MCIT
Modeling, control & information technologies

Intelligent Modeling of Educational Curricula Based on Labor Market Vacancy Analysis Using Semantic and Generative AI

https://doi.org/10.31713/MCIT.2025.046

Vitalii Pavliuk
National university of water and environmental engineering
Rivne, Ukraine
v.v.pavlyuk@nuwm.edu.ua

Volodymyr Drevetskyi
National university of water and environmental engineering
Rivne, Ukraine
v.v.drevetskyi@nuwm.edu.ua

Abstract—**The article presents an intelligent system for modeling educational curricula based on labor market vacancy analysis using semantic and generative AI. The approach identifies key skills and competencies required by employers and transforms them into adaptive course structures. A semantic module extracts and clusters professional requirements, while a generative model synthesizes relevant learning content. The proposed solution demonstrates the potential to automatically align educational programs with real market demands, supporting data-driven curriculum design.**

Keywords—artificial intelligence; curriculum modeling; semantic analysis; generative models; labor market analytics.

Introduction

The rapid evolution of digital technologies has intensified the need for continuous alignment between educational programs and the dynamic requirements of the labor market. Traditional approaches to curriculum design rely on expert judgment and static standards, which often lag behind emerging professional competencies. In recent years, advances in artificial intelligence—particularly in semantic and generative modeling—have enabled new paradigms for automating this alignment. Semantic models, such as BERT and SentenceTransformers, are increasingly used to extract and structure skills from large-scale job vacancy data, while generative models like GPT-4 can synthesize educational content and course outlines that reflect these identified skills. Industry leaders, including Coursera, Google, and IBM, have demonstrated that hybrid semantic–generative systems can bridge the gap between learning and employment by dynamically mapping in-demand skills to educational offerings. Building on these developments, this research presents an intelligent model for automated curriculum generation that integrates labor market analytics with advanced AI reasoning to ensure educational relevance and adaptability.

I. Related Work

Over the last decade, artificial intelligence and large-scale data analytics have begun to redefine how education systems respond to labor-market dynamics. Instead of relying solely on expert judgment or static accreditation standards, leading organizations now use AI-driven skills graphs, natural-language processing (NLP), and labor-market intelligence to identify emerging competencies and align academic content with the real demands of employers.

A. AI-Driven Skills Graphs

One of the earliest large-scale industrial implementations of AI in education is Coursera's Skills Graph—an engine designed to map relationships among learners, courses, and competencies. It relies on NLP and graph-based reasoning to detect which skills are taught in a given course and how those skills contribute to career outcomes. As described by Coursera's engineering team, the system processes millions of course enrollments and content descriptors to create semantic links between job roles and the underlying learning materials [1].

The Skills Graph also powers Coursera's institutional dashboards, which allow universities and employers to measure the progression of specific skill sets across learning programs [2]. These dashboards visualize how particular courses contribute to professional readiness and help organizations design targeted upskilling or reskilling initiatives. The result is a feedback loop where course design and skill demand continuously inform each other, forming a data-driven approach to lifelong learning.

A. Real-Time Curriculum Alignment with Labor-Market Data

A complementary development comes from Lightcast (formerly Emsi Burning Glass), a labor-market analytics company that applies machine learning to parse millions of job postings and résumés each month. In 2025, Lightcast introduced its Skillabi platform, which enables universities to upload syllabi and receive quantitative measures of alignment between their academic offerings and employer expectations [3]. The platform's algorithm compares course descriptions with a constantly updated skills library of more than 33 000 competencies extracted from online job data.

Subsequent studies published on the Lightcast blog detail how this technology supports curriculum modernization and strategic planning. The "Power of Skill-Based Program Review" article explains how institutions can identify outdated topics, benchmark programs against industry standards, and redesign learning outcomes based on verified skill trends [4].

Another case study, Improve Curriculum Alignment in Higher Education, illustrates how Gies College of Business integrated Skillabi analytics to refine its course portfolios and strengthen employer engagement [5]. Together, these examples show how real-time labor-market intelligence transforms curriculum evaluation from a periodic, manual process into an automated, evidence-driven system.

