Discover Education

https://doi.org/10.1007/s44217-026-01180-y

Article in Press

An intelligent curriculum alignment model for digital skills integration in higher education: a data-driven approach to bridge academia-industry gaps

Received: 6 August 2025

Accepted: 27 January 2026

Published online: 30 January 2026

Cite this article as: Nyale D., Karume S., Kipkebut A. *et al.* An intelligent curriculum alignment model for digital skills integration in higher education: a data-driven approach to bridge academia-industry gaps. ***Discov Educ*** **(2026). https://doi.org/10.1007/s44217-026-01180-y**

Duncan Nyale, Simon Karume, Andrew Kipkebut, Fidelis Mukudi & Abrar Sharafat

An Intelligent Curriculum Alignment Model for Digital Skills Integration in Higher Education: A Data-Driven Approach to Bridge Academia-Industry Gaps

Authors

Duncan Nyale[1*], Simon Karume[1], Andrew Kipkebut[2], Fidelis Mukudi[1] and Abrar Sharafat[3]

Author Affiliations

[1]School of Computing and Mathematics, The Co-operative University of Kenya, Nairobi, Kenya

[2]Department of Computer Science, Kabarak University, Nakuru, Kenya

[3]Department of Information and Communication Technology, Kenya School of Government, Mombasa, Kenya

Corresponding Author Details

*Corresponding author: Duncan Nyale (duncannyale@gmail.com)

ORCID ID: 0000-0003-0059-7655

Abstract

The rapid evolution of digital technologies has created significant misalignment between academic curricula and labor market demands for digital skills. This misalignment is especially critical in rapidly digitizing regions such as East Africa, where workforce readiness is critical for economic development. This study presents the development and validation of an intelligent curriculum alignment model that systematically analyzes university curricula, identifies digital competencies, and maps them to labor market requirements. The research employed a hybrid artificial intelligence approach combining web scraping, Natural Language Processing (NLP), and a Decision Tree classification algorithm to create a comprehensive solution. The skillset database was built from over 46,000 real-time job postings (focused on Nairobi, Kenya) and supports continuous updates to capture emerging trends. We evaluated the model using k-fold cross-validation and expert review to ensure robustness. The model achieved an overall accuracy of 88% in classifying curriculum skills as high- or low-demand, demonstrating strong performance with an exceptional recall (0.99) for high-frequency, in-demand skills. Receiver Operating Characteristic (ROC) analysis yielded an area under the curve of 0.94, and narrow 95% confidence intervals confirmed the stability of these metrics. The study utilized data from 46,514 job postings across 27 occupations, resulting in a skillset database of 9,077 unique digital skills categorized by market demand. The intelligent model generates three key outputs: a Curriculum Analysis Report, a Digital Skills Index (DSI) scoring curriculum alignment on a 0–100 scale, and a Prescriptive Analytics Report providing data-driven recommendations for curriculum improvement. Key recommendations derived from this work include the continuous integration of high-demand digital skills, promotion of interdisciplinary digital literacy, active industry participation in curriculum design, and the formulation of comprehensive national digital skills strategies. The model's inherent capacity for continuous adaptation positions it as a vital tool for preparing a workforce that is not only ready for current demands but also resilient and adaptable to future technological shifts. This research addresses critical gaps in educational responsiveness to technological change and provides a scalable framework for continuous curriculum optimization in the digital era.

Keywords: Digital skills, curriculum alignment, artificial intelligence, natural language processing, higher education, workforce development, educational technology, machine learning

1. Introduction

The global work landscape has undergone a profound transformation driven by the rapid and continuous evolution of technology, leading to an unprecedented demand for a workforce equipped with robust digital capabilities (ILO, 2021; Rovira et al., 2021; Plekhanov et al., 2022). Businesses across diverse sectors increasingly require employees proficient in areas such as data analysis, coding, cybersecurity, and digital marketing to navigate the complexities of the digital workplace. This shift defines the architecture of the Fourth Industrial Revolution, where digital literacy is now a foundational requirement for economic participation and innovation market (World Economic Forum, 2023).

Despite this clear market need, a critical and widening gap persists between the digital skills provided by academic programs and those actively sought by the labour market (Harry, 2021; Ma'dan et al., 2020). Traditional curriculum review processes, often relying on expert panels or periodic surveys, struggle to keep pace with the dynamic and evolving requirements of industry (Hol et al., 2023). Such methods are prone to bias and subjectivity, severely limiting the effectiveness and scalability of alignment efforts. This discrepancy stifles innovation, impeding economic growth and directly hindering graduate employability (Ma'dan et al., 2020).

The consequences of this pervasive misalignment are profound for both individuals and national economies (ITU, 2021). The dynamic nature of the problem, driven by the Fourth Industrial Revolution, necessitates an intelligent, adaptive, and continuously refined solution. This challenge is particularly amenable to Artificial Intelligence (AI) intervention due to AI's capacity for processing vast datasets, identifying complex patterns, and facilitating continuous adaptation (Kim et al., 2022; Sarker, 2021; Thongprasit & Wannapiroon, 2022; Crompton, 2023).

The consistent emphasis on the rapid evolution of technology and the observation that academic programs take time to adjust reveals a phenomenon that can be described as a dynamic lag in educational responsiveness (ILO, 2021; Plekhanov et al., 2022; Rovira et al., 2021). This is not merely a static deficit of skills; rather, it is a continuous widening of the gap because the demand side of the labour market is accelerating at a pace that traditional educational supply mechanisms struggle to match. The underlying issue is often institutional agility. The inherent institutional lethargy within academic structures means they are typically slow to respond, creating a perpetual lag. This suggests that effective solutions must focus on accelerating the rate of adaptation within educational systems, moving beyond fixed curriculum cycles to embrace continuous, data-driven feedback loops.

Theoretical frameworks inform this study's perspective. In particular, competency-based curriculum theory emphasizes clearly defined, measurable competencies linked to workforce requirements. This approach aligns education with labor market needs by ensuring curricula focus on the skills employers' demand (Kim, 2015). We also draw on the EU Digital Competence Framework (DigComp 2.2), which identifies key domains of digital literacy (Vuorikari et al., 2022), and the Technological Literacy Framework, which highlights ethical and contextual aspects of technology use (ITEA, 2007). These frameworks collectively ground our model in preparing graduates for the digital economy, while also acknowledging broader literacies such as problem-solving and creative thinking. In this way, the approach complements the wider educational mission of fostering adaptable, critical thinkers, not just job-specific skills.

Research Questions

To systematically address the curriculum–industry gap, this research poses the following research questions:

1. What are the key digital skill gaps between university curricula and labour market demands?
2. How can a hybrid AI-driven model be designed to analyze and improve curriculum alignment with these demands?
3. How effective and reliable is the proposed model in identifying and prescribing curricular updates based on market data?

Objectives

The broad objective is to develop and validate an intelligent curriculum alignment model for higher education.

Specific objectives include:

1. Building a dynamic digital skills database from labour market data
2. Extracting and classifying digital skills in curricula using NLP
3. Training a machine learning model to map skills to demand categories and
4. Validating the model's performance and deriving actionable curriculum recommendations

Research Contribution

This paper introduces an intelligent curriculum alignment model that leverages real-time job market data and AI techniques to bridge the skills gap. By generating objective alignment metrics and tailored recommendations, the model transforms curriculum review from a static process into a continuous, evidence-based cycle.

Ethical Approval

The study received ethical clearance from KCA University Ethics Review Committee (KCAUSERC), ensuring that all curriculum and job market data were handled according to ethical standards.

