ISSN: 1309 – 6575
Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi
Journal of Measurement and Evaluation in Education and Psychology
Research Article; 2024; 15(4); 421-445

Natural Language Processing and Machine Learning Applications For Assessment and Evaluation in Education: Opportunities and New Approaches

Kübra YILMAZ* Kaan Zülfikar DENİZ**

Abstract

This study examines the applications of Artificial Intelligence (AI), Machine Learning (ML) and Natural Language Processing (NLP) technologies in education, particularly in educational assessment and evaluation processes. The study examines the potential of these technologies to contribute to educational assessment and evaluation processes in areas such as automatic item generation, text mining, sentiment analysis, sentence similarity, and providing feedback to students. The study includes both a literature review and sample applications. In the automatic item generation process of the study, language models such as GPT and Gemini are used to generate new educational questions and this process is supported by NLP technologies. The study is enriched with Turkish examples and the results show that these applications can be further developed for Turkish and have potential for other applications.

Keywords: machine learning in education, natural language processing in education, artificial intelligence, educational technologies

Introduction

AI technologies, which continue to develop today, have started to be used in many areas of life. Studies are being conducted on the integration of AI technologies into disciplines such as healthcare, law, architecture, and education, as well as on preventing various risks (Chaudhry & Kazim, 2021; Shin, 2021; Başarır, 2022; Lin, 2023; Ramachandran & Rana, 2024; İlikhan et al.,2024). AI is defined as "systems that perform given complex tasks by imitating human problem-solving abilities" (Newell & Simon, 1956). It is the science and engineering of creating intelligent machines (McCarthy, 2007). In this context, AI is often equated with algorithms. However, the term algorithm is a concept that existed before AI. The term algorithm is derived from the name of the Persian mathematician Muhammad ibn Musa al-Khwarizmi in the 9th century and means instructions developed to perform a calculation or solve a problem (Sheikh et al.,2023).

Russel & Norvig (2010) AI definitions are cetagorized into four groups. These are: thinking like a human, acting like a human, thinking rationally, and acting rationally. The goal of AI include thinking and acting like a human (Kühl et al, 2020). In other words, AI aims not only to understand intelligence but also to create intelligent beings (Russel & Norvig, 2010). Intelligence is defined as the ability to acquire and apply knowledge and skills while AI is defined as the science of creating artificial entities that from experiences, process and use natural language and develop knowlendge (Balas et al.,2020). AI is seen as the effort to endow computers with human-like characteristics such as perception, association, planning, and reasoning (Boden, 2018).

Human beings have been conducting research on human intelligence and cognitive processes for many years. To this day, human intelligence has not been fully deciphered (Deary et al., 2010). For this reason, the definition of "machines imitating complex human skills" has been seen as a strict definition of AI.

* PhD Student., Ankara University, Faculty of Educational Sciences, Ankara-Türkiye, kubrayilmaz.edu@gmail.com, ORCID ID: 0000-0003-1945-0960
** Prof. Dr., Ankara University, Faculty of Educational Sciences, Ankara-Türkiye, zlfkrdnz@yahoo.com, ORCID ID: 0000-0003-0920-538X

To cite this article:
Yılmaz, K., & Deniz, K. Z. (2024). Natural language processing and machine learning applications for assessment and evaluation in education: opportunities and new approaches. *Journal of Measurement and Evaluation in Education and Psychology, 15*(4), 421-445. https://doi.org/10.21031/epod.1551568

Received: 17.09.2024
Accepted: 28.12.2024

In order to make this definition, it is necessary to identify human-specific skills very well before imitation (Sheikh et al.,2023). In the widest sense, AI is defined as systems designed by humans that interpret data collected from the digital and physical worlds to perform complex tasks given by humans. These systems are established to achieve the best performance in reaching a predetermined goal based on the parameters set by the given data (European Commission, 2018).

This study aims to draw a general framework artificial intelligence (AI), natural language processing (NLP) and machine learning (ML) technologies, to examine the studies on the use of these technologies in education, and to concretise them in the minds of the reader with sample applications that may have potential for the use of these technologies in measurement and evaluation. In line with these aims, it aims to contribute to the field. In this study, topics such as automatic item generation, text visualisation, sentiment analysis, sentence similarity and providing feedback to students are discussed and explained with relevant examples. Firstly, the concept of artificial intelligence will be discussed. It is important to examine the developments in this field from past to present in a historical context.

The Historical Development of AI

When examining the historical development of AI, it will be seen that its foundations were laid in the 1950s. Alan Turing, in his article published in Mind journal, posed the question, "Can machines think?" (Turing, 1950). The most serious response to the question "Can machines think?" was given by McCarthy, Minsky, Rochester, and Shannon (1955) at the Dartmouth Conference held in New Hampshire. This conference is considered a written framework for the concept of "AI." The conference addressed topics that shed light on modern AI, such as automatic computers, programming the use of language, and neural networks.

The years 1950-1970 mark the period when the first research in AI was conducted. During these years, the first AI program, “Logic Theorist,” was developed by Newell and Simon (1956). By 1966, Weizenbaum (1966) developed a program called “Eliza,” which is considered a fundamental work in the field of NLP. In 1969, the first mobile robot capable of perceiving its environment, named “Shakey,” was developed by the Stanford Research Institute (SRI) (Nilsson, 1984). 1970-1980 In this period, which was dubbed as the “AI Winter”, governments reduced funding for AI with the impact of the AI report presented by Lighthill (1973) and developments were interrupted. Reasons such as excessive expectations, technological limitations and lack of data led to the AI winter (Hendler, 2008).

The 1980s-2000s were the period when interest in AI revived and new methods started to be developed. In the 1980s, expert systems had a profound impact on the field of AI, and it was suggested that expert systems could produce solutions to real-world problems. These developments revitalized public interest in AI and raised expectations again (Buchanan & Shortliffe, 1984). After 1990, the fields of ML and data mining gained importance with the increasing amount of data. By the 2010s, deep learning models that can discover complex structures in large data sets and perform operations such as image, video and audio processing have made progress in the field with large data sets and powerful processors (LeCun, Bengio, & Hinton, 2015). Today, AI applications have been used in many areas of life with models such as Gemini developed by Google, GPT-3, GPT-4, GPT 4o developed by Open AI.

At this point, it is important to define the basics of ML and NLP and information about their usage areas in order to ground the subject and to have information about the developments.

Machine Learning (ML)

The term ML was coined by Arthur Samuel in 1959 (Burkov, 2019). ML is based on three concepts: data, model and learning (Deisenroth, Faisal, & Ong, 2020). These systems require large data sets (Kış, 2019). After the data is transferred to the computer environment, models are created in a way that the computer can understand. These models are trained with training data and the accuracy of the trained model is tested with test data. For example, ML algorithms are used in learning fraud detection for credit card transactions and in developing accident prevention systems for cars (Shalev-Shwartz & Ben-David, 2014). It is also used in areas such as face recognition and identification. In ML, decision-making processes are automated after learning based on pre-given samples (Yang & Halim, 2022).

Shalev-Shwartz and Ben-David (2014) explain how the machine learns with an example from natural life. For example, when mice encounter a food whose appearance is different from the previous ones, they eat a small amount, and if the food produces a negative effect, the food is associated with disease. Afterwards, the mice do not eat it when a similar food is encountered. A learning mechanism is at work here. Similarly, when the user marks a mail that falls into the mailbox as 'spam', it informs the AI which mail is 'important' and which mail is 'junk'. When a new e-mail arrives, the machine learns whether to put it in the important folder or the junk folder.

In order for machines to be systems that can think like humans, supervised and unsupervised ML models, regression, classification and clustering are used (Shalev-Shwartz & Ben-David, 2014). ML algorithms consist of four categories developed for different purposes. These are: Supervised Algorithms, Unsupervised Algorithms, Semi-supervised Algorithms, Reinforcement Algorithms (Burkov, 2019).

Supervised ML algorithms are algorithms that require some supervision from the developer. The goal of these algorithms is to predict the target variable using a function defined over a set of independent variables (Burkov, 2019; Mahesh, 2018). Linear regression, logistic regression, decision trees, support vector machines (SVM), k-nearest neighbors (KNN) and naive bayes algorithms are examples of supervised ML algorithms (Mahesh, 2018).

Unsupervised ML algorithms are used when the information used to train is not classified or labeled. While there is a goal in supervised learning, there is no goal in unsupervised learning and an inference is reached (Mahesh, 2018). Processes such as clustering, dimensionality reduction, outlier detection are examples of unsupervised ML. Because in these processes, the model works by making inferences from the natural structure of the data and discovering relationships between data instead of pre-labeled data (Burkov, 2019). Semi-supervised ML algorithms use both labeled and unlabeled data for training. They usually use small amounts of labeled data and large amounts of unlabeled data (Chapelle et al., 2006; Burkov, 2019).

