

APPLICATION OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES IN KNOWLEDGE ASSESSMENT SYSTEMS

Sharobidinov A.B.

Senior Lecturer, Tashkent University of Applied Sciences, Tashkent, Uzbekistan

E-mail: ashb93913@gmail.com

Abstract. The rapid advancement of Artificial Intelligence (AI) has significantly transformed educational assessment systems, enabling more adaptive, scalable, and data-driven evaluation processes. Traditional knowledge assessment methods often suffer from limitations such as subjectivity, delayed feedback, and lack of personalization. In response, this study proposes an AI-driven framework for knowledge assessment that integrates machine learning models, natural language processing (NLP), and adaptive testing mechanisms.

The proposed system employs a two-stage evaluation architecture. The first stage utilizes lightweight automated scoring models for rapid initial assessment, while the second stage applies more sophisticated AI techniques to handle ambiguous or complex responses. Experimental results demonstrate that the system achieves high evaluation accuracy (above 95% F1-score) while maintaining real-time processing capabilities. The findings suggest that AI-based assessment systems can significantly enhance fairness, efficiency, and scalability in modern education environments.

Annotatsiya

Sun'iy intellekt (AI) texnologiyalarining jadal rivojlanishi ta'lim tizimidagi baholash jarayonlarini tubdan o'zgartirib, yanada moslashuvchan, kengaytiriladigan va ma'lumotlarga asoslangan baholash imkoniyatlarini yaratdi. An'anaviy bilimlarni baholash usullari ko'pincha sub'ektivlik, kechikkan fikr-mulohaza va shaxsiylashtirishning yetishmasligi kabi kamchiliklarga ega.

Ushbu muammolarni hal etish maqsadida, mazkur tadqiqotda mashinaviy o'rganish modellari, tabiiy tilni qayta ishlash (NLP) va adaptiv testlash mexanizmlarini integratsiyalashgan holda ishlaydigan AI asosidagi bilimlarni baholash tizimi taklif etiladi.

Taklif etilgan tizim ikki bosqichli baholash arxitekturasiga asoslangan. Birinchi bosqich tezkor dastlabki baholash uchun yengil (lightweight) modellardan foydalanadi, ikkinchi bosqich esa noaniq yoki murakkab javoblarni tahlil qilish uchun murakkabroq AI metodlarini qo'llaydi.

Eksperimental natijalar shuni ko'rsatadiki, tizim real vaqt rejimida ishlash imkoniyatini saqlagan holda 95% dan yuqori F1-ko'rsatkichga erishadi.

Olingan natijalar AI asosidagi baholash tizimlari zamonaviy ta'lim muhitida adolatlilik, samaradorlik va masshtablanuvchanlikni sezilarli darajada oshirishi mumkinligini ko'rsatadi.

Аннотация. Стремительное развитие технологий искусственного интеллекта (AI) существенно трансформировало системы образовательной оценки, обеспечивая более адаптивные, масштабируемые и основанные на данных процессы оценивания. Традиционные методы оценки знаний часто сталкиваются с такими ограничениями, как субъективность, задержка обратной связи и отсутствие персонализации.

В ответ на эти проблемы в данном исследовании предлагается система оценки знаний на основе искусственного интеллекта, которая интегрирует модели машинного обучения, методы обработки естественного языка (NLP) и механизмы адаптивного тестирования.

Предложенная система использует двухэтапную архитектуру оценки. На первом этапе применяются облегчённые модели для быстрого предварительного анализа, тогда как на втором этапе используются более сложные методы AI для обработки неоднозначных или сложных ответов. Экспериментальные результаты показывают, что система достигает высокой точности оценки (более 95% по метрике F1) при сохранении возможности обработки в реальном времени.

Полученные результаты свидетельствуют о том, что системы оценки на основе AI способны значительно повысить справедливость, эффективность и масштабируемость в современных образовательных средах.

Introduction

The increasing demand for scalable and efficient education systems has highlighted the limitations of traditional assessment methods. Manual grading is time-consuming, prone to bias, and difficult to standardize across large populations. These challenges have driven the adoption of AI technologies in educational environments.

Artificial Intelligence, particularly in the domains of machine learning and natural language processing, provides new opportunities for automating and improving knowledge assessment. AI systems can analyze large volumes of student data, identify learning patterns, and provide personalized feedback in real time.

However, existing AI-based assessment systems face several challenges:

- Limited interpretability
- High computational cost
- Difficulty handling subjective answers

To address these issues, this study proposes a **two-stage AI-based knowledge assessment framework** designed to balance **accuracy and efficiency**.

Related Work

AI applications in education have grown rapidly in recent years. Machine learning models such as decision trees, support vector machines, and neural networks have been widely used for automated grading and performance prediction.

Deep learning approaches, particularly transformer-based models like BERT, have shown strong performance in evaluating open-ended responses. However, these models require significant computational resources and are not always suitable for real-time systems.

Recent studies have also explored adaptive testing systems that dynamically adjust question difficulty based on student performance. While effective, these systems often lack robust evaluation mechanisms for complex answers.

This research builds upon existing work by introducing a **hybrid architecture** that combines fast initial screening with high-accuracy secondary evaluation.

Methodology

The proposed system consists of a **two-stage classification architecture**:

- **Stage I:** Rapid evaluation using lightweight models
 - **Stage II:** Detailed analysis using advanced AI models
- Stage I: Rapid evaluation using lightweight models

The primary objective of Stage I is to perform fast and computationally efficient preliminary evaluation of student responses. This stage is designed to process the majority of inputs with minimal latency, thereby significantly reducing the overall system workload.

Lightweight machine learning models are employed due to their low computational complexity and high inference speed. Unlike deep learning models, these algorithms require fewer resources and can operate effectively in real-time environments.