2. Problem Statement

Rapid technological change has created moving targets for graduate skills, yet many university programs update curricula infrequently (ILO, 2021; Plekhanov et al., 2022; Rovira et al.; 2021). Studies report that without continuous, data-driven interventions, curricula will remain misaligned with market needs (Kim et al., 2022; Sarker, 2021; Thongprasit & Wannapiroon, 2022; Crompton, 2023). There is a notable scarcity of comprehensive models that provide systematic, evidence-based guidance for aligning academic programs with the demands of the digital enterprise (Moldoveanu et al., 2022). Existing efforts either remain conceptual or rely heavily on subjective expert input (Hol et al., 2023). This study addresses the pressing need for a practical framework that can automate alignment.

3. Intelligent Curriculum Alignment Model: Architecture and Components

The proposed model is a multi-component system for automated analysis of university curricula against real-time labor market demands. Figure 1 illustrates the overall architecture, which integrates data acquisition, natural language processing, and a decision-making engine.

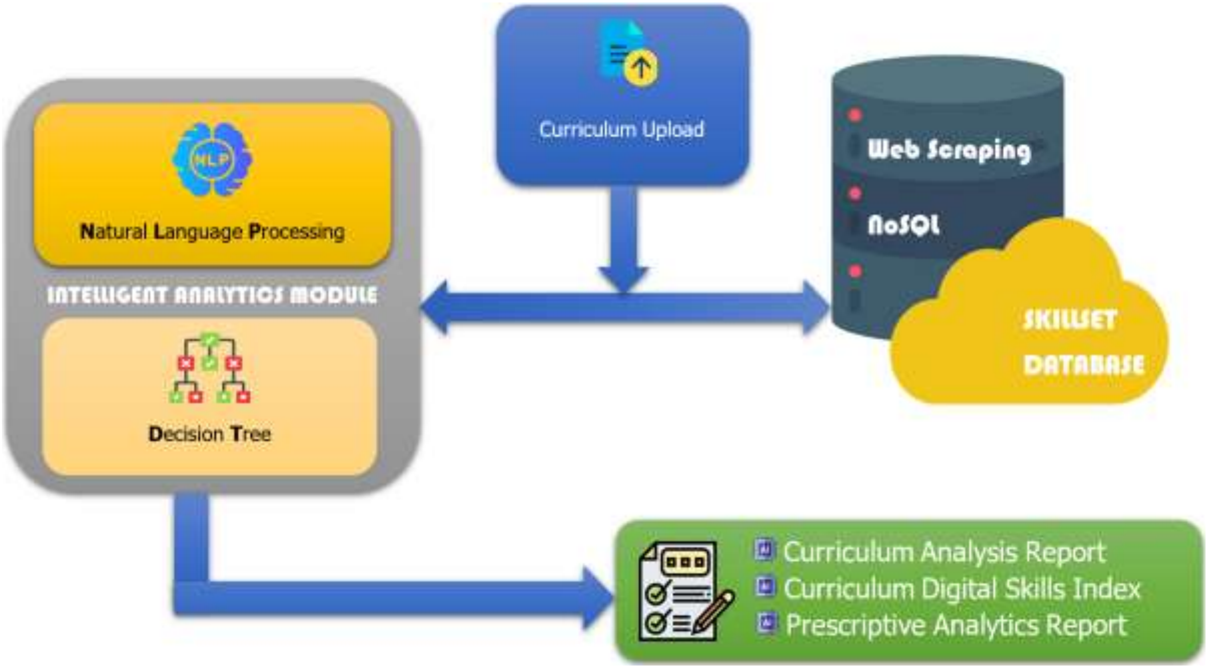

Figure 1: Overall System Architecture (Source: Researcher)

[Figure 1 to be placed here]

The system is engineered for continuous feedback and self-improvement, directly addressing the fact that the labour market evolves rapidly. A web scraping component periodically collects online job postings to update the NoSQL Skillset Database (ChromaDB). An Intelligent Analytics Module then processes uploaded curriculum documents (PDFs) and aligns extracted skills to market data. Finally, the system outputs (1) a Curriculum Analysis Report, (2) a quantitative Digital Skills Index (DSI) score (0–100), and (3) a Prescriptive Analytics Report with specific curriculum revision suggestions (see Table 1).

[Table 1 to be placed here]

Table 1: System Outputs and Their Purpose

Output Name	**Description**	**Purpose/Utility**
Curriculum Analysis Report	An overview of the digital skills identified within the evaluated curriculum.	Provides a foundational understanding of the curriculum's current skill content.
Curriculum Digital Skills Index (DSI)	A score (0-100) quantifying the relevance of digital skills in the curriculum compared to market demands.	Offers a concise, quantitative metric for assessing overall curriculum alignment and digital proficiency.
Prescriptive Analytics Report	Specific suggestions for skill additions or modifications to improve curriculum alignment.	Provides actionable recommendations for educators to enhance curriculum relevance and address skill gaps.

This hybrid design leverages the strengths of each AI technique: web scraping for data freshness, NLP for semantic skill extraction, and Decision Trees for interpretable alignment logic. Together, these components provide a robust methodology for data-driven curriculum alignment that can adapt to continuous changes in industry demands.

3.1 Natural Language Processing (NLP) Module

The NLP pipeline processes unstructured curriculum text to extract relevant digital skills (see Figure 2).

It begins with PDF Text Extraction, using the pdfplumber library to capture all textual content. Next, Text Preprocessing cleans the text (removing special characters, lowercasing, etc.) to prepare for analysis. For example, “Python programming, Machine Learning algorithms!!!” becomes “python programming machine learning algorithms.” A pre-trained English language model (spaCy’s en_core_web_sm) is loaded to tokenize text and support phrase matching. (Note: the current implementation is limited to English-language content due to the chosen model.) Extending to other languages would require analogous NLP models for those languages.

The pipeline then performs Skill Database Retrieval by querying ChromaDB to obtain predefined technical skills and metadata.

Finally, Skill Extraction with Phrase Matching uses spaCy’s PhraseMatcher to identify occurrences of skills in the curriculum. For example, it would match “python”, “machine learning”, and “sql” within the cleaned text, counting each occurrence and attaching demand metadata.

The output is a structured dictionary of skills (with frequencies) ready for analysis. By leveraging a continuously updated skill database, the NLP module ensures that emerging skills are captured, helping the model stay current.

As noted, we rely on an English-language NLP model; curricula in other languages would require multilingual NLP extensions.