Reinforcement ML uses a technique called exploration; the machine interacts with its environment by generating actions, observes the results, and then takes these results into account when performing the next action. The process continues in this way until the algorithm evolves and chooses the right strategy (Mahesh, 2018). Reinforcement learning is used to solve problems with long-term goals where decision-making stages are sequential, such as game playing, resource management, robotics and logistics (Burkov, 2019).

Artificial neural networks, which are inspired by biological neural networks, work similar to the human body's processes of transmitting stimuli to the brain and responding. There are many hidden layers between the input and output layers. In this way, the strength of the network connections determines the output to be transmitted to the next layer by processing the data coming from the input during the learning process of the network, and thus the model becomes capable of making accurate predictions (Yang & Halim, 2022). ML allows us to make various inferences from data by training models with large data sets. NLP is used for processing text data and obtaining meaningful inferences.

Natural Language Processing (NLP)

Language has an important place in human history. People communicate with each other through languages. People's studies on natural languages shed light on today's research on NLP (Oflazer, 2016: Oflazer & Saraçlar, 2018). Developments in information technologies have encouraged people to study languages (Adalı, 2012).

Computers need to use NLP processes in order to understand and communicate with human languages (Şeker, 2015). NLP is a broad set of technologies used to analyze texts semantically and syntactically and to extract meaning (Wijeratne et al., 2009). NLP is a computerized approach to analyzing text. There is no single agreed definition of NLP (Liddy, 2001).

NLP studies and text mining studies are often used together. Text mining studies consist of studies that accept text as a data source. For example, it is used in studies such as classification of texts, extraction of topics from texts, classification, sentiment analysis, text summarization, entity relationship modeling (Şeker, 2015). The branch of science called NLP studies the processing of languages with the help of computers (Adalı, 2012).

One of the most basic concepts of NLP is the concept of corpus. Corpus can be defined as texts written in a language. Corpuses are often used to study language features, train large language models and improve NLP algorithms (Jurafsky & Martin, 2024)**.** Language models are trained with textual data and can learn the structure of the language and language rules with the help of these corpora. Thus, they can make more accurate predictions (Dong, 2023). Turkish has a very rich corpus structure (Sak et al., 2011).

Turkish is in the Turkic languages group of the Altaic language family. Other languages in this language family are Mongolian, Tungusic, Korean and Japanese. For agglutinative languages such as Turkish, there are various difficulties in NLP (Oflazer & Saraçlar, 2018). However, nowadays, various operations can be performed in the field of Turkish NLP with the help of Turkish NLP tools. There are tools developed for Turkish NLP and libraries such as Zemberek-NLP, Turkish Stemmer, TrTokenizer, Mukayese (Merdun et al., 2024; Usta, 2024).

Topics such as NLP and ML are also attracting attention in the field of education and studies are being carried out to integrate them into education (Gierl et al., 2008; Gierl & Lai, 2016, Uysal, 2019; Göloğlu-Demir & Yılmaz, 2018; Mulianingsih et al., 2020; Coelho et al. 2023: Sytnyk & Podlinyayeva, 2024). Other topics to be addressed within the scope of this study are text mining, topic modelling, sentence similarity, student feedback, sentiment analysis, and automatic question generation. See Appendices for all sample applications and code examples.

Text Mining and Topic Modeling

The most common definition of the term data mining is discovering patterns for data (Leskovec et al., 2014). One branch of data mining is text mining. Text mining is concerned with how to determine what the subject of a document is about. Text mining is a method for categorizing texts that contain many topics such as news articles and blog posts into groups according to their topics (Silge & Robinson, 2017). Topic modeling is a method used in NLP applications such as sentiment analysis, document classification, speech recognition, automatic translation. In terms of text mining, topic modeling is based on the bag-of-words assumption (Alghamdi & Alfalqi, 2015).

Topic modeling can provide methods to automatically organize, understand, search and summarize large text data without manual work (Blei et al., 2003). Topic modeling is used to discover patterns of word usage in a document (Alghamdi & Alfalqi, 2015).

Topic modeling is a generative bag-of-words model that learns topics and topic words from frequency measures in texts (Mazidi, 2018). It is a system based on ML and NLP. The LDA model developed by Grun and Hornik (2011) as an R package is used as a topic modeling approach. With the LDA model, subtopics in texts are modeled and thus texts are grouped according to their content similarities. For example, news texts in seven different categories in Turkish (world, economy, culture and arts, health, politics, sports and technology) can be classified under the category they belong to (Yıldırım & Yıldız, 2018). The main point here is that ML learns which category the topics in the text belong to and assigns the texts to the class they belong to.

Text mining is seen as a method that can be used in many areas such as monitoring and evaluating student performance, providing feedback and support to students, and discovering the points where students have difficulty (Ferreira et al., 2020). Word clouds allow some of the findings obtained from text mining to be presented in visualised form. For example; with the help of a word cloud, it is possible to visually see which words are used more by students who answer a question correctly and which words are used more by students who answer incorrectly. In this study, word cloud is introduced with a sample application. See Appendices for application and code examples.

Sentence Similarity

Sentence similarity is one of the most widely used subfields of NLP. It is used to measure the similarity between sentences for tasks such as question answering, information retrieval, summarisation and plagiarism detection (Farouk, 2019). In the sentence similarity approach, words are represented as vectors and the similarities between these vectors are calculated mathematically (Mikolov et al., 2013). Sentence similarity can be interpreted as semantic inferences at the sentence level (Guu et al., 2018). For example, when a customer asks a bank's chatbot a question about a loan, the chatbot matches it with the questions stored in its memory and determines the appropriate response to the customer.

In the sentence similarity approach, the degree of similarity between a reference answer and other answers can be measured. An example of how this method can be used in education is the evaluation of open-ended exam answers by comparing student answers to a reference answer. The closer the similarity score is to 1, the closer the student's answer is to the reference answer and therefore considered correct. For example, Chamidah et al. (2021) the study introduced an essay evaluation system that utilises the similarity between student answers and reference answers using short-answer questions from Indonesia. The method extended the reference answers with synonyms and compared them using Cosine similarity, Jaccard similarity and Dice similarity measures. The questions are categorised into four topics: politics, lifestyle, sport and technology.

In a study developed to evaluate short answers in education, students' answers were compared with reference answers using a semantic similarity measure. The results showed a correlation of 0.70 between manual evaluation and system evaluation. The developed system provides fast and consistent scoring of short answers and significantly supports the reliability of manual scoring (Lubis et al., 2021).

The closer the similarity score is to 1, the closer the student's answer is to the reference answer and therefore considered correct (Wang & Dong, 2020; Chamidah et al., 2021). This method is also seen as a potential tool to detect similarities between student answers and prevent plagiarism. Moreover, this method is seen as a functional tool for measuring language skills and written expression abilities. Providing feedback on students' learning is another important aspect.

Feedback to Students

Feedback is necessary for identifying deficiencies that need to be addressed in education, providing various improvements and developing curricula in this context. The historical development of feedback can be traced back to Thorndike's “Law of Effect” (Lipnevich & Panadero, 2021). Giving feedback to students is important in education. However, giving feedback to each student individually is challenging in terms of time and effort (Cavalcanti et al., 2019).

Effective feedback has some characteristics. Feedback requires that what is expected from the student as a result of the evaluation of the task performed by the student is conveyed to the student in an understandable way. Feedback should be in a way that contributes to the student's learning processes and plays a constructive role. At the same time, it should cover all aspects of the task by focusing on missing gains (Kayalı et al., 2019).

One study is working on a system that gives instant feedback to students. The study is conducted as a pilot study on 800 students studying at a university in India. In this study, it is aimed to give almost real-time feedback to students' writings through the system (Lewkow et al., 2016). In another feedback study, both automatic scoring and giving feedback were studied. The system provides feedback that will allow the student to make the necessary corrections before submitting their work (Woods et al., 2017).

When the number of students is high, it is very difficult to give feedback to each student individually. By using NLP techniques and clustering analysis, common feedback can be given to students who give similar answers. Thus, communication with students will be maintained and time can be saved while doing so. Instead of giving feedback to the students individually, the answers of the students who give similar answers are processed with NLP, converted into mathematical values and clustered according to similarity measures. The same feedback can be given to the student for the answers in the same cluster.

