interactive interface where users can securely log in, upload syllabus documents and teaching materials, and perform automated compliance analysis. The output displays the overall compliance percentage along with the number of fully covered, partially covered, and missing topics. It also presents detailed topic-wise similarity scores and matched content segments. Additionally, the system generates AI-based recommendations to help improve coverage of missing or weakly addressed topics, enabling better curriculum alignment and academic quality improvement.



Fig 4: Compliance Analysis Results



Topic	Status	Similarity	Recommendation
Topic 1	Fully Covered	0.95	
Topic 2	Partially Covered	0.75	Review content
Topic 3	Missing	0.00	Add content
Topic 4	Partially Covered	0.60	Review content
Topic 5	Missing	0.00	Add content
Topic 6	Partially Covered	0.80	Review content
Topic 7	Missing	0.00	Add content
Topic 8	Partially Covered	0.70	Review content
Topic 9	Missing	0.00	Add content
Topic 10	Partially Covered	0.65	Review content

Fig 5: Topic Compliance Details



Fig 6: AI Recommendation

## FUTURE ENHANCEMENTS

While the current system demonstrates effective automated syllabus compliance analysis, several enhancements can further improve the intelligence, scalability, and practical impact of the platform. The following areas highlight potential future developments:

### 1. Adaptive Learning Analytics Integration

Integrate student performance data, assessments, and learning analytics to evaluate not only syllabus coverage but also the effectiveness of content delivery and student understanding.

### 2. Cross-Course Curriculum Consistency Analysis

Extend the system to analyze multiple courses within a program to identify topic redundancies, curriculum gaps, and ensure balanced knowledge distribution across courses.

### 3. Predictive Curriculum Gap Detection

Incorporate machine learning models to predict potential syllabus coverage gaps using historical course data, enabling proactive curriculum planning and improvement.

#### **4. Automated Curriculum Benchmarking**

Enable comparison of institutional syllabi with national and international academic standards or industry requirements to ensure curriculum relevance and competitiveness.

#### **5. Knowledge Graph Based Academic Mapping**

Implement knowledge graph techniques to model relationships between topics, prerequisites, and learning outcomes to support deeper academic insights and structured curriculum planning.

### **BENEFITS**

#### **1. Automated Curriculum Verification**

Eliminates manual syllabus comparison by automatically analyzing teaching materials, saving time and reducing human errors in academic evaluation.

#### **2. Improved Academic Quality Assurance**

Helps institutions identify missing or partially covered topics, ensuring better curriculum delivery and maintaining academic standards.

#### **3. Objective and Data-Driven Evaluation**

Provides quantitative compliance scores and similarity analysis, enabling unbiased and measurable curriculum assessment.

#### **4. Intelligent Recommendations**

Generates AI-based suggestions to improve weak or missing topic coverage, supporting faculty in enhancing course content.

#### **5. Scalability and Efficiency**

Supports analysis of large volumes of academic documents with fast semantic search, making it suitable for institutions with multiple courses and departments.

#### **6. Enhanced Decision Support**

Provides analytical dashboards and reports that assist educators and administrators in making informed curriculum improvement decisions.

### **CONCLUSION**

The AI Syllabus Compliance System presents an intelligent and scalable approach for automating the evaluation of curriculum coverage using advanced Natural Language Processing and semantic similarity techniques. By leveraging transformer-based embeddings and vector similarity search, the system effectively identifies fully covered, partially covered, and missing topics within teaching materials. The platform reduces the limitations of manual verification by providing objective compliance scores, detailed analytical reports, and AI-driven recommendations. This improves transparency, accuracy, and consistency in curriculum assessment while supporting educators in enhancing course delivery. Furthermore, the system's modular and scalable design makes it suitable for institutional quality assurance processes and future integration with advanced academic analytics. Overall, the proposed system demonstrates how AI can significantly improve curriculum monitoring, academic decision-making, and the overall effectiveness of education management systems.

## REFERENCES

1. Albert, L. A. (2025). *Artificial intelligence in systems: Integrating AI into the engineering curriculum*. Available at SSRN 5240570.
2. Ljubić, H., Tomić, Z., & Volarić, T. (2025). *From Vision to Structure: Simplifying Syllabus Development for Educators*. In EDULEARN25 Proceedings (pp. 1576–1581). IATED.
3. Adorni, G., Grosso, D., & Ponzini, D. (2024). *Building a Conversational AI Syllabus for Educator Certification: A Framework for Integrating AI in Educational Practice*. In ICERI2024 Proceedings (pp. 9216–9226). IATED.
4. Mohamed Mahtar, S. N. A. (2015). *Analyzing the quality of artificial intelligence final examination questions according to Bloom's taxonomy and syllabus contents*. Doctoral dissertation, Universiti Teknologi MARA (UiTM).
5. Mittal, A. M., Parthasarathy, P. D., & Joshi, S. (2025). *AI Ethics Education in India: A Syllabus-Level Review of Computing Courses*. In Annual ACM India Compute Conference (pp. 70–85). Springer Nature Switzerland.
6. Dutta, K., Paul, S., & Anand, A. (2024). *GovGPT: An Ethics-Integrated Governance Architecture for Curriculum-Aligned, Child-Centric Educational AI Systems*. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 5(1), 180–185.
7. Schneider, B. (Year not specified). *The Hidden Curriculum of AI Syllabus Statements: How Fear-Based Framing Shapes Student Behaviour and Academic Integrity*.
8. Chomphooyod, P., Jeerapradit, L., Suchato, A., & Punyabukkana, P. (2025). *Multi-agentic AI for Automatic Course Syllabus Generation using LangGraph*. In 2025 10th International STEM Education Conference (iSTEM-Ed) (pp. 1–6). IEEE.
9. Miedema, D., Taipalus, T., Ajanovski, V. V., Alawini, A., Goodfellow, M., Liut, M., & Young, T. (2025). *Data systems education: Curriculum recommendations, course syllabi, and industry needs*. In Working Group Reports on Innovation and Technology in Computer Science Education (pp. 95–123).
10. Sadovnychenko, I., Pastukhova, N., & Miasoiedov, V. (2025). *Ensuring compliance with the principles of academic integrity in an era of distance medical education*. Baltija Publishing. *omputing and Communication: Proceedings of 7th ICICC 2024, Volume 2, 1241, 79*.