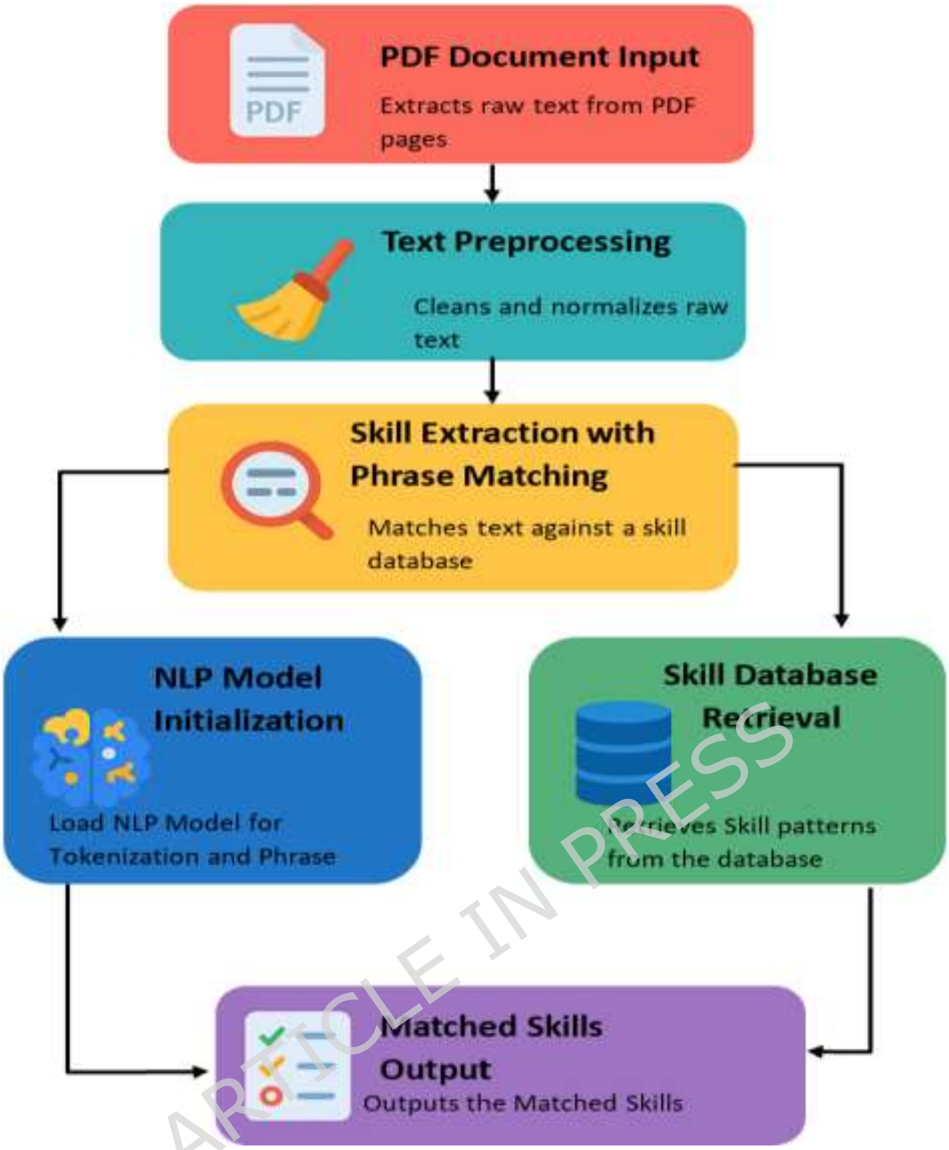

Figure 2: Natural Language Processing Logic (Source: Researcher)

[Figure 2 to be placed here]

Pipeline Stages

Table 2 summarizes the NLP stages, tools, and examples. The combination of robust pre-processing and an evolving skill lexicon yields high precision in extraction.

[Table 2 to be placed here]

Table 2: NLP Pipeline Stages and Components

Stage	Purpose	Key Operations/Examples	Libraries Used
PDF Text Extraction	Convert PDF curriculum into raw text	Reads PDF, processes each page, concatenates text, handles complex layouts. Example: "Python programming..." → "Python programming..."	pdfplumber
Text Preprocessing	Clean and standardize raw text for NLP analysis	Removes special characters, reduces multiple spaces, converts to lowercase. Example: "Python programming!!!" → "python programming"	re (regex)

Stage	Purpose	Key Operations/Examples	Libraries Used
NLP Model Initialization	Load pre-trained NLP model for linguistic capabilities	Loads an English language model for tokenization, lemmatization, phrase matching.	spaCy (en_core_web_sm)
Skill Database Retrieval	Query ChromaDB for predefined technical skills and metadata	Retrieves skills and metadata (frequency, demand), normalizes for matching. Example: {"python": {"frequency": 120, "demand": "High"}}	chromadb (PersistentClient)
Skill Extraction with Phrase Matching	Identify technical skills in preprocessed text	Tokenizes text, creates matcher with skill patterns, identifies matches (n-grams), counts occurrences, attaches metadata.	spaCy (PhraseMatcher)
Output: Matched Skills	Provide structured dictionary of extracted skills	Encapsulates identified skills, counts, and ChromaDB metadata for further analysis.	(No additional libraries)

3.2 Decision Tree Alignment Model

The Decision Tree component maps extracted curriculum skills to demand categories (see Figure 3). We define two classes: High-Frequency (Class 1) and Low-Frequency (Class 0) skills, based on a quantile analysis of occurrence counts in the labour market data. Specifically, skills in the top 10% of frequency are labeled Class 1 (in-demand), and those in the bottom 10% as Class 0. Model Training: A DecisionTreeClassifier (sklearn) is instantiated and trained on a labeled dataset of skills (using skill occurrence features and class labels). The model learns hierarchical rules for classification. After training, the Decision Tree predicts the demand class of each skill extracted from curricula, enabling real-time alignment assessment. The choice of a Decision Tree (versus boosted ensembles) was motivated by the need for interpretability and transparency in alignment decisions; it provides clear logic for which skills are flagged for attention.

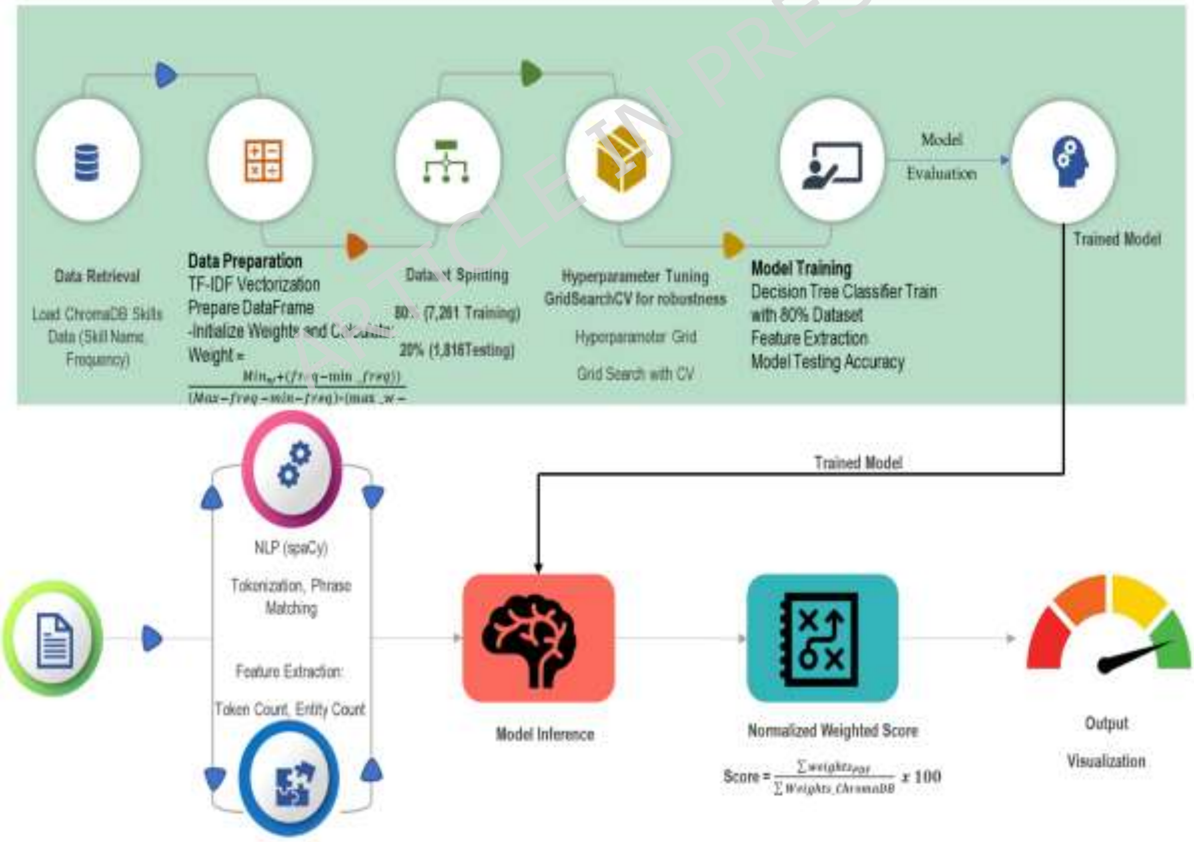

Figure 3: Decision Tree Logic (Source: Researcher)