In addition to giving feedback to the student, various improvements can be made in education with the feedback received from the students about the course. With this feedback, students' emotional states can be determined and steps can be taken to improve education accordingly (Kasumba & Neumann, 2024).

A review of the literature reveals various studies in which automated feedback is used in education, with different techniques being developed for this purpose. For example, Lu and Cutumisu (2021) conducted a two-stage study using deep learning approaches and NLP techniques to provide automated written feedback and assessment in education. In their study, three deep learning models (CNN, CNN + LSTM, and CNN + Bi-LSTM) were tested for automated assessment, with the LSTM model achieving the highest accuracy. This model demonstrated an average performance of 0.73 on the Quadratic Weighted Kappa (QWK) metric, which measures alignment with human evaluation.

For the automated feedback stage, the Constrained Generation by Metropolis-Hastings Sampling (CGMH) method was employed to generate contextually appropriate feedback sentences. These sentences were automatically structured based on errors found in students' writing.

In this study, the students were divided into 3 clusters according to the similarities of their answers and the feedbacks written by the researchers in accordance with the clusters were assigned to the cluster they belonged to by the developed system. With this sample application, a basic level of concretisation for automatic feedback in the reader's mind was aimed to be made. See Appendix for the sample application.

Sentiment Analysis

Sentiment analysis is an application of NLP that examines whether individuals' opinions on a topic are positive, negative or neutral. Since the manual construction and validation of a sentiment lexicon is labor- and time-consuming, many studies have explored automated ways of identifying sentiment-related features in text. According to Dong (2023), two different approaches are used for sentiment analysis. One is a rule-based approach and the other is a ML-based approach. In the ML-based approach, emotions are identified with large amounts of text data to train the model. In the rule-based approach, negative words, emotional words, language features are identified within the framework of predetermined rules and emotions in the text are evaluated as 'positive, negative, neutral'. In addition to ML methods, there are methods developed by researchers for sentiment analysis. Hutto and Gilbert (2014) compared ML and VADER (Valence Aware Dictionary and Sentiment Reasoner) methods on 4000 tweets and found that the VADER method performed better. (Sukmana & Rusydiana, 2023).

In the study conducted by Bostancı and Albayrak (2021), sentiment analysis method was used to extract appropriate advertising content for students during the university preference period. The study was conducted on the comments of 82 twitter and 65 facebook users. Student emotions were classified as optimistic, pessimistic, humorous, productive and extraverted. Accordingly, university advertisement posters were designed for the determined emotional states. In their study, Göloğlu-Demir and Yılmaz (2018) calculated the TF-IDF ratios of the 10 most common words containing 10 positive and 10 negative emotions among 36081 words written by 40 participants for 4 days. As a result of the study, it was concluded that the majority of the students had positive feelings about the project.

In recent years, ML and big language models have also been used in studies (Kasumba & Neumann, 2024; Peña-Torres, 2024). Sentiment analysis can be used in the field of education to obtain students' opinions about a course, subject or teaching method and to make various improvements in educational practices (Peña-Torres, 2024). In recent years, sentiment analysis studies in education have become widespread (Sukmana & Rusydiana, 2023; Lin, 2023; Kasumba & Neumann, 2024; Peña-Torres, 2024). Another subject that is becoming widespread in education is text and question generation (Shin, 2021; Kasumba & Neumann, 2024).

Automated Question Generation

Automatic question generation is divided into template-based and non-template-based approaches (Gierl & Lai, 2016). An example of template-based approaches is IGOR, an automatic item generation tool that allows users to generate a variety of test items and can be applied in areas such as mathematics (Gierl et al., 2008). Template-based approaches use auxiliary information such as text, figures and graphs to generate test items with logical and appropriate values (Singley & Bennett, 2002). Non-template-based approaches produce new texts using NLP techniques. These technologies produce more rational results and add new dimensions to automatic question generation as training data increases. Various tools are being developed to support text generation, one of which is Texar. Texar is a toolkit that can be used in text generation in the field of NLP (Hu et al., 2018). By using neural networks and NLP techniques for automatic question generation, semantic features of texts can be identified and test items can be generated (Shin, 2021).

Today, generative AI tools can be used for this purpose. GPT-3 achieves high success in text generation by processing large amounts of data with 175 billion parameters (Brown et al., 2020). Language models have developed rapidly in recent years. For example, the GPT (Generative Pre-trained Transformer) model introduced by OpenAI is a comprehensive model. GPT-3 adds a new dimension to all the mentioned text generation approaches. For this model, the entire Wikipedia was used as training material. In addition, text data equivalent to 32 times of Wikipedia was also included in the training data of the model. This model was created with 175 billion parameters (Brown et al., 2020).

GPT-3, with its system structure consisting of 175 billion parameters and thousands of introduced texts, can write an article on any topic in seconds, continue any text (in an unprecedented way), and generate both multiple choice and open-ended questions about the text. Machine learning and NLP systems are described as data hungry. The more data they can be trained on, the more rational results they can produce. With the API support provided by OpenAI, researchers can generate various texts in their fields and obtain text-related questions. GPT models can generate previously unseen texts by utilising training data. The most recently optimised GPT model, GPT-4o, is the result of the development of more advanced algorithms by optimising large language models (LLaMEA - Large Language Model Evolutionary Algorithm) (OpenAI, 2023). Another large language model developed by Google is Gemini.

In a study by Zeinalipour et al. (2024), large language models such as GPT-4-Turbo, GPT-3.5-Turbo and Llama were used for automatic question generation in Turkish. The dataset of the study includes various disciplines such as chemistry, biology, geography, philosophy, Turkish language and literature, and history. According to the findings of the study, big language models can be used effectively in the process of creating educational content. In order to provide a broader framework for the use of these technologies in education, the relevant literature has also been examined in education.

The Use of AI Approaches in Education

The intelligence of a learner is often equated with the ability to recall learnt information (Nafea, 2016). This system ignores individual differences, readiness and varying learning speeds. AI applications in education contribute to students' contextual learning and can provide individualised learning experiences for each student (Chaudhry & Kazim, 2021). The application of AI in education has been the focus of academic research for over 30 years (Hamal et al., 2022).

AI technologies can be used to improve learning experiences, improve educational outcomes, develop individualised learning systems to enable students to learn at their own pace, support teachers in material development, etc., and provide instant support to students through gamification, interactive simulations, virtual assistants (Kotlyarova, 2022). In addition, AI applications can be used in many fields such as automating assessment stages, providing instant feedback to students, providing access to space-independent classrooms, medicine, marketing, engineering education, etc. (Sadiku, Ashaolu, Ajayi-Majebi, & Musa, 2021). Tools such as Intelligent Tutoring Systems (ITS) and contextual learning environments (iTalk2Learn and AIDA) are also among the opportunities offered by AI in education (Chaudhry & Kazim, 2021). It is predicted that AI can reduce inequalities among students by providing personalised learning opportunities to individuals (Nkechi et al., 2024). For example, Holstein, McLaren, and Aleven (2018) concluded in their study with 8 teachers and 286 secondary school students in 18 classrooms that it can reduce the gap in learning outcomes between students with different readiness.

It can also be used in areas such as automatic assessment and teacher observation of student progress (Nafea, 2016). Distance education is another area where AI can be used (Coelho et al.,2023). In this context, technologies such as AI, machine learning have a great potential to provide individualised learning experiences in education and increase the efficiency of learning processes and bring new approaches to educators.

For example, these technologies can be utilised in language learning. Perveen (2021) conducted a study on a group of students attending two different courses on English language learning. Word clouds were used in the study. The findings of the study showed that word clouds are a suitable tool for task-based assessment, especially in pre-reading and pre-writing activities.

Another noteworthy area of study is the use of the counterfactual approach to improve the achievement of at-risk students. Cavus and Kuzilek (2024a; 2024b) conducted two studies on counterfactualism. In the first of these studies, counterfactual methods were utilised to increase student achievement in education. These methods are used to provide more accurate counterfactual explanations (what-if scenarios) to students at risk of failure in education. In the study, it is possible to provide meaningful and effective explanations about which factors need to change to increase student achievement. The NICE method was found to be more effective than other methods (Cavus & Kuzilek, 2024a). The second study by Cavus and Kuzilek (2024b) emphasises the importance of the actionability of counterfactual explanations to provide accurate guidance to students at risk of failure.

The aim of the study is to identify specific characteristics that categorise students as at-risk and to show how adjusting for these characteristics can improve their achievement. For this purpose, model-agnostic explanation methods, in particular LIME and SHAP, were used. The results showed that SHAP is more stable and reliable.

