DOI: 10.37943/ILRF5900

Ainur Mukashova
PhD, Assistant Professor, School of Artificial Intelligence and Data Science
A.Mukashova@astanait.edu.kz, orcid.org/0009-0007-2794-4748
Astana IT University, Kazakhstan
Muslim Sergaziyev
PhD, Associate Professor, School of Artificial Intelligence and Data Science
muslim.sergaziyev@astanait.edu.kz, orcid.org/0009-0004-5332-3441
Astana IT University, Kazakhstan.
Mukhanova Ayagoz
PhD, Associate Professor, Department of Information Systems
ayagoz198302@mail.ru, orcid.org/0000-0003-3987-0938
L.N. Gumilyov Eurasian National University, Kazakhstan
Zhanat Kenzhebayeva
Candidate of Technical Sciences, Associate Professor,
Kazakh-German Institute of Sustainable Engineering,
zhanat.kenzhebayeva@yu.edu.kz, orcid.org/0000-0002-1942-4474
Sh. Yesenov Caspian University of Technology and Engineering, Kazakhstan
Ainur Shekerbek
PhD, Senior Lecturer, Department of Information Systems
Shekerbekainur80@gmail.com, orcid.org/0000-0002-1088-4239
L.N.Gumilyov Eurasian National University, Kazakhstan.
Aidana Akhmetova
PhD, Senior Lecturer, Department of Information Systems
akhmetova_azh@mail.ru, orcid.org/0000-0002-3876-1089
L.N. Gumilyov Eurasian National University, Kazakhstan

ENHANCING EDUCATIONAL COMPETENCY FRAMEWORKS IN KAZAKHSTAN THROUGH A HYBRID BLOOM'S TAXONOMY AND GENERATIVE AI APPROACH WITH PCA-BASED VALIDATION

Abstract: Competency-based education frameworks require clear differentiation of learning outcomes across qualification levels to support effective curriculum design and ensure alignment with dynamic labor market requirements. However, national professional standards frequently contain overlapping, ambiguous, or weakly differentiated competency statements, which complicates the systematic development of educational programs. This study proposes a hybrid methodology that integrates the hierarchical framework of Bloom's Taxonomy with generative artificial intelligence to improve the structural differentiation of competency statements within Kazakhstan's professional standards.

A comprehensive dataset of 595 professional standards containing 8,736 competency records was compiled and analyzed using natural language processing techniques. The textual competencies were vectorized utilizing TF-IDF and examined through Principal Component Analysis (PCA) to assess structural differentiation across qualification levels. To address existing gaps and enrich the dataset, 4,024 additional competencies for advanced qualification levels 6-8 were generated using GPT-4. This process was guided by Bloom's Taxonomy action verbs to accurately reflect progressive cognitive complexity and depth of knowledge.

Comparative analysis conducted before and after data augmentation showed a moderate increase in explained variance in the first principal components, rising from approximately 4.5% to 7.5%. This shift suggests improved structural differentiation within the augmented competency dataset. Additionally, classification experiments using multiple machine-learning models

Received: 23.02.2026 Accepted: 17.07.2026 Published: 30.06.2026

demonstrated notably enhanced performance metrics after augmentation, indicating greater separability of competency statements across the targeted qualification levels.

The findings suggest that combining established pedagogical taxonomies with generative AI can efficiently support the systematic refinement of competency frameworks. Rather than replacing expert curriculum designers, generative AI functions as a powerful analytical support tool for identifying structural gaps and improving competency differentiation within large collections of educational standards.

Keywords: generative artificial intelligence; educational standards; competencies; Bloom’s taxonomy; principal component analysis (PCA); machine learning; competency-based education.

Introduction

In the Republic of Kazakhstan (RK), professional standards are key regulatory documents that define the competencies required at different qualification levels. These standards establish the knowledge, skills, and abilities necessary for the design and evaluation of educational programs and play a central role in aligning higher education with labor market needs [1-2]. However, existing professional standards often exhibit structural inconsistencies, including overlapping competency formulations and insufficient differentiation between qualification levels.

A popular dimensionality reduction method in data science, principal component analysis (PCA) finds patterns and variations in high-dimensional data, such as text-based competencies. By vectorizing textual descriptions, PCA can reveal how well competencies cluster or differentiate by levels. Principal Component Analysis (PCA) is recognized as one of the most effective methods for dimensionality reduction and uncovering hidden structures in complex datasets. The study [3-6] demonstrated that the application of PCA in text-based classification tasks can help identify structural patterns in high-dimensional textual data and support exploratory analysis of competency frameworks. Reduces computational costs and identifies the most informative features. For the analysis of professional standards and competency generation, this approach also proves valuable, as it helps to identify key variations and optimize the performance of generative AI algorithms.

The development and assessment of learning outcomes occupy a central position in curriculum design, teaching methodologies, and evaluation practices in higher education, providing institutions with a foundation for documenting and enhancing student learning. Learning outcomes clearly define what learners are expected to know and be able to do upon completion of a course or program. This clarity enables educational institutions to align curricula with workforce demands, thereby ensuring that graduates are well-prepared for the complexities of modern careers [7-8].

A fundamental framework that is frequently used in the creation of curricula, learning objectives, and instructional goals across disciplines is Bloom's Taxonomy of Educational Objectives. Remember, understand, apply, analyze, evaluate, and create are its hierarchical stages, which provide an organized route for the development of cognitive abilities [9-11]. Shaw and Holmes [12] contend that Bloom's lasting relevance stems from its capacity to promote critical thinking rather than only specify educational goals, despite the fact that it was first designed for general teaching and assessment. Bloom's framework supports a balanced approach to the development of cognitive, affective, and linguistic skills in language education by helping to categorize instructional goals. In nations like Nepal, where indigenous curriculum must accommodate a variety of educational demands, this is especially crucial [13].

The studies [14-17] conclude that the integration of generative AI tools into education can improve critical thinking in cognitive, affective and metacognitive domains, indicating a complex relationship between the use of AI and skill development.