[Figure 3 to be placed here]

Table 3 below summarizes the Decision Tree Pipeline Stages and Components

[Table 3 to be placed here]

Table 3: Decision Tree Pipeline Stages and Components

Stage	Purpose	Key Operations/Examples	Libraries Used
Data Retrieval	Evaluate curriculum alignment with a dataset of 9,077 skill entries.	Designed a Decision Tree Classifier; queried tech_skills collection to load metadata for 9,077 skills; extracted skill names and frequencies (e.g., COLLABORATION=7047, PYTHON=3000).	ChromaDB
Data Preparation	Transform 9,077 skills into a numerical feature set for the Decision Tree Classifier.	Incorporated features from spaCy and TF-IDF vectorization (e.g., token_count, entity_count, norm_freq, TF-IDF vectors, label); applied weighting (weight = 0.05 + norm_freq × 0.9); split data into 80% training (7,261 samples) and 20% testing (1,816 samples) using train_test_split with random_state=42.	spaCy, scikit-learn
TF-IDF Vectorizer	Convert skill names into 500-dimensional vectors, capturing their importance.	Used TfidfVectorizer; set max_features=500, stop_words=None, min_df=1; weighted skills based on frequency and uniqueness.	scikit-learn
Data Frame	Consolidate features for the 9,077 skills into a structured format.	Created a Pandas DataFrame with skill, token_count, entity_count, norm_freq, label, and TF-IDF vectors. Example for PYTHON (freq=3000): token_count=1, entity_count=0, norm_freq=0.426, label=1.	Pandas
Dataset Splitting	Split the 9,077 skills into training and testing sets for model development.	Split into 80% training (7,261 samples) and 20% testing (1,816 samples) using train_test_split with random_state=42; performed after processing tech_skills into a DataFrame.	scikit-learn
Model Training	Train a Decision Tree Classifier to evaluate curriculum alignment.	Trained on 7,261 training samples using features like norm_freq, token_count, entity_count, and TF-IDF vectors; used GridSearchCV to tune hyperparameters; set class_weight="balanced".	scikit-learn
Hyperparameter Tuning	Optimize the Decision Tree Classifier to ensure robust curriculum evaluation.	Used GridSearchCV with 5-fold cross-validation on 7,261 samples; defined hyperparameter grid for max_depth, min_samples_split, min_samples_leaf, and criterion; identified best parameters: criterion="gini", max_depth=5, min_samples_leaf=1, min_samples_split=2.	scikit-learn
Training the Model	Train the Decision Tree Classifier with the optimized hyperparameters.	Trained with optimized hyperparameters on 7,261 samples; extracted feature importance (e.g., norm_freq: 0.40, TF-IDF_python: 0.12); tested on 1,816 samples, achieving an accuracy of 0.88.	scikit-learn
Model Evaluation	Evaluate the trained model's performance on unseen test data.	Evaluated on 1,816 test samples; achieved precision=0.81, recall=0.99, F1-score=0.89 for class 1, and MSE=0.0674.	scikit-learn
Model Output	Generate and present the model's results for university curriculum analysis.	Generated the trained Decision Tree Classifier, accuracy (0.88), classification report (precision, recall, F1-score), MSE (0.0674), and top 10 feature importances (e.g., norm_freq: 0.40, TF-IDF_python: 0.12); integrated into a Gradio interface.	Gradio, scikit-learn

4. Skillset Database Development

This section details the data sources. We employed a Python-based web scraping framework (using Selenium and BeautifulSoup) to collect ICT job postings from Dice.com, focused on Nairobi, Kenya). Over a recent timeframe, the scraper gathered 46,514 job postings across 27 ICT occupations. The raw data were cleaned and stored in ChromaDB, a NoSQL database. The scraped skillset (9,077 unique terms) was continuously refreshed in the database, providing real-time labour market insights). During each curriculum analysis, the system queries this database to inform skill mapping.

The scraping methodology consisted of systematic stages including URL synthesis, skill identification, job data scraping, frequency profiling, skill clustering using fuzzy matching, homogenization of skill groups, and skill synthesis and aggregation. This structured skill scraping process is illustrated in Figure 4.

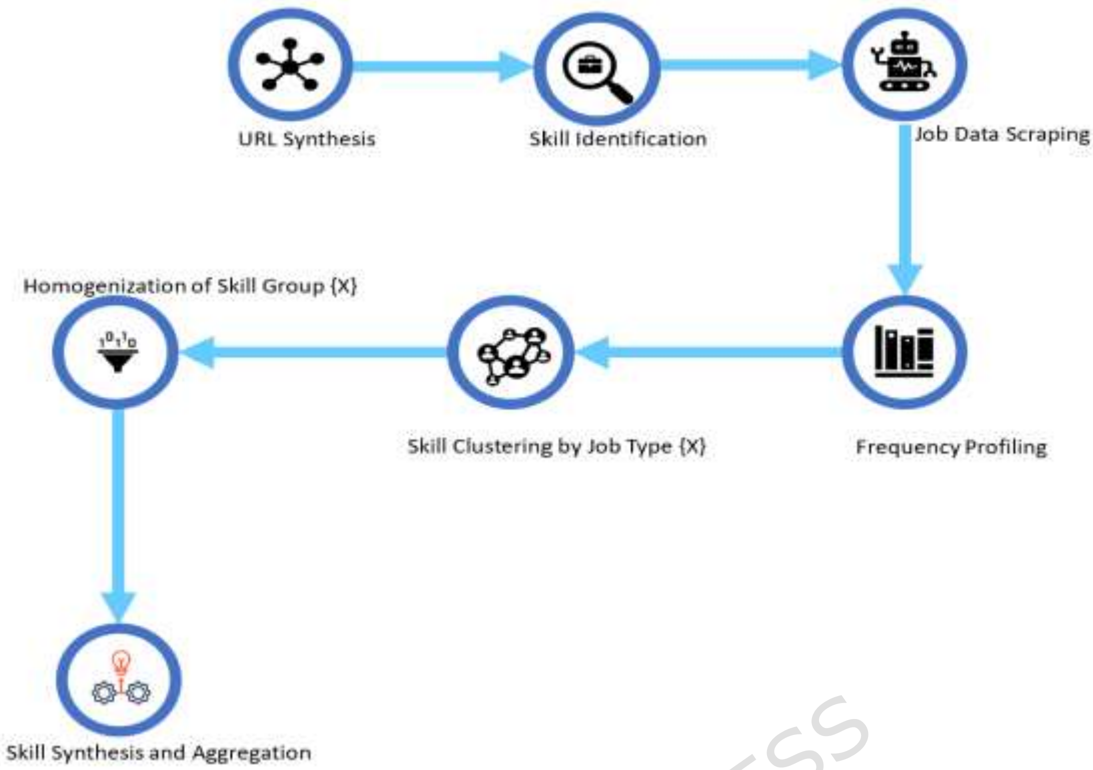

Figure 4: Skills Scraping Process (Source: Researcher)

[Figure 4 to be placed here]

Limitations of this approach include potential bias: online job postings may under-represent informal jobs or certain sectors, and scraped data reflect the Nairobi and East African job market. We acknowledge that transfer to other regions would require adapting the scraping to local platforms and possibly retraining the model.