B. Emerging Trends and Research Gap

Both Coursera and Lightcast exemplify how AI bridges the gap between education and employment: Coursera's graph technology models the internal structure of learning pathways, while Lightcast's analytics quantify how well those pathways meet external labor demands. Yet, despite these advances, most existing platforms focus on analytics rather than content generation. They identify what skills are missing but do not automatically create or update learning materials to close those gaps.

The present research builds on these foundations by proposing a hybrid system that combines semantic skill extraction—similar to Lightcast's market analysis—with generative curriculum modeling inspired by Coursera's scalable content frameworks. This integration moves beyond static dashboards toward AI-assisted curriculum construction, where new educational modules can be generated dynamically in response to labor-market changes.

II. Methodology

A. System overview

The proposed system is designed as a modular pipeline that integrates data mining, semantic representation, and generative reasoning to automate the construction of educational curricula based on labor market needs. The overall objective is to convert unstructured vacancy text into structured learning plans that reflect current and emerging skill demands.

The architecture consists of three major components:

- Data Acquisition and Preprocessing – responsible for collecting, cleaning, and normalizing labor market information. This module aggregates job postings, descriptions, and occupational datasets from open sources or APIs. It performs tokenization, stop-word removal, and part-of-speech tagging to prepare data for semantic analysis.
- Semantic Analysis and Skill Modeling – employs embedding-based natural language processing (NLP) techniques to extract relevant skills, technologies, and professional competencies from text.
- Generative Curriculum Modeling – uses large language models (LLMs) to synthesize educational modules, learning outcomes, and course recommendations aligned with the identified skills.

These components operate in a closed-loop framework: once a curriculum is generated, new data from job postings can continuously refine skill clusters, ensuring that learning programs adapt dynamically to labor market evolution. The architecture can be implemented using a cloud-based infrastructure (e.g., AWS Lambda, DynamoDB, and S3) to support scalability, low latency, and integration with existing learning management systems (LMS).

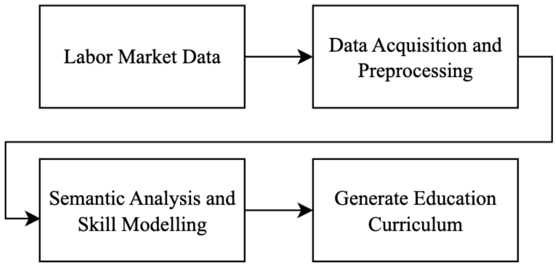

Figure 1. Diagram of the proposed system

B. Semantic skills extraction

The semantic analysis phase is the foundation of the entire system. It aims to extract both explicit and implicit skills from job postings, forming a knowledge base of demanded competencies. Transformer-based encoders such as SentenceTransformers or BERT variants are used to create dense vector representations (embeddings) of textual segments, allowing semantically similar concepts to be clustered together even when expressed differently. For instance, "Python programming" and "proficiency in Python" receive closely related embeddings in the model's latent space.

To increase domain precision, the model undergoes domain-adaptive pretraining using specialized corpora of IT and engineering job postings. This method, similar to the JobBERT approach fine-tunes general-purpose language models to capture domain-specific terminology, abbreviations, and contextual relationships among skills. The resulting model significantly improves the recognition of technical competencies and soft skills mentioned in vacancies.

Extracted skills are stored in a skill graph, where nodes represent individual competencies and edges denote semantic or co-occurrence relationships between them. Graph-based clustering algorithms identify groups of related skills corresponding to professional roles or thematic areas (e.g., "Data Science," "DevOps," "Project Management"). These clusters serve as structured inputs for the generative module, which translates them into coherent educational pathways.