The use of AI within Bloom's Taxonomy helps educators structure learning sequences that foster independence and key competencies [18]. Waladi and Lamarti [17] analyze how artificial intelligence can be applied in the design of adaptive assessment systems within the framework of competence-based learning. Lubbe et al. [19] note that the integration of AI and Bloom's taxonomy creates new cognitive levels and strategies for assessing knowledge.

Despite the growing interest in artificial intelligence in education, relatively few studies have examined how generative AI can be systematically integrated with established pedagogical frameworks for improving competency structures in professional standards. In particular, there remains a lack of empirical studies evaluating whether AI-generated competencies contribute to measurable improvements in the structural differentiation of competency datasets.

This study aims to:

1. Analyze the structural differentiation of competencies in Kazakhstan’s professional standards using Principal Component Analysis applied to vectorized textual representations.
2. Generate additional competency statements for qualification levels 6, 7, 8 using a generative AI model constrained by Bloom’s Taxonomy action verbs.
3. Evaluate whether AI-assisted augmentation improves the structural separability of competency statements using dimensionality reduction and machine-learning classification.

Methods and Materials

A multi-stage methodological framework that included data formalization, text vectorization, dimensionality reduction, and classification model creation was created in order to accomplish the research objectives.

Dataset Compilation

Using the Republic of Kazakhstan's (RK) professional standards, which are normative documents outlining requirements for professional activities, knowledge, abilities, and competencies, a comprehensive dataset was created as the empirical basis of this study. These standards were obtained from the official platform and updated following revisions introduced in 2023. The dataset covers a wide range of professional domains, including education, healthcare, IT, and engineering, and includes 595 professional standards across 17 training directions.

After data structuring, a unified table consisting of 8,736 records and 14 attributes was created, where each row represents a single competency or learning outcome record extracted from a professional standard. Key attributes include qualification level, learning outcomes, knowledge, and associated skills. To ensure consistency for further analysis, the dataset was standardized in Russian, which is the primary language of RK professional standards [20]. Table 1 summarizes the main characteristics of the dataset used in this study.

Table 1. Descriptive statistics of the competency dataset

Dataset characteristic	**Value**
Professional standards analyzed	595
Training directions	17
Initial competency records (LOCs)	8736
Generated competencies	4024
Total dataset size after augmentation	12760
Qualification levels represented	1 – 8
Qualification levels augmented	6 – 8
Dataset language	Russian

The overall dataset structure and processing pipeline are illustrated in Figure 1.

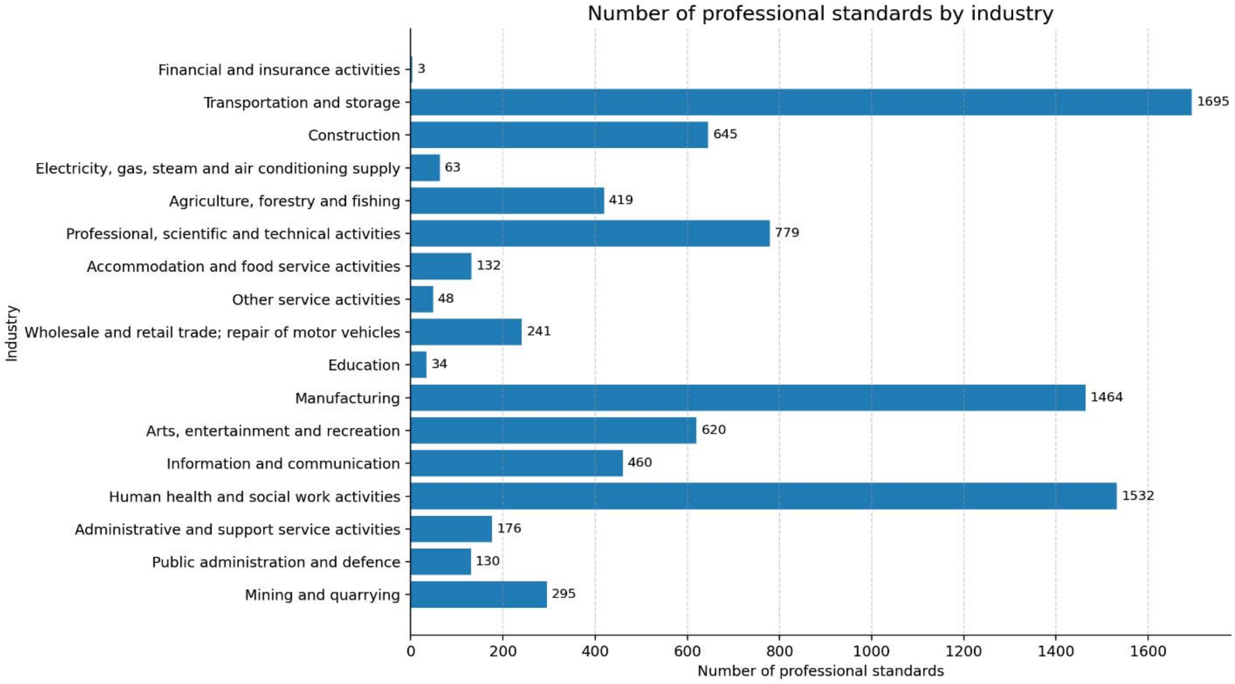

Figure 1. Structure of the Dataset Compiled from RK Professional Standards

Generation of New Competencies

The Republic of Kazakhstan's (RK) professional standards, which were drawn from the 595 standards updated in 2023, included comprehensive explanations of functional responsibilities, knowledge, skills, and abilities. A transformer-based large language model, specifically a GPT architecture via the OpenAI API, was used to extract, formalize, and refine competency requirements from unstructured textual sources for the autonomous development of learning outcomes. The generation parameters were configured to ensure a balance between creativity, coherence, and strict adherence to professional standards: the temperature was set to 1.0 to generate diverse yet relevant competency formulations, the maximum tokens limit was strictly set to 300 to ensure concise outputs and prevent overly verbose content, and Top-p sampling was employed to control the probability distribution of word selection. The model was invoked with targeted system instructions to process the textual input and automatically generate structured competency statements.