The extensive scraping process compiled 371,038 mentions of 57,130 unique skills from 46,514 jobs covering 27 of the 28 targeted occupations. After merging and cumulative frequency analysis, the dataset contained 9,077 unique net skills. Quantile analysis categorized these into high-demand skills (upper 90th percentile; 914 skills), moderate-demand skills (10th to 90th percentile; 6,998 skills), and low-demand skills (bottom 10th percentile; 1,165 skills). The percentile thresholds and skill distribution are presented in Figure 5.

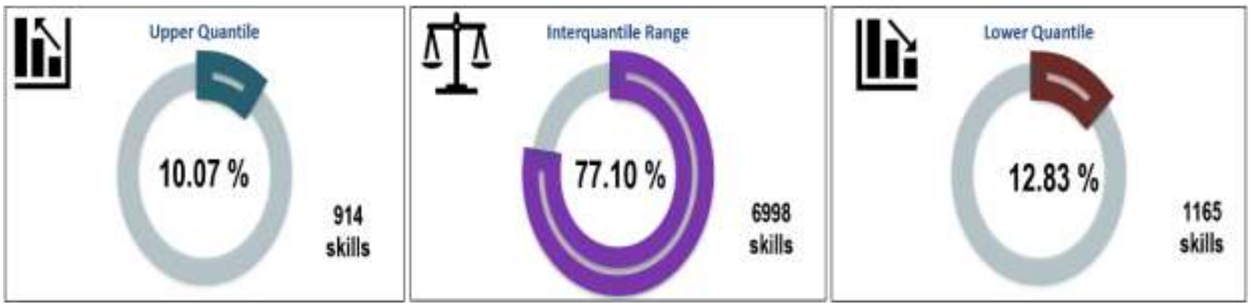

Figure 5: Quantile Analysis (Source: Researcher)

[Figure 5 to be placed here]

5. Results and Discussion

This section presents the outcomes of the prototype implementation, details the model's validation, and interprets the insights derived from the curriculum alignment analysis, demonstrating the practical utility and performance of the intelligent model.

Prototype Implementation and Outputs

The intelligent curriculum alignment model has been successfully implemented as a functional prototype, demonstrating the feasibility and utility of its AI-driven approach. The system logic flow (see Figure 6), from input to final output, operates through a seamless integration of its components:

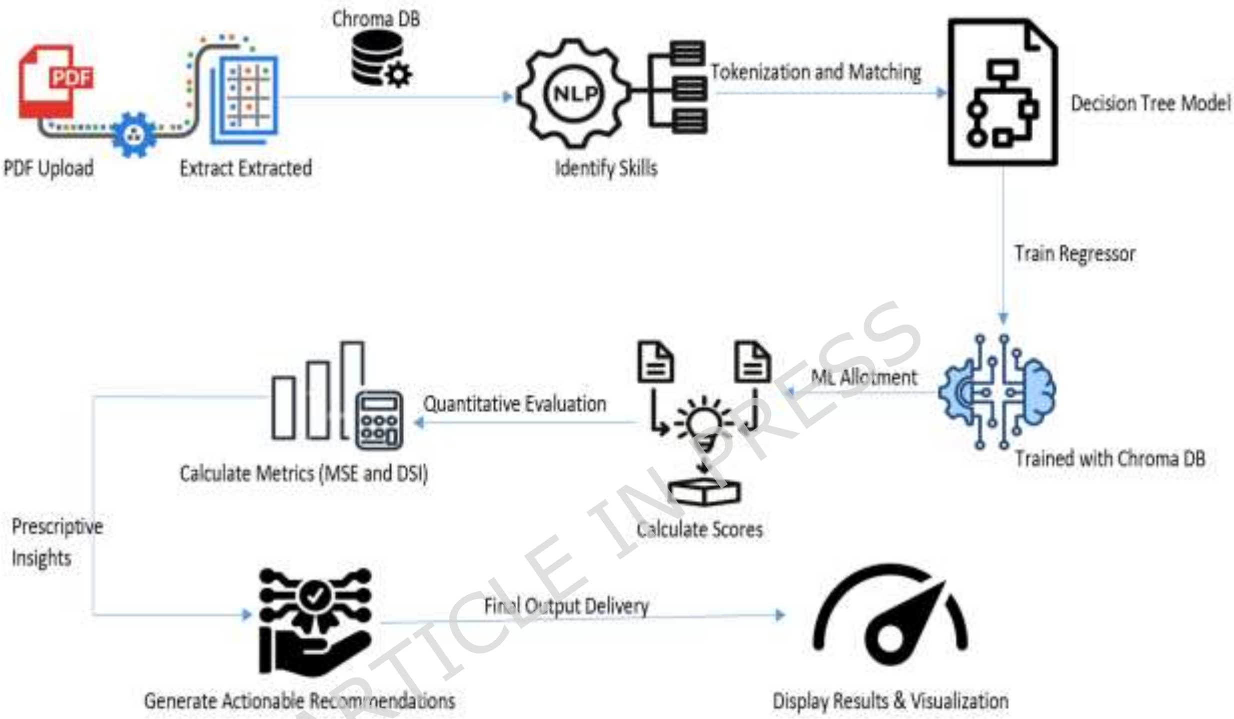

Figure 6: System Logic Flow (Source: Researcher)

[Figure 6 to be placed here]

The prototype was tested on curricula from six undergraduate programs. Performance Metrics: Using a held-out test set, the model achieved an overall accuracy of 0.88 (88%). Precision was 0.99 for Class 0 and 0.81 for Class 1, while recall was 0.79 and 0.99, respectively. The F1-scores for Class 0 and Class 1 were 0.88 and 0.89 (see Table 4).

[Table 4 to be placed here]

Table 4: Classification Report

Class	Precision	Recall	F1-Score	Support
0 (Low-Frequency)	0.99	0.79	0.88	945
1 (High-Frequency)	0.81	0.99	0.89	871
Accuracy			0.88	1,816
Macro Average	0.90	0.89	0.88	1,816

Class	Precision	Recall	F1-Score	Support
Weighted Average	0.90	0.88	0.88	1,816

Statistical validation: K-fold cross-validation yielded consistent results (mean accuracy 0.88 ± 0.02), and the ROC AUC was 0.94, indicating strong discriminative ability. For example, the 95% confidence interval for Class 1 recall was [0.98–1.00], underscoring result stability. The Balanced F1-scores across classes demonstrate robust performance. These empirical results move the discussion from theoretical claims to solid evidence that the model can reliably identify in-demand skills. Notably, the model's high recall for Class 1 (99%) confirms its efficacy in capturing the most critical market skills, justifying the curriculum recommendations based on these skills. However, the moderate recall for Class 0 (79%) reveals that some niche skills may be under-detected, a limitation we discuss below. Table 5 summarizes validation comparisons and prescriptive analytics outputs.