The use of counterfactual explanations provides individualised support for students at risk of failing a course (Smith et al., 2022). In a study aiming to understand the impact of counterfactual explanations on student performance, using data from 134 successful and 148 unsuccessful students, it was concluded that the developed system provides individualised recommendations for each student (Tsiakmaki & Ragos, 2021). It is seen that artificial intelligence technologies can be used in subjects such as individual learning, automatic assessment, instant feedback, and counterfactual explanations can be useful to increase student achievement. Young (2024) mentioned the advantages and disadvantages of integrating AI into education. Accordingly, it is stated that AI provides great improvements in education such as personal learning, assessment, creating relevant educational content and virtual reality. All these study examples show that the use of AI in education is effective and beneficial. There is still a lack of research on the inclusion of AI in educational applications. In this study, topics such as automatic item generation, text visualisation, sentiment analysis, sentence similarity and providing feedback to students are discussed and explained with relevant examples. See Appendices for all sample applications and code examples.

Methods

Research Design

This study is designed to draw a general framework for technologies such as artificial intelligence, machine learning and NLP and to examine the use of these technologies in education. The research also aims to concretise the use of these technologies in the minds of the readers by supporting them with sample applications.

This study examines the potential of artificial intelligence, machine learning and NLP technologies to contribute to measurement and evaluation processes in education. Figure 1 shows the flow chart for the literature review and Figure 2 shows the flow chart for the application. Various searches were made in national and international databases (such as scopus, google scholar, national thesis center) to reach the articles to be analyzed. A total of 90 sources were examined and detailed explanations were provided for the selected examples. The distribution of the sources is as follows:

Table 1.
Distribution of Sources by Category

Main Category	Count
AI	20
AI in Education	16
NLP	12
Sentiment Analysis	7
Question Generation	7
ML	6
Text Mining	6
Feedback	4
Automated Written Scoring	3
Turkish NLP	3
Sentence Similarity	2
GPT-Gemini Models	2
NLP Library	2
Total	90

Search Criteria

The keywords used in the research are: ‘artificial intelligence’, ‘machine learning’, ‘natural language processing’, ‘artificial intelligence in education’. Databases such as ‘Scopus, Google Scholar, National Thesis Centre’ were used in the study. Research tools such as ‘Connected Papers and Typeset’ were also used in the literature review.

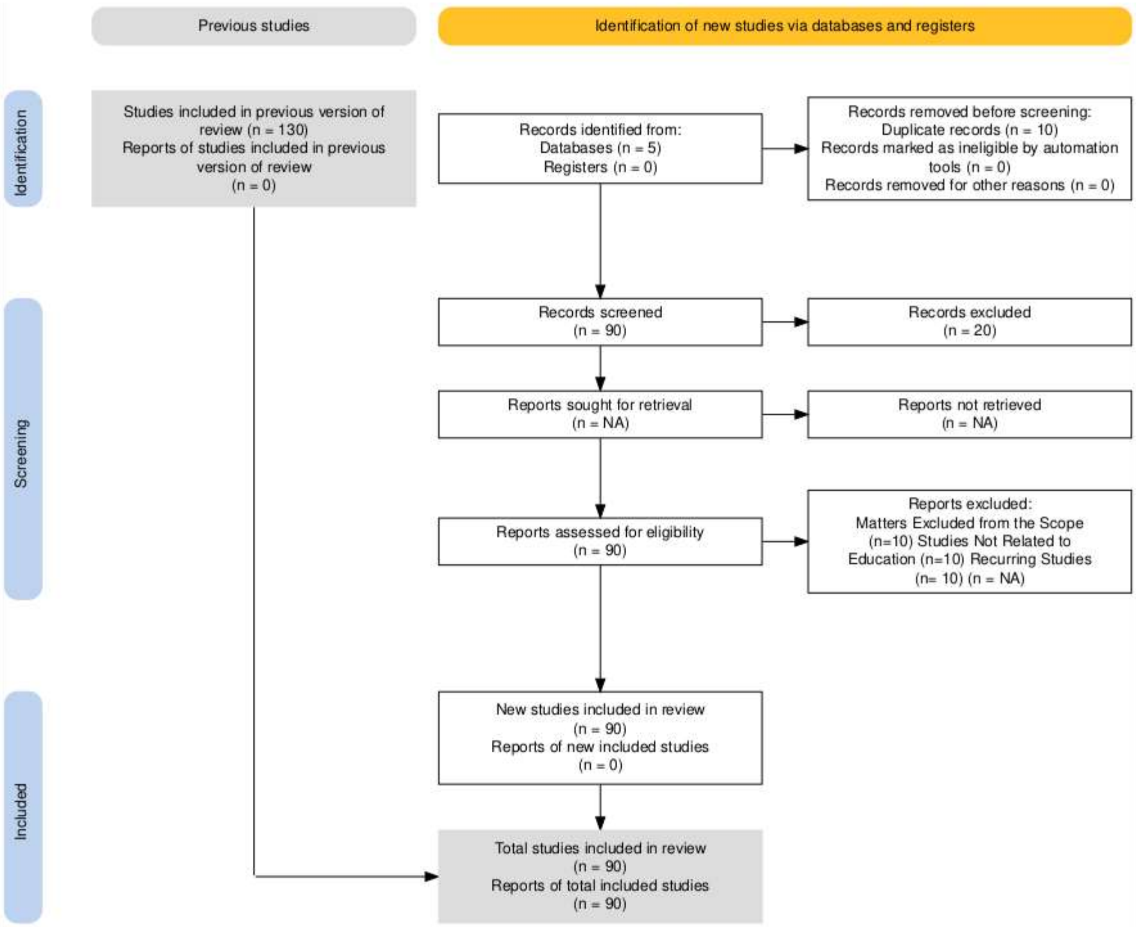
Search Process

90 studies related to the keywords were included in the review, and non-related studies were eliminated. The details of this process are given in the prisma diagram in the following section.

Selection Criteria

In the articles selected for the review, attention was paid to the potential use of artificial intelligence, machine learning and NLP technologies in the field of measurement and evaluation in education and to include application examples. The flow chart of the application steps is as follows:

Figure 1.
Research Process Prisma Flow Diagram (Haddaway, Page, Pritchard and McGuinness, 2022).



Data Analysis

In addition to the literature review, sample applications were also included in the study. The methods and tools used in the study are as follows.

Tools and Methods

Sentence similarity is related to the identification and comparison of text features (Mohler & Mihalcea, 2009). Sentence Transformers model was used for sentence similarity. Turkish Zeyrek (Zeyrek, 2020) and Pandas (McKinney, 2010) libraries were used for sentiment analysis. In addition, Seaborn (Waskom, 2020) and Matplotlib (Hunter, 2007) libraries were used for data visualisation.

In this step, the TF-IDF (Term Frequency-Inverse Document Frequency) vectorisation method was used. In the text vectorisation stage, TF expresses the frequency of occurrence of the word in the document, while IDF measures the ability of the word to distinguish between categories. In this study, texts were converted into numerical values by TF-IDF vectorisation (Shin, 2021; Chen el al., 2016). After this process, student responses were clustered. KMeans algorithm in Python 3.6.11 was used for clustering analysis.

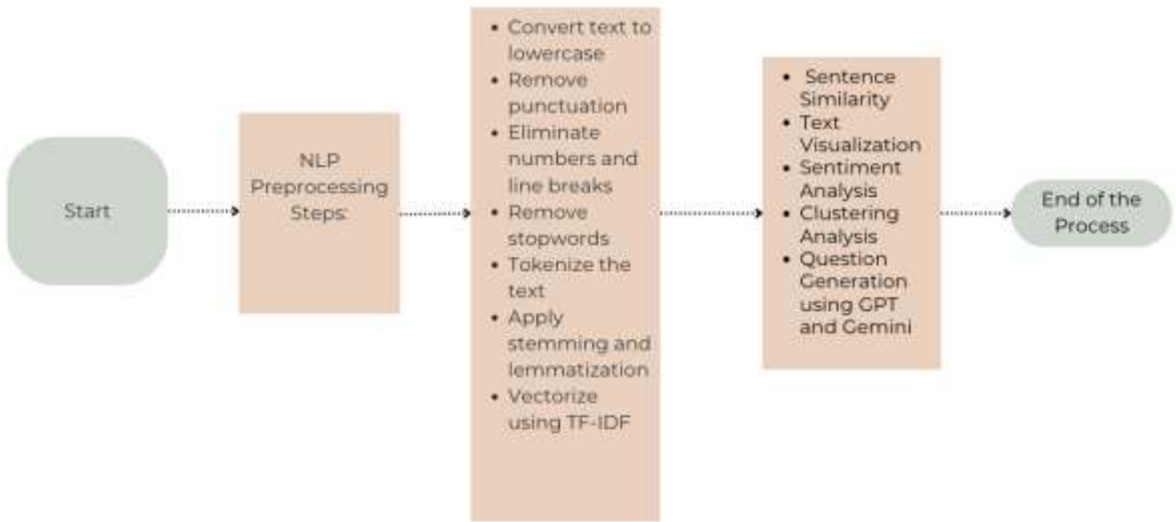
In this step, NLP data preprocessing steps were performed by the researchers. Feedbacks were written by the researchers in accordance with the student responses divided into 3 clusters. For feedback, Scikit-Learn (Pedregosa et al., 2011) and Pandas libraries were used to analyse student responses and assign appropriate feedback to these responses.