Bloom's Taxonomy verbs [10] were incorporated as a restricting mechanism to guarantee structural consistency and cognitive progression in the generated content. With particular distinction across certification levels 6, 7, and 8, this method made it easier to develop competencies in line with RK educational requirements. The verbs were chosen to represent increasing cognitive demands:

- Level 6: 2,063 competences with an emphasis on analyze (n = 12), evaluate (n = 8), and compare (n = 5).
- Level 7: 1,728 competences are produced by concentrating on design (n = 15), integration (n = 7), and appraisal (n = 10).
- Level 8: Creating 233 capabilities by giving priority to invent (n = 18), lead (n = 12), and pioneer (n = 5).

In order to improve distinction and scalability for higher-level certifications, prompt engineering that was adapted to RK sectoral contexts (such as IT and healthcare) made sure that the developed competences remained relevant and followed the taxonomic hierarchy. The dataset for additional study was greatly expanded, with 4,024 newly created skills in total.

Text Vectorization

Using the scikit-learn toolkit, the texts from the dataset underwent Term Frequency-Inverse Document Frequency (TF-IDF) vectorization to transform qualitative descriptors into a numerical feature space appropriate for multivariate analysis. This procedure allowed for the quantitative evaluation of competency distributions by converting the textual data into an organized manner. The 8,736 records in the dataset and the resultant word vocabulary were reflected in the feature space's dimensionality of 8,736 × 3,142.

I. Pipeline for Preprocessing

To improve the quality and relevance of the data, a strict preprocessing pipeline was put in place:

1. Tokenization and Lowercasing: To guarantee consistency and lessen case sensitivity, texts were divided into discrete tokens (words or phrases) using the nltk.word_tokenize function. All characters were then transformed to lowercase.

2. Stop-Word Removal: To reduce noise in the vector space, common Russian stop-words were eliminated using the Russian National Corpus (RNC) stop-word list, which was then augmented with domain-specific items unrelated to competency analysis (such as articles and prepositions).

3. Lemmatization: To normalize morphological variations and reduce lexical redundancy, words were lemmatized using the Pymorphy3 morphological analyzer for the Russian language. Lemmatization converts words to their base dictionary forms (lemmas), allowing different grammatical forms of the same word (e.g., анализировать, анализировал, анализирует) to be treated as a single lexical unit.

4. Term Filtering: In order to maximize the vocabulary size and concentrate on informative features, the parameters Min_df=2 (minimum document frequency of 2) and Max_df=0.95 (maximum document frequency of 95%) were selected to reject uncommon phrases (such as mistakes) and excessively common terms (such as "профессиональный").
5. Normalization: To guarantee that terms contribute equally across texts and account for different text lengths, TF-IDF weights were normalized using the L2 norm.

II. Vectorization Outcome

A word-document matrix with dimensions of 8,736 × 3,142 was produced as a result of the preprocessing; each row represented a competency record, and each column belonged to a distinct term weighted by its TF-IDF score. This matrix captured the semantic structure of competencies across RK standards and was used as the input for further PCA analysis.

Mathematical Formulation of Competency Generation

Formally, a professional standard is represented as a structured textual dataset:

$$PS = \{TF, K, S, A\} \qquad (1)$$

where:
$TF = \{tf_1, tf_2, \dots, tf_n\}$ is the set of task functions,
$K = \{k_1, k_2, \dots, k_m\}$ is the set of knowledge elements,
$S = \{s_1, s_2, \dots, s_p\}$ is the set of skills,
$A = \{a_1, a_2, \dots, a_q\}$ is the set of abilities.

All components are extracted from officially published professional standards using rule-based parsing and expert-validated structural markers (sections «Labor Functions», «Knowledge», «Skills», «Abilities»).

Text Representation and Vectorization

Each textual component is preprocessed using standard NLP procedures (tokenization, lemmatization, stop-word removal). The aggregated semantic representation is defined as:

$$V = f(TF, K, S, A) \qquad (2)$$

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

where $f(\cdot)$ denotes a text vectorization function implemented using TF-IDF.

Competency Generation Model

Competency generation is formulated as a conditional text generation task:

$$C = G(B) \qquad (3)$$

where:
- $G(\cdot)$ is a generative language model (LLM)
- $B = \{b_1, \dots, b_r\}$ is a controlled vocabulary of Bloom's Taxonomy verbs.

Bloom's verbs are injected into the generation prompt as hard lexical constraints, ensuring alignment with cognitive complexity levels.

Probabilistic Formulation

The optimal competency formulation is obtained by maximizing the conditional likelihood:

$$C^* = (TF, K, A, B) \qquad (4)$$

where $P(\cdot)$ is approximated by the token-level likelihood produced by the generative model. Multiple candidate competencies are generated using nucleus sampling, after which the best candidate is selected based on semantic similarity between c and V, measured via cosine similarity.

Qualification Level Classification

To assign a qualification level, a mapping function is introduced:

$$L: C \to \{6,7,8\} \qquad (5)$$

where levels correspond to Bachelor, Master and Doctoral qualifications.

The function $L(\cdot)$ is implemented as a supervised classifier trained on labeled competency statements, using: Bloom's verb categories, syntactic complexity, semantic embedding features.

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) was used as a dimensionality reduction approach to examine the dataset's internal structure and visualize its multidimensional space. The scikit-learn toolkit (Pedregosa et al., 2011, version 1.3.0) was used to implement PCA, which preserved the maximum variance while converting the 8,736 × 3,142 TF-IDF vectorized matrix into a lower-dimensional representation. With an emphasis on the first two principal components (PC1 and PC2) for visualization and comprehension, the analysis was set up with n_components=50 to capture the predominant patterns.

The procedure included:

1. Feature scaling: Because TF-IDF matrices are sparse, feature scaling was applied using StandardScaler(with_mean=False) to preserve the sparse structure of the matrix. This approach prevents densification of the TF-IDF matrix and ensures compatibility with subsequent dimensionality reduction procedures.