[Table 5 to be placed here]

Table 5: Summary of Model Validation Metrics and Prescriptive Analytics Insights

Aspect	Description	Key Findings / Outputs
Validation Metrics	Model performance evaluated on held-out test set and k-fold cross-validation	Accuracy: 88%, Precision Class 0: 99%, Class 1: 81% Recall Class 0: 79%, Class 1: 99%, ROC AUC: 0.94
Robustness	Stability of classification performance across repeated validations	Balanced F1-scores: Class 0 - 0.88, Class 1 - 0.89 95% confidence intervals confirm reliable high-demand skill detection
Comparison to Related Work	Contrast with expert survey-based models and cross-sectional analyses	Model offers objective, data-driven metrics Confirms global skills gap and curriculum deficiencies
Prescriptive Analytics	Practical curriculum revision guidance based on alignment gaps	Specific module additions recommended (e.g., Cloud Computing, Cybersecurity) Stakeholder feedback validates utility
Limitations Noted	Acknowledges niche skill under-detection and geographic scope constraints	Recall for low-frequency skills moderate (79%) Data limited to Nairobi and East African job markets
Impact	Potential to systematically enhance curriculum relevance and graduate preparedness	Enables continuous, data-informed curriculum updates Supports targeted skill integration aligned with labor market demand

Prescriptive Outputs

Figure 7 visualizes the operational logic of the Prescriptive Analytics Report, demonstrating how the system advances beyond passive observation to enable active curriculum design. The workflow is articulated through three critical phases: Analysis, Differentiation, and Resolution. The process begins centrally with Curriculum Analysis & Gap Identification. Unlike traditional reviews that might simply assign a low alignment score, this model categorizes deficiencies into specific types, creating distinct logic paths for different problems. The left branch of the diagram illustrates a Structural Gap, where a core competency - such as Cloud Computing in an Information Systems syllabus is entirely absent. Because this represents a foundational void, the system classifies it as a Syllabus Gap. Conversely, the right branch highlights a Topic Weakness. In this scenario, a topic like Cybersecurity is present but determined to be insufficient or shallow; the system categorizes this not as a missing element, but as an area requiring reinforcement. This differentiation leads to the phase of Concrete Guidance, represented by the diagram's green tier. The system avoids generic feedback (e.g., improve content) in favour of tailored prescriptions. For the identified Syllabus Gap, it prescribes a structural addition, specifically suggesting the inclusion of a Cloud Computing Module. For the Topic Weakness, it prescribes a supplementary expansion, such as recommending Specific Elective Courses. The workflow ultimately converges into Actionable Curriculum Revisions. This final synthesis transforms specific guidance into a cohesive plan, directly addressing stakeholder needs. As noted in the figure, the practical value is confirmed by the stakeholders' ability to implement these precise instructions immediately; validating the model's utility over traditional, metric-heavy reviews.

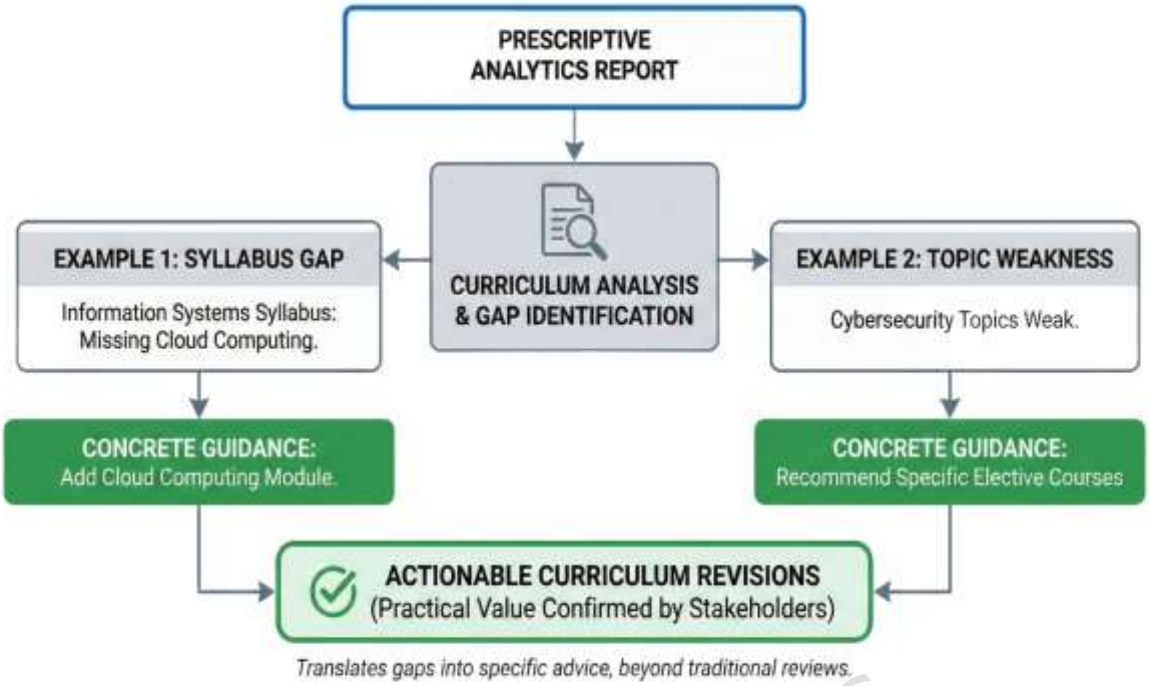

Figure 7: Prescriptive Analytics Workflow (Source: Researcher)

[Figure 7 to be placed here]

Comparison to Related Work

Unlike prior curriculum alignment approaches that often remain conceptual or rely primarily on expert opinion, our model represents a fully implemented and empirically validated solution. For example, Hol et al. (2023) proposed a curriculum matching model based on expert surveys, which introduced subjectivity and lacked quantitative validation, limiting scalability and objectivity. In contrast, our data-driven approach leverages labour market analytics combined with natural language processing and machine learning to produce objective metrics for curriculum alignment evaluation. This approach aligns with recent advances in AI-driven curriculum augmentation aimed at personalized learning and adaptive curriculum development (Sajja et al., 2024; Mariyono & Nur Alif Hd, 2025).

Similarly, Moldoveanu et al. (2022) highlighted the global digital skills gap through cross-sectional analyses but did not provide a dynamic model responsive to real-time labour market changes. Our findings build on this by quantifying specific skill gaps across syllabi. This supports calls for integrating emerging technology competencies in education (Al Farsi, 2023; Al Farsi et al., 2023a) and is consistent with frameworks for digital transformation skills (Mhaske et al., 2025). The importance of equitable digital skills alignment is emphasized by Zou et al. (2025) who argue for inclusive approaches to digital education to avoid worsening inequalities.

Furthermore, our model's focus on employability and curriculum relevance is aligned with evidence on persistent workforce challenges when education programs lag behind technological advancement (Ma'dan et al., 2020). Additional validation of AI and virtual reality technologies in education indicates their growing potential to enhance learning engagement and skill acquisition (Al Farsi et al., 2023b; Krishna Pasupuleti, 2024). These technologies support the movement towards more dynamic and learner-centered curricula, complementing our model's adaptive capabilities.

By integrating scalable AI tools with a continuous feedback loop from job market data, our framework not only identifies current skill deficits but positions itself as an innovative solution for ongoing curriculum updates in response to rapid digital transformation aligning with findings from studies by Kholifah et al., (2025) and UNESCO, (2025). This sets it apart from existing, largely static models, ensuring alignment with both industry demands and evolving digital competencies.

Table 6 summarizes key studies and initiatives related to AI-enabled curriculum alignment in higher education.

It highlights each study's focus, the methodological approach, and how it aligns or contrasts with the current project. The goal is to demonstrate the landscape of ongoing efforts and position our model within this active research field, emphasizing its empirical validation and real-time responsiveness.