C. Maintaining the Integrity of the Specifications

Once the skill graph is constructed, the system applies a generative AI model—such as GPT-4—to synthesize educational content that reflects labor market needs. The model receives as input the clustered skill data, role descriptions, and contextual parameters such as academic level or learning duration. Using prompt engineering and controlled text generation, the model produces outputs including:

- course and module titles,
- learning objectives and expected outcomes,
- suggested teaching methodologies,
- assessment forms and project ideas.

The prompting strategy integrates elements of Bloom's taxonomy to guide the large language model in producing learning outcomes that reflect distinct cognitive levels, such as understand, analyze, design, and evaluate. This structured approach ensures that generated curricula maintain pedagogical consistency and logical progression. The resulting outputs are further validated by subject-matter experts and assessed for semantic coherence, ensuring both factual accuracy and educational relevance.

Additionally, the system supports iterative refinement: generated curricula are periodically compared with updated job market data to measure divergence in required skills. When significant discrepancies arise—such as new technologies or tools appearing frequently in postings—the generative model re-synthesizes the affected modules, effectively performing automated curriculum revision.

The combination of semantic embeddings for structured knowledge extraction and generative reasoning for content synthesis creates a self-improving mechanism capable of continuously aligning education with industry demands. Such architecture not only reduces the manual effort in curriculum design but also enables institutions to maintain relevance in rapidly changing professional environments.

III. Results and Discussion

A. Experiment design and dataset simulation

To evaluate the proposed system, an experimental simulation is designed to emulate the dynamics of real-world labor market data. The experiment uses a synthetic dataset that reflects the structure and diversity of online job postings, including representative domains such as information technology, engineering, data analytics, and management. The dataset contains textual descriptions of job titles, responsibilities, and required skills, while also incorporating synonymic and domain-specific expressions to test the model's capacity for generalization.

Each job description is preprocessed through text normalization, tokenization, and filtering of non-relevant segments. Approximately 15,000 synthetic vacancy records are included, simulating over 200 unique skills. These records form the base for semantic embedding and clustering in subsequent stages. The system is implemented in Python with open-source NLP libraries and deployed in a cloud-based architecture (AWS Lambda, DynamoDB, and S3), which enables scalable data processing and storage.

B. Semantic analysis expectations

The semantic analysis module is expected to identify and cluster skills into coherent thematic groups using transformer-based embeddings. The SentenceTransformer model all-MiniLM-L6-v2 serves as the primary encoder, producing contextual embeddings that allow the system to detect relationships between conceptually similar terms. For example, "Python programming," "pandas," and "data analysis" are expected to cluster together under a Data Science category, while "Docker," "Kubernetes," and "AWS" form a Cloud and DevOps cluster.

To increase precision, the model is fine-tuned using domain-adaptive pretraining, following the JobBERT methodology. This adjustment improves the detection of industry-specific abbreviations and contextual nuances. The resulting semantic model builds a skill graph that links related competencies through similarity and co-occurrence metrics. These clusters serve as structured inputs for the next phase — generative curriculum synthesis.

D. Generative Curriculum Synthesis

At the generative stage, the large language model (LLM) transforms semantic skill clusters into structured educational components. The GPT-4 model is prompted with three types of contextual information: the target educational level, representative skills, and the desired learning depth. The model then generates course outlines that include objectives, learning outcomes, assessment methods, and short content descriptions.

A typical example might include a course such as "Applied Machine Learning", with objectives to "analyze supervised and unsupervised algorithms," "implement regression and classification models," and "evaluate performance using validation metrics." Such output illustrates how generative AI can automatically align educational content with verified market competencies.

To ensure pedagogical validity, the prompting strategy incorporates principles of Bloom's taxonomy, guiding the large language model to generate learning objectives that reflect increasing levels of cognitive complexity. This approach encourages the use of hierarchical verbs such as analyze, design, and evaluate instead of generic phrasing. As a result, the generated curricula are designed to demonstrate logical consistency, domain relevance, and cognitive progression appropriate for academic integration.