2. Eigenvalue decomposition: Principal component analysis (PCA) identified the principal components explaining the largest variance by calculating eigenvectors.

3. Dimensionality reduction: By retaining the 50 most important components, computational complexity was reduced without losing important structural information.

With special emphasis on levels 6-8 post-augmentation, this approach made it possible to identify competency clusters corresponding to various qualification levels (1-8). The analysis of underlying patterns was made easier by visualizing PC1 and PC2 using 2D scatter plots (created using Matplotlib), which highlighted overlaps and intersections between classes across certification levels. To measure the percentage of total variance captured, the explained variance

DOI: 10.37943/ILRF5900

ratio for each component was computed. This ratio was used as a metric to evaluate differentiation both before and after the introduction of created competences.

Classification

In order to enable the automatic determination of qualification levels for learning outcomes and competences (LOCs), a number of machine learning models were created and assessed at the end of the investigation. The scikit-learn package was used to implement the models, which included Multinomial Naive Bayes (MultinomialNB), Gaussian Naive Bayes (GaussianNB), Multinomial Logistic Regression, and Decision Tree. The pre-augmentation dataset, which included 8,736 original LOCs from RK professional standards, and the post-augmentation dataset, which increased to 12,760 LOCs with the addition of 4,024 new competences, were used to train and test these models.

Results

This section presents the results of a comparative analysis between the original competencies documented in the educational and professional standards of the Republic of Kazakhstan and the competencies generated using Generative Artificial Intelligence (GenAI) guided by Bloom's Taxonomy verbs. To uncover the latent structure of the texts and assess the degree of differentiation, the Principal Component Analysis (PCA) method was applied

Analysis of Original Competencies

At the first stage, the competency statements were vectorized using the TF-IDF model. The resulting vectors were analyzed through PCA, allowing for the visualization of the distribution of competencies within a two-dimensional space.

The results are shown in Figure 2. As observed, the explained variance for the first two principal components was relatively low: PC1 = 2.61% and PC2 = 1.87% (a total of approximately 4.48%). This distribution indicates a considerable overlap among clusters of competencies across qualification levels. In other words, the original formulations exhibit a high degree of similarity, which complicates the clear differentiation between levels 1–8. This effect is particularly pronounced for levels 6–8, where dense clustering and point overlap are evident.

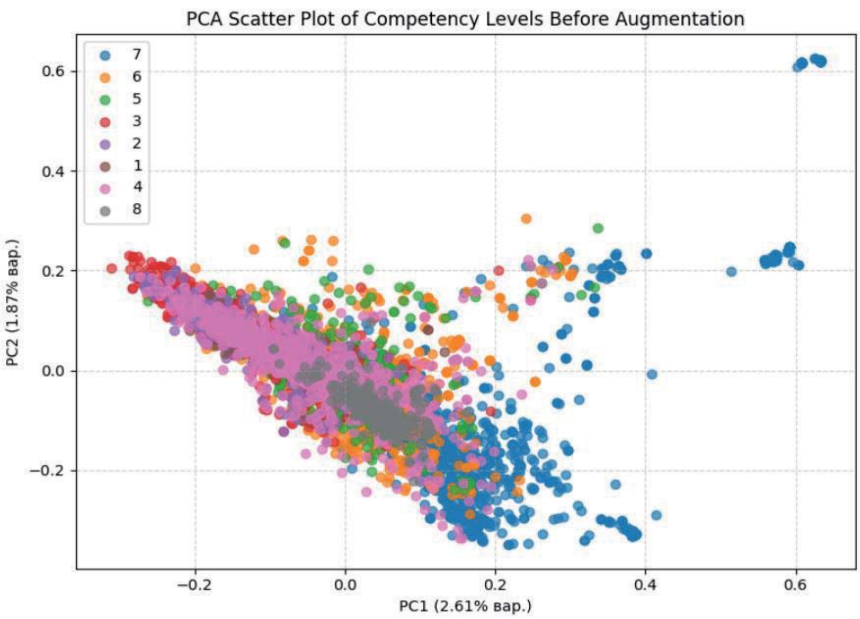

Figure 2. PCA scatter plot of competency levels before augmentation

Thus, the baseline analysis confirmed the hypothesis regarding the insufficient structural organization of the original competencies within the professional standards.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

Generation of New Competencies for Levels 6-8

At the next stage, new competencies were generated for qualification levels 6 (Bachelor), 7 (Master), and 8 (Doctoral). During the generation process, Generative Artificial Intelligence (GenAI) was employed with specific constraints on the selection of verbs that strictly correspond to the cognitive levels defined in Bloom's Taxonomy. This approach enabled the creation of learning outcomes that reflect the required depth of cognitive processes - ranging from analysis and synthesis to evaluation and creation.

Repeated PCA Analysis

After generation, the competencies were re-vectorized and reanalyzed using the Principal Component Analysis (PCA) method. The results are presented in Figure 3. In contrast to the original dataset, the explained variance increased significantly: PC1 = 5.19% and PC2 = 2.31% (a total of approximately 7.50%).

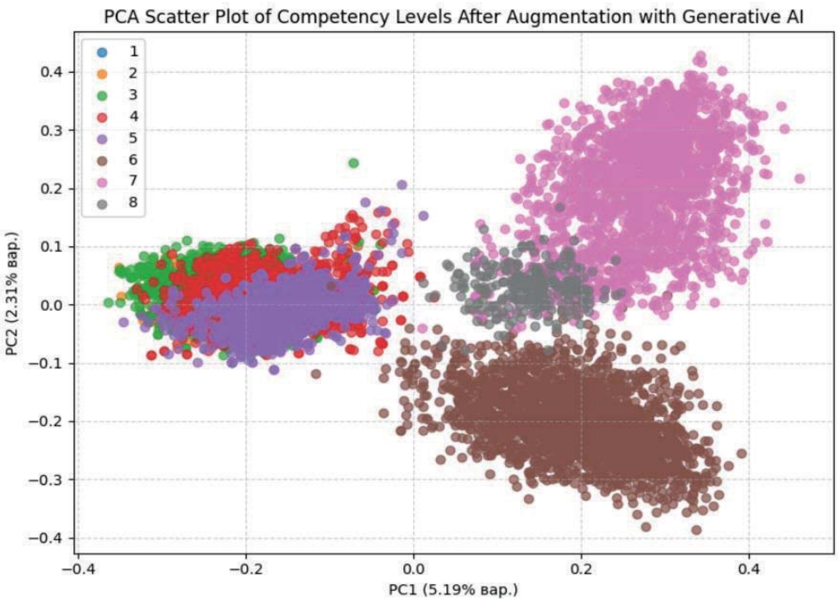

Figure 3. PCA scatter plot of competency levels after augmentation with generative AI