[Table 6 to be placed here]

Table 6 Related Work on AI-Enhanced Curriculum Alignment in Higher Education

Study/Authors	Focus	Key Findings/Limitations	Relation to Current Project
Hol et al. (2023)	Curriculum matching based on expert surveys	Introduced subjectivity, lacked quantitative validation	Cited as an example of conceptual AI curriculum alignment; contrasted with empirical validation in manuscript
Moldoveanu et al. (2022)	Cross-sectional analysis of digital skills gap	Static, no dynamic model responsive to labour market changes	References global skills gap; manuscript quantifies specific gaps, expands dynamic responsiveness
Sajja et al. (2024), Mariyono & Nur Alif Hd (2025)	AI-driven curriculum augmentation and personalized learning	Advances in AI for adaptive curricula enhancing personalization	Manuscript aligns with these cutting-edge AI approaches for personalized and adaptive curriculum development
Al Farsi (2023; 2023a,b)	Emerging technology competencies in education	Advocates integrating virtual reality & AI for engagement	Supports manuscript calls for incorporating emerging tech competencies in curricula
Mhaske et al. (2025)	Digital transformation skills framework	Framework to guide digital skills integration	Consistent with manuscript framework for digital skills alignment
Zou et al. (2025)	Inclusive digital education	Emphasizes equitable digital skills alignment to prevent inequalities	Supports manuscript's emphasis on equitable skills alignment
Ma'dan et al. (2020)	Employability challenges from curricular lag	Highlights workforce challenges due to outdated curricula	Supports manuscript's focus on employability and curriculum relevance
Al Farsi et al. (2023b), Krishna Pasupuleti (2024)	AI and VR technologies enhancing learning engagement	Shows potential of VR and AI for learner engagement and skill acquisition	Complement manuscript's adaptive curriculum capabilities
Kholifah et al. (2025), UNESCO (2025)	Ongoing curriculum updates in digital transformation context	Advocate continuous curriculum adaptation with AI tools	Manuscript integrates continuous market feedback with scalable AI, aligning with these contemporary visions

6. Limitations and Future Work

While the model provides a strong foundation, several limitations affect scope and generalizability.

Geographic Context: Our implementation focused on Kenya and East Africa. Although many digital skills are globally relevant, local labour market conditions and regulatory contexts vary. Adapting the model to other regions (with different job markets or languages) would require re-scraping region-specific job data and possibly retraining the NLP and classification components.

Language and Culture: The current NLP pipeline supports only English (via spaCy's English model). This excludes non-English curricula or local dialects. Future work should integrate multilingual NLP capabilities to broaden applicability.

Data Bias and Coverage: Our data derives from online ICT job postings, which may omit informal sector positions or emerging start-up roles. Such biases could skew the skill database. We attempted to mitigate this by focusing on a broad spectrum of ICT occupations and cleaning data (e.g., normalizing terminology), but some bias likely remains. Addressing this fully would involve incorporating additional data sources (e.g., employer surveys or alternative job platforms).

Educational Scope: The model emphasizes employability-aligned competencies (a human-capital perspective), consistent with competency-based education theory. However, we recognize that universities also aim to cultivate broader values and critical thinking. It is important to note that competency-based frameworks can include such goals; for example, DigComp incorporates problem-solving and creative thinking as part of digital competence. Thus, our model is intended to complement traditional education goals, not replace them. Stakeholders should use the model's outputs alongside broader curricular aims.

Future enhancements include integrating transformer-based embeddings (e.g., BERT) for deeper semantic analysis, which may improve extraction accuracy at the cost of higher computational requirements. Ensemble learning (e.g., boosted trees) could be explored to potentially increase classification performance. We also plan longitudinal studies to measure the impact of model-guided curriculum changes on graduate outcomes over time.

Table 7 below presents key limitations and future directions for the intelligent curriculum alignment model developed in this study.

[Table 7 to be placed here]

Table 7 Limitations and Future Work in Curriculum Alignment Model

Limitation/Aspect	Description	Remarks/Future Work
Geographic Context	Model implementation focused on Kenya and East Africa; skills have global relevance but local markets and regulations vary.	Adaptation requires re-scraping region-specific job data and retraining NLP and classification components.
Language and Culture	NLP pipeline currently supports only English via spaCy; excludes non-English curricula and local dialects.	Future work should integrate multilingual NLP capabilities to broaden applicability.
Data Bias and Coverage	Data from online ICT job postings may omit informal sector and start-up roles, introducing bias despite mitigation efforts.	Incorporate additional data sources like employer surveys or alternative job platforms to address bias.
Educational Scope	Focus on employability-aligned competencies consistent with competency-based theory; universities also value broader critical skills.	Model intended to complement, not replace, traditional goals; use outputs alongside wider curricular aims.
Future Enhancements	Plans to integrate transformer-based embeddings (e.g., BERT) for deeper analysis, explore ensemble learning, and conduct longitudinal studies.	Aims to improve accuracy and evaluate impact of curriculum changes on graduate outcomes over time.

7. Conclusion and Recommendations

This study has developed and validated an intelligent model that bridges the persistent gap between academic programs and evolving digital labor market demands. By automating the analysis of real-time job market data and curriculum content, the model offers universities evidence-based insights and actionable recommendations to proactively update curricula. The strong performance metrics (accuracy 0.88, recall for key skills 0.99) provide confidence that the DSI and prescriptive reports are reliable tools for decision-making.

7.1 Synthesis of Findings Regarding Research Questions

1. RQ1 (Digital Skill Gaps): The study identified significant gaps in the integration of high-demand skills such as Cloud Computing, AI, and Cybersecurity within traditional curricula. The quantile analysis revealed that while universities often cover foundational skills, they frequently lag in adopting rapidly emerging High-Frequency skills (Class 1) identified by the real-time data.
2. RQ2 (Hybrid Model Design): We demonstrated that a hybrid architecture combining Web Scraping (for real-time data), NLP (for semantic extraction), and Decision Trees (for interpretable classification) effectively creates a dynamic alignment mechanism. This design successfully overcomes the static nature of traditional curriculum reviews.
3. RQ3 (Model Reliability): The model proved to be highly effective and reliable, achieving an 88% overall accuracy and an exceptional 99% recall for high-demand skills. This confirms that the AI-driven approach can accurately identify and prescribe necessary curricular updates with a high degree of confidence.

Recommendations

Based on the findings, we offer the following:

For Academic Institutions: Continuously integrate the high-demand digital skills identified by the model (e.g., Python, Cloud Computing, AI/ML, Cybersecurity) into core curricula. Foster interdisciplinary digital literacy by embedding these competencies across programs. Provide faculty development to build instructors' capacity in emerging technologies. Actively use the Prescriptive Analytics Reports to guide curriculum revisions.

For Industry: Engage systematically in curriculum design by providing up-to-date input on required skills and offering practical experiences (internships, project collaborations) that align with curricular content. Share anonymized job and skill data to help refine the model's database.

For Policymakers: Develop national digital skills strategies that align educational policy with market needs, using models like this one as diagnostic tools. Invest in digital infrastructure and training for educators. Encourage academia–industry partnerships through funding and incentives.

We emphasize that high-frequency skills (Class 1) should be prioritized, but low-frequency skills (Class 0) are also valuable for adaptability. For instance, while the model highlights Python and data analysis as crucial, policymakers should also promote broader digital literacies (critical thinking, ethical AI) to ensure graduates can navigate unforeseen future demands.

By providing a hybrid AI-driven framework, this research represents a paradigm shift toward continuous, data-informed curriculum optimization. We envision that widespread adoption of such tools will help higher education remain agile and relevant in the fast-changing digital era.

A summary of the recommendations is provided in table 8 below.

[Table 8 to be placed here]

Table 8: Recommendations for Bridging Academic-Industry Digital Skills Gaps

Recommendation Category	Specific Recommendations	Target Audience
Academic Institutions	- Continuously integrate high-demand digital skills (e.g., Python, Cloud Computing, AI/ML, Cybersecurity). - Foster interdisciplinary digital literacy across programs. - Provide faculty development for emerging technologies. - Use Prescriptive Analytics Reports to guide curriculum revisions.	Academic Institutions
Industry	- Engage systematically in curriculum design with up-to-date skill input. - Offer practical experiences aligning with curricular content (internships, collaborations). - Share anonymized job and skill data to refine skill databases.	Industry
Policymakers	- Develop national digital skills strategies aligned with market needs. - Invest in digital infrastructure and educator training. - Encourage academia-industry partnerships via funding and incentives.	Policymakers
General	- Prioritize high-frequency skills but recognize low-frequency skills' adaptability value. - Promote broader digital literacies including critical thinking and ethical AI use.	All

Acknowledgements
Not applicable

Author contributions
DN provided a concise overview of the existing literature, formulated the methodology, collated and analyzed data. SM AK and FM helped refine the content. AS helped refine the methodology, collection and analysis of data. All five authors collaborated on the selection of the final paper collection. The final version of the paper received approval from all authors.