Using ChatGPT and GPT-4

In the clustering analysis part, coding assistance was obtained from ChatGPT. In the sentiment analysis stage, student responses were generated with ChatGTP and sentences were grouped according to the appropriate emotion (positive, negative, neutral). Language models such as GPT-4 and Gemini were actively used for question generation.

In this study, data preprocessing steps, the first step of NLP, were performed. Sentence similarity and word cloud operations were coded by the researchers. NLP data preprocessing steps are shown in Figure 2. Sample applications were implemented using Python programming language and these applications are presented in the Appendix with code samples. The flowchart of the application steps is as follows:

Figure 2.
Application Flowchart



Conclusions

Text Visualisation Examples

The use of word clouds helps students to grasp the main theme of the text by highlighting words that occur frequently in sentences. Important findings in the text visualisation literature were presented by Perveen (2021). In this study, an example of a word cloud visualised from the 'Anthem of Independence' was given. A word cloud was created with the most frequently mentioned words in the National Anthem. In addition, word cloud examples containing the correct and incorrect answers given by the students to the question 'What is the meaning and importance of the National Anthem?' were created using GPT. Especially in the 'İstiklal Marşı' example, it is foreseen that it can provide an opportunity to evaluate students' understanding of the main themes of the anthem such as patriotism, freedom and unity. Visual objects can be useful in terms of memorisation. The words that stand out in students' correct and incorrect answers can also be evaluated.

Sentence Similarity Examples

Two examples implemented in Python programming language are given in the Appendix. In the first example, the sentence similarity score is calculated as 0.77. In the other example, the similarity score between two sentences is calculated as 0.91. A similarity score of 0.77 indicates a moderate level of similarity between the two sentences, while a higher score of 0.91 indicates that the sentences share almost the same meaning.

The results obtained in this study (0.77; 0.91) are higher than the results obtained by Lubis et al., (2021) (0.70) for both sentence pairs. When used in automated assessment systems, these scores reflect how close the student's answer is to the model answer, allowing an objective assessment to be made. This method can be used to help graders in the evaluation of open-ended questions and to speed up the process.

Student Feedback Example

In the example below, the text and responses were generated using GPT-3.5. In the analysis of the text, Turkish NLP techniques and clustering analysis, an unsupervised ML model, were used in Python programming language. Student responses were categorised into three clusters according to their similarities. Predetermined feedbacks for each cluster were automatically added next to the responses in that cluster. In the literature, Lu and Cutumisu (2021) conducted more extensive studies on automatic written feedback in education. In this study, a basic example for automatic feedback is presented and an example of the use of both ML and NLP is created.

Sentiment Analysis Example

In this study, a sample application was made at a basic level. Student views on a mathematics lesson created using GPT-4o, a code sample for sentiment analysis and a pie chart of student views were created. Accordingly, 4 out of 10 students expressed positive, 4 negative and 2 undecided opinions about the mathematics lesson. Research on sentiment analysis emphasises the potential of these technologies in education (Kort et al., 2001; Peña-Torres, 2024). It is seen that the use of sentiment analysis in education will be beneficial in studies conducted on real students with larger data sets.

Item Generation Example

In studies on question generation (Gierl et al., 2008; Gierl & Lai, 2016; Shin, 2021), it is seen that models developed by researchers are used. However, studies on generating questions with GPT have also increased in recent years (Smith et al., 2024; Berger et al., 2024). In this study, a text was created with GPT4 in the question generation phase and open-ended, multiple-choice questions were generated based on this text. Similarly, both open-ended and multiple-choice questions were generated with Gemini. In future studies, questions created with both tools can be applied to real students and the results can be evaluated. See Appendices for all sample applications and code examples.

Discussion and Suggestions

This study addresses the historical development of AI, ML, NLP techniques, and their implications for education, including both a literature review and practical applications. The results indicate that approaches like ML and NLP are applicable in educational settings (Zeinalipour et al., 2024; Smith et al., 2024; Berger et al., 2024). Student feedback plays a critical role in improving the learning process in education; Lu and Cutumisu (2021) and Kasumba and Neumann (2024) provide significant findings on this topic. Through counterfactual validity, factors that contribute to improving student success can be focused on (Cavus & Kuzilek, 2024a). In addition to traditional teaching methods, new approaches such as ML and NLP can contribute to students' learning processes by supporting each other (Nafea, 2016).

If we consider the points that all these applications will contribute to the field of measurement and evaluation in education; personalised learning environments will provide individuals with the opportunity to evaluate themselves and progress at their own pace. With tests adapted to the individual, the individual will be evaluated with questions appropriate to his/her own level, which will allow each student to catch his/her own success scale without ignoring individual differences. This has the potential to contribute to the constructivist approach. These tools will also save time for question writers and practitioners and can be used as an auxiliary tool in the process of preparing questions suitable for each level. Of course, the questions developed with these tools need to be tested on appropriate samples and used in a controlled manner by humans. Thanks to automatic scoring and feedback, the use of open-ended items will become widespread, and more information can be obtained in the measurement of high-level cognitive skills by conducting an interactive process with the student. At the same time, it will contribute to the formative assessment approach and prepare the ground for making necessary improvements in education in the light of this feedback from students. In addition to feedback, emotion analysis, which is another issue addressed in this study, will support these improvements by enabling more information to be obtained about the student in accordance with the formative assessment approach.

Text visualisation can be interesting for students. In addition, it is foreseen that interactive environments such as simulation, virtual reality, game-based learning that can attract students' interest will also be included in learning and assessment processes. AI technologies have the potential to bring very useful innovations for individuals with special needs. For example, it is predicted that virtual assistants can be designed to provide reading support to students with dyslexia.

AI technologies can be easily used both in K-12 and higher education when the necessary infrastructures are provided. While simpler, easy-to-apply and easy-to-use tools are integrated into educational processes at K-12 level, it is predicted that more complex structures and advanced technologies can be used at higher education level. Scaling issues can be addressed by using open source materials at both levels and developing modular systems suitable for each level. The integration of these tools into education can be tested with pilot applications for both levels of education and the results can be evaluated.

More detailed information about students can be obtained by conducting various studies (ML, NLP) on student data obtained from learning management systems (LMS). In the light of these data, learning environments suitable for students can be designed. Customised chatbots can be developed where students can ask questions on any subject at any time.

All these innovations bring with them ethical and security issues. Young (2024) addressed these issues as data privacy and security, bias and fairness, job loss and equality in education, accessibility and inclusiveness. At this point, ethical principles and guidelines regarding the use of AI in education should be determined, copy detection software should be developed, data privacy and security should be ensured. Accountability and necessary transparency should be provided on how the systems work. At the same time, educator training should be emphasised at the point of AI and educators should be given competence in these issues.

In conclusion, it is evident that the use of these technologies in assessment and evaluation in education has been increasing and holds potential for enhancing student success and improving the quality of education. Future research should focus on further studies regarding the integration of these technologies into assessment and evaluation in education.

The sample questions and student answers in this study were generated with generative artificial intelligence tools such as GPT and Gemini. In future studies, it is suggested that the questions generated with these tools should be applied on real student sets and necessary psychometric studies should be carried out.

In this study, an example of automatic feedback to students was given. In future studies, feedback from students about the courses can be processed and evaluated what kind of improvements can be made in this direction. Generalisability of these technologies can be ensured with studies conducted on different languages.

Declarations

Gen-AI Use: The authors of this article declare that Gen-AI tools have NOT been used in any capacity for content creation in this work.
Author Contribution: The first author led the study and contributed to conceptualization, methodology, data modeling, analysis, and visualization, interpretation, and writing. All the other authors played critical roles in shaping the study by contributing to concept, methodology, interpretation, or revision.
Conflict of Interest: No potential conflict of interest was reported by the authors.
Consent to Publish: Written consent was sought from each author to publish the manuscript.
Competing Interests: The authors have no relevant financial or non-financial interests to disclose.