Furthermore, the visualization revealed a clearer separation of competencies by qualification levels. The clusters corresponding to levels 6, 7, and 8 became more distinct, with fewer overlaps between them. This indicates that the generated competencies possess greater semantic variability and more accurately reflect the specificity of each qualification level.

Comparative Analysis

A comparison between the original and generated competencies led to the following conclusions:

- The increase in explained variance indicates that the new competencies are more diverse and structurally organized.
- The distinct formation of clusters confirms an improvement in the quality of level differentiation.
- The use of Bloom's Taxonomy verbs ensures the cognitive hierarchy of learning outcomes, thereby strengthening the methodological rigor of educational standards.

Thus, the results of the study confirm the effectiveness of applying Generative AI for the enhancement of educational and professional standards.

To evaluate the performance of the classification models, the following metrics were used: Precision, Recall, F1-score, and Support (see Table 2). These metrics provide a comprehensive assessment of prediction accuracy, class coverage, and overall balance of results.

Table 2. Classification results by qualification levels (1-8) for different models

Model	**Class**	**Pre-Augmentation**				**Post-Augmentation (with GenAI)**			
		Precision	**Recall**	**F1-Score**	**Support**	**Precision**	**Recall**	**F1-Score**	**Support**
	1	0,578947	0,366667	0,44898	30	0,296296	0,228571	0,258065	35
	2	0,361111	0,425234	0,390558	214	0,404959	0,396761	0,400818	247
	3	0,544699	0,540206	0,542443	485	0,560175	0,538947	0,549356	475
	4	0,404011	0,374005	0,38843	377	0,493606	0,546742	0,518817	353
	5	0,39576	0,374582	0,38488	299	0,640777	0,626582	0,6336	316
	6	0,594556	0,653543	0,622656	635	0.998172	0.996352	0.999683	626
Decision Tree	7	0,834025	0,77907	0,805611	516	0.995236	0.99987	0.999412	495
	8	0,950425	0,9	0,947368	80	0.998745	0.99941	0.999123	74
	1	1	0,033333	0,064516	30	1	0,114286	0,205128	35
	2	0,441379	0,299065	0,356546	214	0,513514	0,230769	0,318436	247
	3	0,582343	0,707216	0,638734	485	0,585406	0,743158	0,654917	475
	4	0,552239	0,392573	0,458915	377	0,561983	0,577904	0,569832	353
	5	0,540284	0,381271	0,447059	299	0,658046	0,724684	0,689759	316
	6	0,578947	0,866142	0,694006	635	0,9984	0,996805	0,997602	626
	7	0,890909	0,75969	0,820084	516	0,99798	0,99798	0,99798	495
Logistic Regression	8	0,96875	0,3875	0,553571	80	0,98564	0,972973	0,986301	74
	1	1	0,033333	0,064516	30	1	0,114286	0,205128	35
	2	0,441379	0,299065	0,356546	214	0,513514	0,230769	0,318436	247
	3	0,582343	0,707216	0,638734	485	0,585406	0,743158	0,654917	475
	4	0,552239	0,392573	0,458915	377	0,561983	0,577904	0,569832	353
	5	0,540284	0,381271	0,447059	299	0,658046	0,724684	0,689759	316
	6	0,578947	0,866142	0,694006	635	0,9984	0,996805	0,997602	626
Multinomial NB	7	0,890909	0,75969	0,820084	516	0,99798	0,99798	0,99798	495

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

	8	0,96875	0,3875	0,553571	80	0,97856	0,972973	0,986301	74
Gaussian Naive Bayes	1	1	0,033333	0,064516	30	1	0,114286	0,205128	35
	2	0,441379	0,299065	0,356546	214	0,513514	0,230769	0,318436	247
	3	0,582343	0,707216	0,638734	485	0,585406	0,743158	0,654917	475
	4	0,552239	0,392573	0,458915	377	0,561983	0,577904	0,569832	353
	5	0,540284	0,381271	0,447059	299	0,658046	0,724684	0,689759	316
	6	0,578947	0,866142	0,694006	635	0,9984	0,996805	0,997602	626
	7	0,890909	0,75969	0,820084	516	0,99798	0,99798	0,99798	495
	8	0,96875	0,3875	0,553571	80	0,99652	0,972973	0,986301	74

The results of the classification by qualification levels (1-8) for different models are presented in Table 2.

The analysis of the obtained results demonstrates the following:

- High-level classes (Levels 6-8) demonstrate very high classification performance across all models, with F1-scores approaching 0.99. This result suggests that competency statements at these qualification levels contain distinctive lexical and semantic features that facilitate model discrimination. However, this effect may also be partially influenced by the structured use of Bloom's Taxonomy verbs during competency generation, which introduces clearer cognitive differentiation between levels.

To reduce the risk of data leakage, the training and testing split (80/20) was performed prior to the augmentation process. Generated competencies derived from training examples were excluded from the test dataset. This procedure ensured that the classification models were evaluated on unseen competency statements.