Funding
The authors received no funding for this research work.

Data Availability

The data and materials used in this paper are available upon request. The comprehensive list of included studies, along with relevant data extracted from these studies, is available from the corresponding author upon request.

Declarations

Ethics approval

A research ethical clearance was obtained from the KCA University Scientific Ethics Review Committee (KCAUSERC), with the approval code KCAUSERC/ EXTER013. The research was carried out following the guidelines of the KCAUSERC.

Consent to participate

This study did not involve human participants. Data was web scraped exclusively from publicly accessible websites. No personal data was collected, ensuring the privacy and anonymity of individuals. As such, informed consent from individuals is not applicable to this research.

Consent for publication

No human participants were involved in this study. The web scraped data was only used for scholarly and educational purposes, in accordance with ethical research guidelines and institutional policies. The paper's publication is permitted by all of the authors.

Competing interests

The authors declare no competing interests.

Clinical trial number

Not applicable

References

Al Farsi, G., bin Mohd. Yusof, A., Bin Rusli, M. E., & AlSinani, M. (2023a). The Current Perceptions About Instructional Tools in Educational Towards Adoption of Virtual Reality Among Undergraduate Students. International Journal of Interactive Mobile Technologies (iJIM), 17(08), pp. 169–180. https://doi.org/10.3991/ijim.v17i08.3693

Al Farsi, G., Yusof, A. B. M., Rusli, M. E. B., & AlSinani, M. (2023b). Virtual reality environments with the education sector. In M. Pandit, M. K. Gaur, & S. Kumar (Eds.), Artificial intelligence and sustainable computing: ICSISCET 2022 (Algorithms for Intelligent Systems). Springer. https://doi.org/10.1007/978-981-99-1431-9_49

Al Farsi, G. (2023). The efficiency of UTAUT2 model in predicting student's acceptance of using virtual reality technology. *International Journal of Interactive Mobile Technologies (iJIM), 17*(12), 17-27. https://doi.org/10.3991/ijim.v17i12.36951

Crompton, H. (2023). Artificial Intelligence in Higher Education: The State of the Field. Int J Educ Technol High Educ, 1(20). https://doi.org/10.1186/s41239-023-00392-8

Harry, S. C. (2021). Four Strategies for Aligning the Curriculum with Technology. www.linkedin.com. https://www.linkedin.com/pulse/four-strategies-aligning-curriculum-technology-dr-harry-saint-cyr

Hol, A., Richardson, J., McGovern, J., & Hamilton, M. (2023). A New Sustainable Model for Aligning Industry Requirements and University Programs. ACM Inroads, 14(1), 30–39. https://doi.org/10.1145/3583086

ILO. (2021) Changing demand for skills in digital economies and societies: Literature review and case studies from low- and middle-income countries. ILO Publications. https://www.ilo.org/skills/areas/skills-training-for-poverty-reduction/WCMS_831372/lang--en/index.htm

ITEA. (2007). Standards for technological literacy: Content for the study of technology. *International Technology Education Association*. https://wwwcp.umes.edu/tech/wp-content/uploads/sites/94/2021/09/xstnd.pdf

ITU. (2021). Skills Development for the digital economy. ITU Publications. https://www.itu.int/en/mediacentre/backgrounders/Pages/skills-development-digital-economy.aspx

Kholifah, N., Nurtanto, M., Sutrisno, V. L., Majid, N. W., Subakti, H., Daryono, R. W., & Achmadi, A. (2025). Unlocking workforce readiness through digital employability skills in vocational education graduates: A PLS-SEM analysis based on human capital theory. *Social Sciences & Humanities Open*, *11*, 101625. https://doi.org/10.1016/j.ssaho.2025.101625

Kim, J., Lee, H., Cho, Y. M. (2022). Learning Design to Support Student-AI Collaboration: Perspectives of Leading Teachers for AI in Education. Educ Inf Technol, 5(27), 6069-6104. https://doi.org/10.1007/s10639-021-10831-6

Kim, J. (2015). Competency-based curriculum: An effective approach to digital Curation education. *Journal of Education for Library and Information Science Online*, *56*(4), 283-297. https://doi.org/10.3138/jelis.56.4.2

Krishna Pasupuleti, M. (2024). AI-enhanced virtual reality: Transforming immersive training and simulation for real-world skill development. *AI-Powered Virtual Reality: Enhancing Immersive Experiences for Training and Simulation*, 96-113. https://doi.org/10.62311/nesx/66241

Ma'dan, M., Ismail, M. T., & Daud, S. (2020). Strategies to Enhance Graduate Employability: Insights from Malaysian Public University Policy-Makers. Malaysian Journal of Learning and Instruction, 17(2), 137–165. https://doi.org/10.32890/mjli2020.17.2.

Mariyono, D., & Nur Alif Hd, A. (2025). AI's role in transforming learning environments: A review of collaborative approaches and innovations. *Quality Education for All*, *2*(1), 265-288. https://doi.org/10.1108/qea-08-2024-0071

Mhaske, P., Bhattacharjee, B., Haldar, N., Upadhyay, P., & Mandal, A. (2025). Bridging digital skill gaps in the global workforce: A synthesis and conceptual framework building. *Research in Globalization*, *11*, 100311. https://doi.org/10.1016/j.resglo.2025.100311

Moldoveanu, M., Frey, K., & Moritz, B. (2022). 4 Ways to Bridge the Global Skills Gap. Harvard Business Review. https://hbr.org/2022/03/4-ways-to-bridge-the-global-skills-gap

Plekhanov, D., Franke, H., Netland, T. H. (2022). Digital Transformation: A Review and Research Agenda. European Management Journal. https://doi.org/10.1016/j.emj.2022.09.007

Rovira, S., Peres, W., & Saporito, N. (2021). Digital technologies for a new future. United Nations, Economic Commission for Latin America and the Caribbean. https://www.cepal.org/sites/default/files/publication/files/46817/S2000960_en.pdf

Sajja, R., Sermet, Y., Cikmaz, M., Cwiertny, D., & Demir, I. (2024). Artificial Intelligence-Enabled Intelligent Assistant for Personalized and Adaptive Learning in Higher Education. *Information*, *15*(10), 596. https://doi.org/10.3390/info15100596

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-world Applications and Sarker Research Directions. SN COMPUT. SCI., 3(2). https://doi.org/10.1007/s42979-021-00592-x

Thongprasit, J., Wannapiroon, P. (2022). Framework of Artificial Intelligence Learning Platform for Education. IES, 1(15), 76. https://doi.org/10.5539/ies.v15n1p76

UNESCO. (2025). Artificial intelligence in education: Challenges and opportunities. UNESCO. https://www.unesco.org/en/digital-education/artificial-intelligence

Vuorikari, R., Kluzer, S., & Punie, Y. (2022). DigComp 2.2: The digital competence framework for citizens—With new examples of knowledge, skills and attitudes (EUR 31006 EN). *Publications Office of the European Union*. https://doi.org/10.2760/115376

World Economic Forum. (2023). The Future of Jobs Report 2023. World Economic Forum. https://www.weforum.org/reports/the-future-of-jobs-report-2023/digest/

Zou, Y., Kuek, F., Feng, W., & Cheng, X. (2025). undefined. *Frontiers in Education*, *10*. https://doi.org/10.3389/feduc.2025.1562391