References

Adalı, E. (2012). Doğal dil işleme. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 5*(2). https://dergipark.org.tr/tr/pub/tbbmd/issue/22245/238797

Alghamdi, R., & Alfalqi, K. (2015). A survey of topic modeling in text mining. *International Journal of Advanced Computer Science and Applications (IJACSA), 6*(1), 147-153.

Balas, V. E., Kumar, R., & Srivastava, R. (Eds.). (2020). *Recent trends and advances in AI and internet of things*. Springer Nature. https://doi.org/10.1007/978-3-030-32644-9.

Başarır, L. (2022). Modelling AI in architectural education. *Gazi University Journal of Science, 35*(4), 1260-1278. https://doi.org/10.35378/gujs.967981

Berger, M., Kinsley, A., & Chawla, S. (2024). A novel multi-stage prompting approach for language agnostic MCQ generation using GPT. *arXiv preprint arXiv:2401.07098*. https://arxiv.org/abs/2401.07098

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of ML Research, 3*, 993-1022. https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf

Boden, M. A. (2018). What is AI? In *AI: A Very Short Introduction*. Oxford University Press. https://doi.org/10.1093/actrade/9780199602919.003.0001

Bostancı, B., & Albayrak, A. (2021). Duygu Analizi İle Kişiye Özel İçerik Önermek. Veri Bilimi, 4(1), 53-60. https://dergipark.org.tr/tr/pub/veri/issue/59505/777675#article_cite

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., & Amodei, D. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*. https://arxiv.org/abs/2005.14165

Buchanan, B. G., & Shortliffe, E. H. (1984). *Rule-based expert systems*. Addison-Wesley.

Burkov, A. (2019). *The hundred-page machine learning book.* Andriy Burkov.

Cavalcanti, A. P., Ferreira Leite de Mello, R., Rolim, V., André, M., Freitas, F., & Gaševic, D. (2019). An analysis of the use of good feedback practices in online learning courses. In *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)* (pp. 153-157). Maceio, Brazil. https://doi.org/10.1109/ICALT.2019.00061

Cavus, M., & Kuzilek, J. (2024a). *An effect analysis of the balancing techniques on the counterfactual explanations of student success prediction models*. arXiv preprint arXiv:2408.00676.

Cavus, M., & Kuzilek, J. (2024b). *The Actionable Explanations for Student Success Prediction Models: A Benchmark Study on the Quality of Counterfactual Methods*. arXiv preprint arXiv:2405.14016.

Chamidah, N., Santoni, M. M., Irmanda, H. N., Astriratma, R., Tua, L. M. & Yuniati, T. (2021). Word Expansion using Synonyms in Indonesian Short Essay Auto Scoring. *International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS).* doi:10.1109/ICIMCIS53775.2021.9699374

Chapelle, O., Schölkopf, B., & Zien, A. (2006). *Semi-Supervised Learning*. MIT Press. Retrieved from http://www.acad.bg/ebook/ml/MITPress-%20SemiSupervised%20Learning.pdf

Chaudhry, M. A., & Kazim, E. (2021). AI in Education (AIEd): a high-level academic and industry note 2021. *AI and Ethics*, 2(157-165). https://link.springer.com/article/10.1007/s43681-021-00074-z

Chen, J., Chen, C., & Liang, Z. (2016). Optimized TF-IDF algorithm with the adaptive weight of position of word. In *2nd International Conference on AI and Industrial Engineering (AIIE2016) Advances in Intelligent Systems Research* (Vol. 133).

Coelho, A. M. L., da Silva, H. F., da Silva, L. A. C., Andrade, M. E., & Rodrigues, R. G. da S. (2023). Inteligência artificial: Suas vantagens e limites em cursos à distância. *Revista Ilustração*, 4(2), 23-27. https://doi.org/10.46550/ilustracao.v4i2.150

Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Reviews Neuroscience*, *11*(3), 201-211. https://doi.org/10.1038/nrn2793

Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). *Mathematics for ML*. Cambridge University Press. https://mml-book.com

Dong, J. (2023). NLP pretraining language model for computer intelligent recognition technology. *ACM Transactions on Asian and Low-Resource Language Information Processing*. Retrieved from : https://dl.acm.org/doi/pdf/10.1145/3605210

European Commission. (2018). *Definition of AI*. High-Level Expert Group on AI (AI HLEG). Retrieved from https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf.

Farouk, M. (2019). Measuring sentences similarity: A survey. *Indian Journal of Science and Technology*, *12*(25). https://doi.org/10.17485/ijst/2019/v12i25/143977

Ferreira, R., André, M., Pinheiro, A., Costa, E., & Romero, C. (2020). Text Mining in Education. Retrieved from: https://arxiv.org/pdf/2403.00769

Gierl, M. J., & Lai, H. (2016). A process for reviewing and evaluating generated test items. *Educational Measurement: Issues and Practice*, *35*(4), 6-20. Retrieved from https://doi.org/10.1111/emip.12136.

Gierl, M. J., Zhou, J., & Alves, C. (2008). Developing a taxonomy of item model types to promote assessment engineering. *The Journal of Technology, Learning and Assessment*, *7*(2).

Göloğlu Demir, C., & Yılmaz, H. (2018). Sınıf dışı eğitim faaliyetlerinin öğrencilerin bilim ve teknolojiye yönelik tutumlarına etkisi ve duygu analizi. *İnsan ve Toplum Bilimleri Araştırmaları Dergisi*, *7*(5), 101-116. https://doi.org/10.15869/itobiad.483404

Grun, B., & Hornik, K. (2011). topicmodels: An R package for fitting topic models. Retrieved from https://cran.r-project.org/web/packages/topicmodels/vignettes/topicmodels.pdf

Guu, H., Hashimoto, T. B., & Oren, Y. (2018). Generating sentences by editing prototypes. *Transactions of the Association for Computational Linguistics*, *6*, 437-450. https://doi.org/10.1162/tacl_a_00030

Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis Campbell Systematic Reviews, 18, e1230. https://doi.org/10.1002/cl2.1230

Hamal, O., El Faddouli, N., Alaoui Harouni, M. H., & Lu, J. (2022). AI in Education. *Sustainability*, *14*(2862). https://doi.org/10.3390/su14052862

Hendler, J. (2008). Avoiding another AI winter. *IEEE Intelligent Systems*, *23*(2), 2-4.

Holstein, K., McLaren, B. M., & Aleven, V. (2018). Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. *AI in Education: 19th International Conference*, 154-168. https://doi.org/10.1007/978-3-319-93843-1_12

Hu, Z., Yang, Z., Shi, H., Tan, B., Zhao, T., He, J., Liang, X., Wang, W., Yu, X., Wang, D., Qin, L., Ma, X., Liu, H., Singh, D., Zhu, W., & Xing, E. P. (2018). Texar: A modularized, versatile, and extensible toolbox for text generation. *Proceedings of Workshop for NLP Open Source Software*, 13-22. https://doi.org/10.18653/v1/W18-2503

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90-95.

Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In E. Adar & P. Resnick (Eds.), *Proceedings of the eighth international AAAI* *conference on weblogs and social media*. Retrieved from https://ojs.aaai.org/index.php/ICWSM/issue/view/274

İlikhan, S., Özer, M., Tanberkan, H., & Bozkurt, V. (2024). How to mitigate the risks of deployment of AI in medicine? *Turkish Journal of Medical Sciences*, 54(3), 483-492. https://doi.org/10.55730/1300-0144.5814

Jurafsky, D., & Martin, J. H. (2024). *Speech and language processing: An introduction to NLP, computational linguistics, and speech recognition*.

Kasumba, R., & Neumman, M. (2024). Practical Sentiment Analysis for Education: The Power of Student Crowdsourcing. *Proceedings of the AAAI Conference on AI, 38*(21), 23110-23118. https://doi.org/10.1609/aaai.v38i21.30356

Kayalı, B., Balat, Ş., Kurşun, E., & Karaman, S. (2019). Lisansüstü eğitimde etkili ve nitelikli geribildirim. *Journal of Instructional Technologies & Teacher Education, 1*(8), 10-20.

Kış, A. (2019). Eğitimde yapay zeka. In *14. Uluslararası Eğitim Yönetimi Kongresi Tam Metin Bildiri Kitabı*.