- Mid-level classes (Levels 3-5) show stable results, particularly for Logistic Regression and Naive Bayes models, with F1-scores ranging from 0.65 to 0.69. The Decision Tree model performs less consistently, yielding lower scores (e.g., F1-score = 0.39 for Level 2).
- Low-level classes (Levels 1-2) proved to be the most challenging for classification. These classes exhibit high precision but low recall, indicating difficulties in correctly identifying all instances within these categories. For example, in Level 1, both Logistic Regression and Naive Bayes models achieve Precision = 1.0, but Recall = 0.11.

Consequently, it can be concluded that Logistic Regression and Gaussian Naive Bayes provide the most balanced results across the majority of classes.

Figure 4 illustrates a comparative analysis of the performance of various classification algorithms based on the F1-score metric for each competency level (1-8).

The graph clearly shows that Logistic Regression and Gaussian Naive Bayes models demonstrate higher and more stable performance across most classes, particularly from level 3 and above.

In contrast, the Decision Tree model performs significantly worse for the lower classes (levels 1-2), which supports the conclusions drawn from Table 2. For levels 6-8, all models achieve near-perfect F1-score values, indicating a well-defined structural representation of features within the competencies of these qualification levels. Table 3 presents a comparative analysis of the performance metrics for four machine learning models - Multinomial Naive Bayes, Gaussian

Naive Bayes, Multinomial Logistic Regression, and Decision Tree - before and after the application of Generative AI (GenAI) to the competency dataset.

Table 3. Comparative performance of classification models before and after integration of Generative AI (GenAI)

Model	**Pre-Augmentation**				**Post-Augmentation (with GenAI)**			
	Accuracy	**F1-Score**	**Precision**	**Recall**	**Accuracy**	**F1-Score**	**Precision**	**Recall**
Multinomial NB	0,545144	0,455979	0,593074	0,545144	0,679893	0,613336	0,599992	0,679893
Gaussian Naive Bayes	0,597876	0,59558	0,604545	0,597876	0,690576	0,687645	0,690974	0,690576
Multinomial Logistic Regression	0,623293	0,60806	0,637887	0,623293	0,777184	0,765936	0,778034	0,777184
Decision Tree	0,57132	0,572277	0,575633	0,57132	0,734452	0,734335	0,734363	0,734452

The results demonstrate a consistent improvement across all models after the integration of GenAI-enhanced competencies. The Multinomial Logistic Regression model achieved the highest overall performance, with an Accuracy of 0.777, F1-score of 0.766, and Precision of 0.778, indicating superior balance between precision and recall. The Decision Tree model also showed an increase in accuracy from 0.571 to 0.734 after the inclusion of generated competencies, indicating improved structural organization of competency statements within the dataset.

Both Naive Bayes models exhibited moderate but stable improvements, confirming the positive impact of semantic enrichment and the inclusion of Bloom's Taxonomy verbs during generation. Overall, the post-generation results highlight the potential of Generative AI to enhance the semantic diversity, consistency, and separability of competency formulations, thereby improving the performance of classification algorithms in educational data analysis. Figure 5 presents the comparative performance of four machine-learning models - Decision Tree, Logistic Regression, Multinomial Naive Bayes, and Gaussian Naive Bayes - evaluated by the metrics Precision, Recall, and F1-Score for each qualification class.

The graphs demonstrate a clear upward trend in all metrics as the qualification level increases.

- For lower levels (1-2), all models exhibit limited performance due to high textual similarity and overlapping competencies.
- Beginning with level 3, the metrics rise steadily, with Logistic Regression and Gaussian Naive Bayes maintaining the highest and most consistent results.
- At levels 6-8, all models reach near-perfect values (Precision, Recall, F1 $\approx$ 0.99), reflecting strong structural distinctiveness of the generated competencies.

These results confirm that integrating Generative AI and Bloom's Taxonomy verbs enhances the semantic separability of competencies and improves classification performance across all models.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

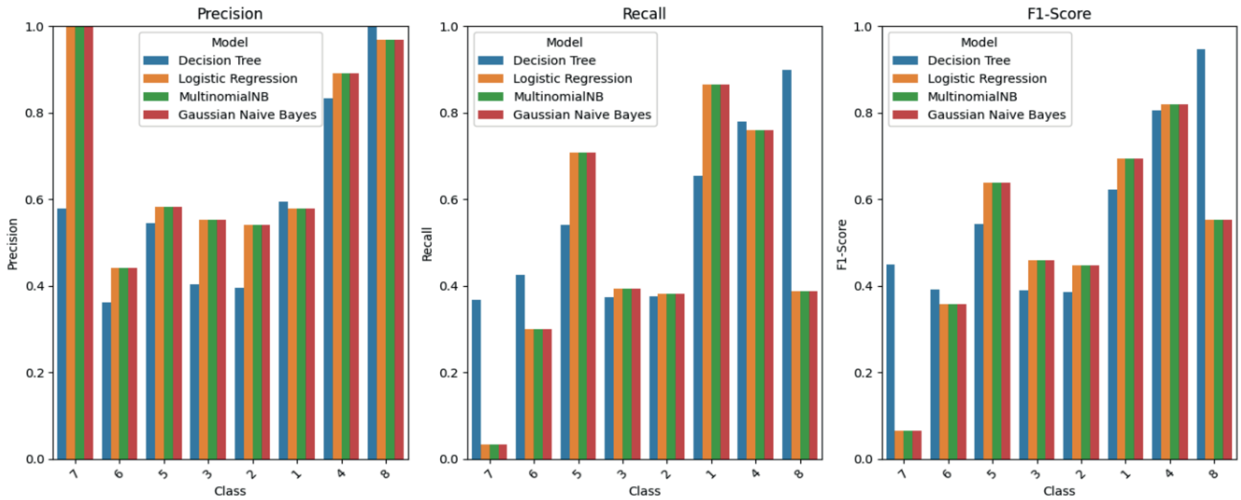

Figure 4. F1-score distribution across qualification levels (1-8) for different classification models.