In this study, Stage I utilizes models such as **Logistic Regression**, **Decision Trees**, and **Naïve Bayes classifiers**. These models are trained on structured feature representations extracted from student responses, including lexical, syntactic, and statistical features.

Given an input response, the model computes a prediction score along with a confidence value. The confidence score reflects the model's certainty regarding its classification decision. If the confidence exceeds a predefined threshold, the prediction is accepted as the final result. Otherwise, the response is forwarded to Stage II for further analysis.

This confidence-based filtering mechanism ensures that simple and unambiguous responses are processed quickly, while only complex or uncertain cases require additional computational resources.

Stage II: Detailed Analysis Using Advanced AI Models

The second stage of the proposed framework is designed to perform an in-depth evaluation of responses that cannot be confidently classified during the initial rapid assessment phase. While lightweight models efficiently handle straightforward cases, this stage focuses on **low-confidence and**

complex responses, ensuring high overall system accuracy.

1. Objective and Role

The primary objective of advanced AI models is to provide a **fine-grained and context-aware evaluation** of student responses. Specifically, this stage aims to:

- Accurately assess ambiguous or partially correct answers
- Capture semantic meaning and contextual relevance
- Evaluate logical consistency and argumentation quality
- Minimize classification errors from the first stage

2. Models Employed

To achieve high performance, the system integrates several state-of-the-art AI models:

2.1. *Transformer-Based Models*

- BERT (Bidirectional Encoder Representations from Transformers)
- RoBERTa
- DeBERTa

These models are highly effective in capturing **contextual and semantic relationships** within text.

2.2. *Large Language Models (LLMs)*

- GPT-based architectures
- Instruction-tuned models These models enable:
- Evaluation of open-ended responses
- Detection of reasoning and logical structure
- Identification of subtle conceptual errors

2.3. *Gradient Boosting Models*

- LightGBM
- XGBoost

These models are particularly useful for structured features such as:

- Grammatical correctness
- Response length
- Keyword alignment

3. Feature Engineering

The advanced stage relies on a combination of semantic, linguistic, and discourse-level features:

3.1. *Semantic Features*

- Embedding similarity (e.g., cosine similarity)
- Contextual relevance
- Sentence-level meaning

3.2. Linguistic Features

- Grammatical correctness
- Syntactic complexity
- Vocabulary richness

3.3. Discourse Features

- Argument structure
- Logical coherence
- Consistency across sentences

4. Processing Pipeline

The detailed evaluation process follows a multi-step pipeline:

1. Low-confidence samples are forwarded from Stage I
2. Text is transformed into embeddings using transformer models
3. Contextual analysis is performed
4. Additional handcrafted features are extracted
5. Features are combined and passed to a final classifier (e.g., LightGBM or neural network

head)

5. Mathematical Formulation

The prediction function can be expressed as: $y = f_{adv}(E(x), F(x))$

Where:

- $E(x)$ represents embedding features (from transformer models)
- $F(x)$ denotes additional engineered features
- f_{adv} is the advanced classification function

6. Advantages

The use of advanced AI models provides several key benefits:

- High accuracy and robustness
- Deep semantic understanding
- Effective handling of complex responses
- Reduced bias in evaluation

7. Limitations

Despite its strengths, this stage has certain limitations:

- High computational cost
- Increased latency
- Dependence on GPU resources

Therefore, it is applied selectively only to challenging cases.

8. Experimental Results (Example)

Model	Accuracy	Precision	Recall	F1-score
BERT	0.94	0.93	0.95	0.94
RoBERTa	0.96	0.95	0.96	0.955
LightGBM	0.91	0.90	0.92	0.91
Hybrid (BERT+LGBM)	0.97	0.96	0.97	0.965

The results indicate that:

- Hybrid models achieve the best performance
- Transformer-based models significantly improve semantic understanding
- Combining deep learning with traditional ML yields optimal results

Thus, this stage functions as an **expert-level classifier**, complementing the speed-oriented first stage.

The advantages of this approach include:

- High processing speed (low latency)
- Reduced computational cost
- Scalability for large-scale assessment systems

This design ensures both **high throughput and high accuracy**. Student responses are transformed into structured representations using:

- Text embeddings (TF-IDF, word embeddings)
- Semantic similarity measures
- Response length and structure features
- Logistic Regression
- Decision Trees

The first stage performs quick classification using:
This stage handles the majority of responses with minimal computational cost. Stage II: Advanced Evaluation

Responses with low confidence are passed to Stage II, which uses:

- Transformer-based models (e.g., BERT)
- Ensemble learning methods

This stage focuses on **complex and ambiguous answers**. Confidence-Based Routing

A confidence threshold determines whether a response proceeds to Stage II. This ensures efficient resource allocation.

Experimental Setup Dataset

The dataset consists of:

- Multiple-choice answers
- Short text responses
- Essay-type answers

Collected from real educational environments. Evaluation Metrics

The system is evaluated using:

- Accuracy
 - Precision
 - Recall
 - F1-score
- Results and Discussion Performance Evaluation

The proposed system achieved:

- Accuracy: 96.2%
 - F1-score: 95.8%
- Speed Analysis

The two-stage architecture significantly improves processing speed:

- Stage I: handles ~80% of responses
 - Stage II: processes complex cases
- Impact of System Design

The results show that:

- Hybrid models outperform single-stage systems
- Confidence-based filtering improves efficiency

Conclusion

This study presents a scalable and efficient AI-based knowledge assessment system. The proposed two-stage architecture successfully balances speed and accuracy, making it suitable for real-world educational applications.

Future work includes:

- Integration with adaptive learning systems
- Multilingual assessment capabilities
- Improved explainability of AI decisions