Kort, B., Reilly, R., & Picard, R. W. (2001). An affective model of interplay between emotions and learning: Reengineering educational pedagogy—Building a learning companion. *Proceedings IEEE International Conference on Advanced Learning Technologies*, 43-46. https://doi.org/10.1109/ICALT.2001.943850

Kotlyarova, I. O. (2022). AI technologies in education. *Bulletin of the South*

Kühl, N., Goutier, M., Hirt, R., & Satzger, G. (2020). Machine learning in artificial intelligence: Towards a common understanding. *arXiv:2004.04686 [cs.LG]*. https://doi.org/10.48550/arXiv.2004.04686

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*.

Leskovec, J., Rajaraman, A., & Ullman, J. D. (2014). *Mining of Massive Datasets*. Cambridge University Press. Retrieved from http://infolab.stanford.edu/~ullman/mmds/book0n.pdf

Lewkow, N., Kode, S., Feild, J., Zimmerman, N., Riedesel, M., Essa, A., Boulanger, D., Seanosky, J., Kumar, V., & Kinshuk. (2016). A scalable learning analytics platform for automated writing feedback. In *Proceedings of the Third (2016) ACM Conference on Learning @ Scale - L@S '16*. https://doi.org/10.1145/2876034.2893380

Liddy, E. D. (2001). NLP. In *Encyclopedia of library and information science* (2nd ed.). Marcel Decker, Inc. Retrieved from https://surface.syr.edu/cgi/viewcontent.cgi?article=1043&context=istpub

Lin, F. (2023). Sentiment analysis in online education: An analytical approach and application. *Proceedings of the 2023 International Conference on ML and Automation*. Retrieved from https://doi.org/10.54254/2755-2721/33/20230225.

Lipnevich, A. A., & Panadero, E. (2021). A review of feedback models and theories: Descriptions, definitions, and conclusions. *Frontiers in Education, 6*. https://doi.org/10.3389/feduc.2021.720195

Lighthill, J. (1973). *Artificial intelligence: A general survey*. In *Artificial intelligence: A paper symposium* (pp. 1-77). Science Research Council. https://www.aiai.ed.ac.uk/events/lighthill1973/lighthill.pdf

Lu, C., & Cutumisu, M. (2021). *Integrating deep learning into an automated feedback generation system for automated essay scoring*. Paper presented at the International Conference on Educational Data Mining (EDM). International Educational Data Mining Society. https://files.eric.ed.gov/fulltext/ED615567.pdf

Lubis, F. F., Mutaqin, A. P., Waskita, D., Sulistyaningtyas, T., Arman, A. A., & Rosmansyah, Y. (2021). Automated short-answer grading using semantic similarity based on word embedding. *International Journal of Technology, 12*(3), 571-581.

Mahesh, B. (2018). ML algorithms - A review. *International Journal of Science and Research (IJSR), 9*(1). https://www.ijsr.net/archive/v9i1/ART20203995.pdf

Mazidi, K. (2018). Automatic Question Generation From Passages. In A. Gelbukh (Ed.), *CICLing 2017, LNCS 10762* (pp. 655-665). Springer. https://doi.org/10.1007/978-3-319-77116-8_49

Merdun, G., Okçular, E., Altınok, D., & Akkurt, F. (2024). *Turkish NLP Resources*. GitHub repository. Retrieved July 10, 2024, from https://github.com/agmmnn/turkish-nlp-resources

McCarthy, J. (2007). What is AI? Computer Science Department, Stanford University. Retrieved from http://www-formal.stanford.edu/jmc/whatisai.pdf

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1956). A proposal for the Dartmouth summer research project on AI. Retrieved from http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf

McKinney, W. (2010). Data analysis in Python. *Proceedings of the 9th Python in Science Conference* (pp. 51-56).

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Mulianingsih, F., Anwar, K., Shintasiwi, F. A., & Rahma, A. J. (2020). AI dengan Pembentukan Nilai dan Karakter di Bidang Pendidikan. Ijtimaiya: Journal of Social Science Teaching, 4(2), 148-154. Retrieved from http://journal.stainkudus.ac.id/index.php/Ijtimaia

Mohler, M., & Mihalcea, R. (2009). Text-to-text semantic similarity for automatic short answer grading. *In Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, 567–575. Association for Computational Linguistics.

Nafea, I. T. (2016). ML in educational technology. In ML- Advanced techniques and emerging applications. IntechOpen. http://dx.doi.org/10.5772/intechopen.72906

Newell, A., & Simon, H. A. (1956). The logic theory machine. *IRE Transactions on Information Theory, 2*(3), 61-79.

Nilsson, N. J. (1984). Shakey the robot. SRI International. Retrieved from https://www.sri.com/publication/artificial-intelligence-pubs/shakey-the-robot-pub/

Nkechi, A. A., Ojo, A. O., & Eneh, O. A. (2024). Impact of AI in Achieving Quality Education. IntechOpen. https://doi.org/10.5772/intechopen.1004871

Oflazer, K. (2016). Türkçe ve doğal dil işleme. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 5*(2). https://dergipark.org.tr/tr/pub/tbbmd/issue/22245/238795

Oflazer, K., & Saraçlar, M. (2018). *Turkish NLP*. Springer.

OpenAI. (2023). GPT-4 Turbo and GPT-4. OpenAI. https://platform.openai.com/docs/models/gpt-4-turbo-and-gpt-4

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830. https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

Peña-Torres, J. A. (2024). Towards an improved teaching practice using sentiment analysis in student evaluation. *Ingeniería y Competitividad, 26*(2), e-21013759. Retrieved from: https://www.researchgate.net/publication/381598280_Towards_an_improved_of_teaching_practice_using_Sentiment_Analysis_in_Student_Evaluation.

Perveen, A. (2021). Use of word clouds for task-based assessment in asynchronous e-language learning. *MEXTESOL Journal, 45*(2).

Ramachandran, D., & Rana, R. S. (2024). AI for legal system: Jurisprudence in the digital age. *International Journal of Advanced Academic Studies, 6*(5), 03-13. https://doi.org/10.33545/27068919.2024.v6.i5a.1158

Russell, S., & Norvig, P. (2010). *AI: A Modern Approach* (3rd ed.). Pearson.

Sadiku, M. N. O., Ashaolu, T. J., Ajayi-Majebi, A., & Musa, S. M. (2021). AI in education. *International Journal of Scientific Advances, 2*(1), 1-11. https://typeset.io/pdf/artificial-intelligence-in-education-5ggabmq2kf.pdf

Sak, H., Güngör, T. & Saraçlar, M. (2011) Resources for Turkish morphological processing. *Lang Resources & Evaluation* **45**, 249–261. https://doi.org/10.1007/s10579-010-9128-6

Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781107298019

Sheikh, H., Prins, C., & Schrijvers, E. (2023). *Mission AI: The new system technology*. Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-031-21448-6

Shin, E. (2021). *Automated item generation by combining the non-template and template-based approaches to generate reading inference test items* (Doctoral dissertation, University of Alberta). Department of Educational Psychology.

Singley, M. K., & Bennett, R. E. (2002). Item generation and beyond: Applications of schema theory to mathematics assessment. In S. H. Irvine & P. C. Kyllonen (Eds.), *Item generation for test development* (pp. 361–384). Routledge. https://doi.org/10.4324/9781410602145

Silge, J., & Robinson, D. (2017). *Text mining with R: A tidy approach*. O'Reilly Media.