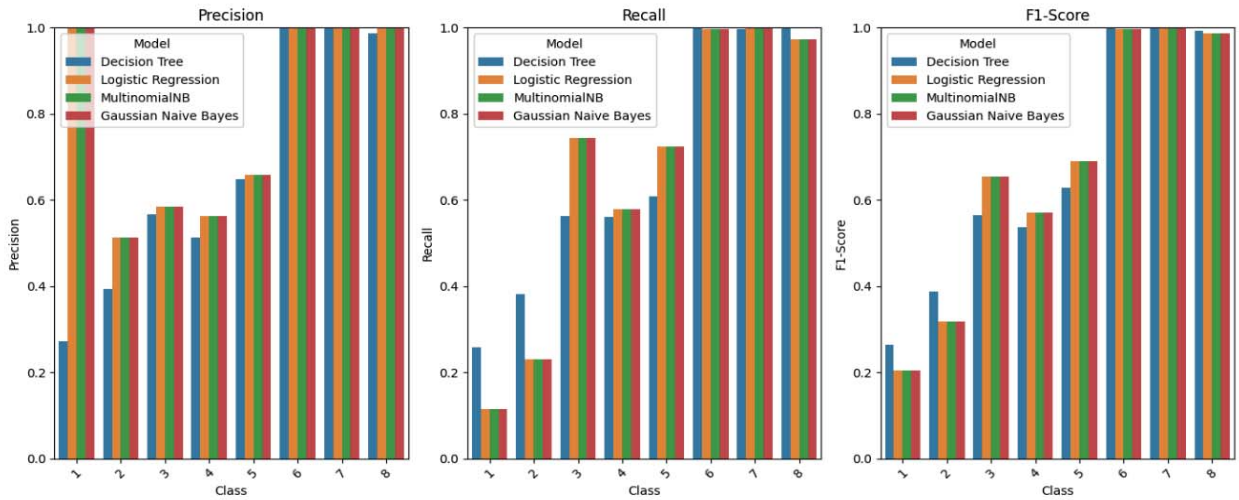

Figure 5. Comparative analysis of Precision, Recall, and F1-Score values for different classification models across qualification levels (1-8).

Discussion

The findings of this study suggest that integrating a hybrid Bloom's Taxonomy and Generative AI approach may improve the structural clarity and differentiation of educational competency frameworks in Kazakhstan. The initial PCA analysis, revealing a cumulative explained variance of approximately 4.5% (PC1: 2.61%, PC2: 1.87%), indicated substantial overlap and limited differentiation across qualification levels in the original 8,736 learning outcomes extracted from RK professional standards. Such low variance is consistent with observations of generic phrasing and inconsistent cognitive progression in traditional competency formulations.

The introduction of 4,024 competencies generated using GPT-4 with Bloom's Taxonomy constraints contributed to improved structural differentiation of the dataset. Post-augmentation PCA results showed an increase in explained variance to approximately 7.5% (PC1: 5.19%, PC2: 2.31%), suggesting clearer clustering patterns, particularly for levels 6–8. This result indicates that AI-assisted competency generation guided by Bloom's action verbs may contribute to improved structural organization of competency statements. The progressive use of Bloom’s verbs-from analyze to innovate-introduces clearer cognitive differentiation across qualification levels, which is reflected in reduced overlap in PCA projections and improved classification performance for higher qualification levels.

Classification experiments provide additional evidence supporting this observation. Prior to augmentation, the Decision Tree model produced F1-scores ranging from 0.449 to 0.948, reflecting moderate separability of competency statements across levels. After augmentation, classification performance for levels 6-8 approached near-perfect values. This effect may reflect clearer lexical and semantic distinctions introduced through structured competency generation based on Bloom's Taxonomy verbs. However, such results should be interpreted cautiously, as the use of structured verb hierarchies may introduce strong lexical signals that facilitate model discrimination.

Several limitations should also be considered. The dataset covers 17 training directions, which may limit the generalizability of the results to other sectors or national educational systems. In addition, generative models such as GPT-4 may introduce biases originating from their training data, which could influence the semantics of generated competencies. Furthermore, the observed increase in explained variance, although indicative of improved structural differentiation, remains relatively modest and suggests that additional analytical techniques may reveal further patterns in competency datasets.

Future research could extend this approach by incorporating expert validation procedures to assess the pedagogical relevance and practical applicability of AI-generated competencies. Additional experiments using alternative embedding techniques or nonlinear dimensionality-reduction methods may also provide deeper insights into competency structure. At the policy level, AI-assisted tools could support the revision of professional standards and curriculum design processes, particularly in large-scale competency frameworks.

Ethics and Governance Considerations

The application of generative artificial intelligence in the development of competency frameworks introduces important ethical and governance considerations. While AI-assisted systems can significantly enhance the efficiency of competency generation and analysis, generative models may occasionally produce inaccurate or contextually inappropriate competency formulations.

In the context of national professional standards, such outputs should therefore be interpreted as recommendations rather than authoritative standards. The integration of AI-generated competencies into official educational frameworks requires expert validation by curriculum designers and domain specialists.

Another consideration concerns potential cultural or contextual mismatches, as large language models are trained on global datasets that may not fully reflect the institutional context of national educational systems. For this reason, AI-generated competencies should be carefully reviewed within the regulatory and educational framework of the Republic of Kazakhstan.

The dataset used in this study consists exclusively of publicly available professional standards and does not include personal or sensitive information. Nevertheless, transparency in data processing and generation procedures remains essential to ensure reproducibility and responsible use of AI in educational policy development.

In practice, generative AI should be viewed as a decision-support tool that assists experts in analyzing and improving competency frameworks rather than replacing human judgment in the design of educational standards.

Conclusion

This study explores the potential of a hybrid Bloom's Taxonomy and Generative AI approach for improving the structural organization of educational competency frameworks within the Republic of Kazakhstan. The initial analysis of 8,736 competency statements extracted from professional standards revealed limited differentiation across qualification levels, with PCA explaining approximately 4.5% of the total variance in the first two components.

The generation of 4,024 additional competencies for qualification levels 6-8, guided by Bloom's Taxonomy verbs, resulted in a moderate increase in explained variance to approximately 7.5%, suggesting improved structural differentiation of competency statements. Classification

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

experiments also demonstrated improved separability of competency levels after dataset augmentation.