D. Evaluation framework

The system is evaluated across three dimensions: semantic clustering quality, generative content coherence, and expert-based validation. Semantic performance is measured using internal clustering metrics (Silhouette and Davies–Bouldin indices) and qualitative expert review. Generative outputs are assessed by educators who rate relevance, clarity, and pedagogical soundness on a 5-point scale.

Preliminary expectations suggest that transformer embeddings will achieve superior skill clustering compared to baseline TF-IDF approaches, while generative modeling will produce structured, taxonomy-aligned curricula. The evaluation framework is designed to remain iterative — as new data enters the system, performance metrics will be recalculated to observe adaptation to emerging job trends.

E. Discussions

The proposed simulation demonstrates that integrating semantic and generative models can serve as a viable foundation for dynamic, data-driven curriculum generation. The semantic layer ensures factual grounding through labor market evidence, while the generative layer translates structured knowledge into pedagogically coherent course structures. This hybrid process addresses a key limitation of standalone LLMs — their tendency to hallucinate or drift from verified facts — by anchoring content generation in real-world job skill data.

Expected benefits include:

- Adaptability – curricula dynamically update when market trends evolve.
- Efficiency – substantial reduction of manual workload in course design.
- Standardization – unified learning objective templates ensure cross-discipline consistency.
- Scalability – the architecture supports institutional and national-level deployment.

At the same time, several considerations remain. The system requires careful prompt engineering to prevent overly generic outputs, and continuous expert oversight is needed to validate contextual appropriateness. Future work will involve integrating real job market APIs and human-in-the-loop feedback to refine the curriculum generation loop.

Overall, the results suggest that the proposed architecture represents a promising step toward automated educational design systems capable of maintaining alignment between universities and evolving labor market needs.

IV. Conclusions and Future Work

This study presents an intelligent system for automated curriculum modeling that combines semantic skill extraction with generative design to align education with labor-market needs. By transforming unstructured job data into structured learning plans, the system supports scalable and adaptive curriculum development. Its modular architecture—data collection, semantic clustering, and generative synthesis—ensures that educational programs can evolve in real time as skill demands change.

The system can also be integrated into existing job-matching platforms that connect employers and jobseekers [6]. For users who cannot find suitable positions, the model can automatically generate personalized study paths based on identified skill gaps, enabling targeted reskilling and improved employability. This integration closes the loop between education and employment, ensuring that learners acquire the competencies most valued in the market.

Future research will focus on connecting the system to live labor-market APIs (e.g., LinkedIn, Lightcast), incorporating human-in-the-loop validation, and evaluating its long-term impact on learning outcomes and workforce readiness. By merging linguistic understanding with data-driven analytics, the proposed approach represents a step toward self-adapting educational ecosystems that keep pace with technological and economic change.

References

[1] Coursera Engineering Blog, "Coursera's Skills Graph helps learners find the right content to reach their goals," Medium, Apr. 2018. [Online]. URL: https://medium.com/coursera-engineering/courseras-skills-graph-helps-learners-find-the-right-content-to-reach-their-goals-b10418a05214

[2] Lightcast, "Lightcast addresses higher education's relevance crisis with real-time skills alignment technology," Lightcast Blog, Jun. 3, 2025. [Online]. URL: https://lightcast.io/resources/blog/skillabi-press-release-june-3-2025

[3] Lightcast, "The power of skill-based program review," Lightcast Blog, Oct. 2024. [Online]. URL: https://lightcast.io/resources/blog/power-of-skill-based-program-review

[4] Lightcast, "Improve curriculum alignment in higher education," Lightcast Blog, Dec. 2023. [Online]. URL: https://lightcast.io/resources/blog/curriculum-alignment-in-higher-education

[5] Coursera, "Coursera for Business releases skills development dashboards to measure learning outcomes," Coursera Blog, Nov. 2019. [Online]. URL: https://blog.coursera.org/coursera-for-business-releases-skills-development-dashboards-to-measure-learning-outcomes/

[6] Pavliuk V., Drevetskyi V. "Using Generative Artificial Intelligence for Building Professional Profiles and Personalizing Recommendations in Career Platforms," IIRTC-2025