These results indicate that integrating generative AI with pedagogical taxonomies may support the systematic refinement of competency frameworks. Rather than replacing human expertise, generative AI can function as an analytical support tool that assists curriculum designers in identifying structural gaps and improving competency differentiation.

Despite several limitations, including potential biases in generative models and the sectoral scope of the dataset, the proposed framework provides a promising direction for AI-assisted educational standards analysis. Future research should incorporate expert validation and explore additional analytical techniques to further strengthen the reliability of AI-supported competency design.

References

[1] Yaskevich B., Kuspangaliev B., Tagliabue L., & Umar T. (2024). Designing a master degree program in building information modeling (BIM) using national professional standards: a case study in Kazakhstan. *Bulletin of L.N. Gumilyov Eurasian National University Technical Science and Technology Series*, *149*(4), 326–350. https://doi.org/10.32523/2616-7263-2024-149-4-326-350

[2] Magauova, A., Makhambetova, Z., Iskakova, A., Yessim, U., & Narkulova, B. (2021). *Experience in developing professional competences of social pedagogues in higher educational institutions of Kazakhstan*. *98*, 01001. https://doi.org/10.1051/SHSCONF/20219801001

[3] Li, G. & Qin, Y. (2024). An Exploration of the Application of Principal Component Analysis in Big Data Processing. *Applied Mathematics and Nonlinear Sciences*, *9*(1). https://doi.org/10.2478/amns-2024-0664

[4] Ramasubramanian, S., Sowmyarani, C. N., Athreya, A. J., Devarajan, A., Shankar, A. U., & Kumar, R. (2024, May). Data dimensionality reduction using principal component analysis: a case study. In *2024 1st International Conference on Communications and Computer Science (InCCCS)* (pp. 1-6). IEEE. https://doi.org/10.1109/InCCCS60947.2024.10593421.

[5] Tonin, F., Tao, Q., Patrinos, P., & Suykens, J. A. K. (2024). Deep Kernel Principal Component Analysis for multi-level feature learning. *Neural Networks*, *170*, 578–595. https://doi.org/10.1016/j.neunet.2023.11.045

[6] Allam, H., Makubvure, L., Gyamfi, B., Graham, K. N., & Akinwolere, K. (2025). Text classification: How machine learning is revolutionizing text categorization. *Information*, 16(2), 130. https://doi.org/10.3390/info16020130

[7] Liu, O. L. (2023). *Assessing student learning outcomes of higher education* (pp. 283–295). Edward Elgar Publishing. https://doi.org/10.4337/9781802204193.00029

[8] Nazina, L. I., Kleymenova, N. L., Nikulcheva, O. S., Pegina, A. N., Pribytkova, O. V., & Egorova, G. N. (2020). Assessment of Learning Outcomes in the University Quality Management System. *Conference Digital Economy*, 478-483. https://doi.org/10.2991/AEBMR.K.200730.087

[9] Adhikari, Y. (2024). A Review of Revised Bloom's Taxonomy of Educational Objectives. Education Review Journal, 1(1), 115-126. https://doi.org/10.3126/erj.v1i1.82852

[10] Gfrerer, Margareth. (2026). Bloom's Taxonomy to Innovate Learning Objectives at Ethiopian University of Applied Sciences Study Programmes. https://link.springer.com/chapter/10.1007/978-3-032-18891-5_31

[11] Calma, A., & Cotronei-Baird, V. (2021). Assessing critical thinking in business education: Key issues and practical solutions. The International Journal of Management Education, 19(3), 100531. https://doi.org/10.1016/j.ijme.2021.100531

[12] Shaw, C. S., & Holmes, K. E. (2014). Critical thinking and online supplemental instruction: A case study. Learning Assistance Review, 19(1), 99–119. https://eric.ed.gov/?id=EJ1034026

[13] Shaikh, F. (2025). Triangulating Effective Teaching: Flipped Learning, Bloom's Taxonomy and Outcome Based Education in Review. *International Journal of Psychological Science*, *5*(1), 8–17. https://doi.org/10.11648/j.ijps.20250501.12

[14] Gonsalves, C. (2024). Generative AI's Impact on Critical Thinking: Revisiting Bloom's Taxonomy. *Journal of Marketing Education*. https://doi.org/10.1177/02734753241305980

[15] Elim, E. H. S. Y. (2026). Promoting cognitive skills in AI-supported learning environments: the integration of bloom's taxonomy. *Education 3-13*, *54*(3), 612–622. https://doi.org/10.1080/03004279.2024.2332469

[16] Helal, M. Y., Elgendy, I. A., Albashrawi, M. A., Dwivedi, Y. K., Al-Ahmadi, M. S., & Jeon, I. (2025). The impact of generative AI on critical thinking skills: a systematic review, conceptual framework and future research directions. *Information Discovery and Delivery*. https://doi.org/10.1108/IDD-05-2025-0125

[17] Waladi, C. & Lamarti, M. S. (2024). Adaptive AI-Driven Assessment for Competency-Based Learning Scenarios. In M. Khaldi (Ed.), *Innovative Instructional Design Methods and Tools for Improved Teaching* (pp. 215-226). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-3128-6.ch010

[18] Lindín, C. (2024). Estrategias para la incorporación de la inteligencia artificial en educación a partir de ChatGPT: Oportunidades y dilemas para profesorado, alumnado e investigación-publicación. *Didacticae*. https://doi.org/10.1344/did.43107

[19] Lubbe, A., Marais, E., & Kruger, D. (2025). Cultivating independent thinkers: The triad of artificial intelligence, Bloom's taxonomy and critical thinking in assessment pedagogy. *Education and Information Technologies*, *30*(12), 17589-17622. https://doi.org/10.1007/s10639-025-13476-x

[20] Mukashova, A. (2025). *Professional Standards Structured Dataset* [Data set]. Kaggle. https://www.kaggle.com/datasets/ainurmukashova/professional-standards-structured-dataset