Smith, B. I., Chimedza, C., & Bührmann, J. H. (2022). Individualized help for at-risk students using model-agnostic and counterfactual explanations. *Educational and Information Technologies, 27*(2), 1539–1558. https://doi.org/10.1007/s10639-021-10661-6

Smith, J., Li, H., & Patel, R. (2024). Automated generation of multiple-choice cloze questions for assessing English vocabulary using GPT-turbo 3.5. *arXiv preprint arXiv:2403.02078*. https://arxiv.org/abs/2403.02078

Sukmana, R., & Rusydiana, A. S. (2023). Social media sentiment analysis on waqf and education. *Islamic Marketing Review*, 2(2). Retrieved from http://journals.smartinsight.id/index.php/IMR

Sytnyk, L., & Podlinyayeva, O. (2024). AI in education: Main possibilities and challenges. In Proceedings of the 8th International Scientific and Practical Conference "International Scientific Discussion: Problems, Tasks and Prospects" (pp. 569-579). Brighton, United Kingdom. https://doi.org/10.51582/interconf.19-20.05.2024.058

Şeker, S. E. (2015). Metin madenciliği (Text mining). *YBS Ansiklopedi*, *2*(3). https://ybsansiklopedi.com/wp-content/uploads/2015/08/MetinMadenciligi30_32.pdf

Tsiakmaki, M., & Ragos, O. (2021). A case study of interpretable counterfactual explanations for the task of predicting student academic performance. *2021 25th International Conference on Circuits, Systems, Communications and Computers (CSCC)*, 120-125. https://doi.org/10.1109/CSCC53858.2021.00029

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433-460. Retrieved From: https://phil415.pbworks.com/f/TuringComputing.pdf

Ural State University. Ser. Education. Educational Sciences, 14(3), 69-82. https://vestnik.susu.ru/ped/article/view/12330

Usta, Y. (2024). *Awesome Turkish NLP*. GitHub repository. Retrieved July 10, 2024, from https://github.com/yusufusta/awesome-turkish-nlp

Uysal, İ. (2019). *Açık uçlu maddelerde otomatik puanlamanın güvenirliği ve test eşitleme hatalarına etkisi* (Doktora tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü). YÖK Ulusal Tez Merkezi. https://tez.yok.gov.tr/UlusalTezMerkezi/tezDetay.jsp?id=55GArTnn6vLwQ3HOxnwo_w&no=gj27xzLBIdoGFgSUJzjT6Q

Wang, J., & Dong, Y. (2020). Measurement of Text Similarity: A Survey. *Information, 11*(9), 421. doi:10.3390/info11090421

Waskom, M. (2020). *Seaborn: Statistical data visualization*. Erişim adresi: https://seaborn.pydata.org

Weizenbaum, J. (1966). ELIZA—A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36-45.

Wijeratne, Y., Silva, N., & Shanmugarajah, Y. (2009). NLP for government: Problems and potential. Retrieved from https://lirneasia.net/wp-content/uploads/2019/04/Natural_Language_Processing_for_Government__Problems_and_Potential.pdf

Woods, B., Adamson, D., Miel, S., & Mayfield, E. (2017). Beyond Automated Essay Scoring: Forecasting and Improving Outcomes in Middle and High School Writing. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '17)* (pp. 2071-2080). ACM. https://doi.org/10.1145/3097983.3098160

Yang, A., & Halim, S. (2022). Natural language generation using ML techniques. *Journal of Student Research, 11*(2). Retrieved from https://typeset.io/papers/natural-language-generation-using-machine-learning-1itplfnn.

Yıldırım, S., & Yıldız, T. (2018). A Comparison of Different Approaches to Document Representation in Turkish Language. *Journal of Natural and Applied Sciences, 22*(2), 569-576. https://doi.org/10.19113/sdufbed.15893

Young, J. (2024). The rise of AI in education. *International Journal of Innovative Research & Development, 13*(2), 74. https://www.internationaljournalcorner.com/index.php/ijird_ojs/article/view/173518/118319

Zeinalipour, K., Keptiğ, Y. G., Maggini, M., & Gori, M. (2024). Automating Turkish educational quiz generation using large language models. https://doi.org/10.48550/arXiv.2406.03397

Zeyrek, A. (2020). *Zeyrek: Morphological analysis for Turkish*. GitHub. https://github.com/ahmetaa/zeyrek

Appendix

Appendix 1.

Examples of Text Visualization

Sample application and code example for text visualisation examples are given below.


```
# Kelime listelerini oluşturma
dogru_kelimeler = set(dogru_cevaplar.split())
yanlis_kelimeler = set(yanlis_cevaplar.split())

# Ortak kelimeleri çıkartma
dogru_kelimeler = dogru_kelimeler - yanlis_kelimeler
yanlis_kelimeler = yanlis_kelimeler - dogru_kelimeler

# Güncellenmiş metinler
dogru_cevaplar = ' '.join(dogru_kelimeler)
yanlis_cevaplar = ' '.join(yanlis_kelimeler)
```

Appendix 2.

Examples of Sentence Similarity

Examples of sentence similarity sample application and code example are given below.

```
cümle1 = "İstanbul 1453 yılında fetholundu."
cümle2 ="Fatih Sultan Mehmet İstanbulu 1453 yılında aldı."
benzerlik oranı:([[0.7761]])
cümle1 = "Atatürk 19 Mayıs 1919'da Samsuna'a ayak baktı."
cümle2 ="19 Mayıs 1919 Atatürk'ün Samsun'a gittiği tarihtir."
benzerlik oranı:([[0.9192]])
```

Appendix 3.

Examples of Sentence Similarity

Student Feedback sample text, application and code example are given below.

Sample Text

Çocukluk arkadaşlarının hafızasında kalan anılar, zamanla solmayan nadir hazinelerdir. Oyunlar, gülüşmeler, hatta küçük kavgalar bile yıllar geçse de unutulmaz. Ahmet ve Mehmet, çocukluklarını aynı mahallede, aynı sokakta geçirmiş iki dosttu. İki arkadaş, her akşam saatlerce sokakta top oynar, macera ararlardı. Bir gün, kocaman bir ağaç gördüler. Gövdesi sağlam, dalları uzanmıştı gökyüzüne. Ahmet, hemen tırmanmaya başladı. Mehmet, cesareti topladı ve arkadaşının ardından ağaca tırmandı. Birlikte tepesine çıktıklarında, etraflarını seyrettiler. Küçük mahalleleri, yeşillikler içindeki parkı, uzaklarda görünen caminin minaresini görebiliyorlardı. O an, ikisi de birbirine gülümsedi. Bu anı, yıllar geçse de unutulmayacak anılardan biri olacaktı.

Question: Metinde Ahmet ve Mehmet'in en unutulmaz anısını sizce ne yaratmış olabilir?

Appendix Table 1.

Example of Student Responses and Feedback

Response	Cluster	Feedback
1. Ağacın tepesindeki sessizliğin ve huzurun tadını çıkarmaları.	2	Tebrikler. Daha fazla okuma çalışması yaparak yorumlama yeteneğini artırabilirsin. Bu noktada okuman için .... kitaplarını önerebilirim.
2. Ahmet'in ağacın tepesinden bulutların şekillerini tahmin etmeye çalışması ve Mehmet'in onunla bakması.	1	Metni tekrar okumanı ve yeni çıkarımlar yapmanı önerebilirim. Konuyu farklı yönleriyle ele almanın yorumlama yeteneğini geliştireceğini umuyorum.
3. İki arkadaşın ağacın tepesindeyken birlikte yıldızları saymaları ve hangi yıldızın hangi burcu temsil ettiğini konuşmaları.	2	Tebrikler. Daha fazla okuma çalışması yaparak yorumlama yeteneğini artırabilirsin. Bu noktada okuman için .... kitaplarını önerebilirim.
4. Ağacın tepesinde otururken Ahmet'in eline düşen yaprağı yakalamaya çalışması ve Mehmet'in ona yardım etmesi.	1	Metni tekrar okumanı ve yeni çıkarımlar yapmanı önerebilirim. Konuyu farklı yönleriyle ele almanın yorumlama yeteneğini geliştireceğini umuyorum.
5. İki arkadaşın ağacın tepesinden çevredeki ağaçlardaki kuşların cinslerini tahmin etmeye çalışmaları.	2	Tebrikler. Daha fazla okuma çalışması yaparak yorumlama yeteneğini artırabilirsin. Bu noktada okuman için .... kitaplarını önerebilirim.
6. Tepeden gördükleri manzarayı birlikte betimleyerek bir hikaye oluşturmaya çalışmaları.	2	Tebrikler. Daha fazla okuma çalışması yaparak yorumlama yeteneğini artırabilirsin. Bu noktada okuman için .... kitaplarını önerebilirim.
7. Ahmet'in ağacın tepesinden aşağıya bir şey atmaya cesaret edememesi ve Mehmet'in onu desteklemesi.	1	Metni tekrar okumanı ve yeni çıkarımlar yapmanı önerebilirim. Konuyu farklı yönleriyle ele almanın yorumlama yeteneğini geliştireceğini umuyorum.
8. İki arkadaşın ağaçta otururken geçmişte yaşadıkları komik anıları hatırlamaları.	3	Keyifli bir yorum getirmişsin. Konuya hakim görünüyorsun. İnsan yaşam boyu öğrencidir. Yeni şeyler öğrenmekten hep keyif alman dileğiyle.

9. İki arkadaşın ağaçta otururken bulutların hareketini izleyerek hayal kurmaları.	3	Keyifli bir yorum getirmişsin. Konuya hakim görünüyorsun. İnsan yaşam boyu öğrencidir. Yeni şeyler öğrenmekten hep keyif alman dileğiyle.

In this study, feedback was generated by clustering responses based on answer similarities.

```
df = pd.read_excel('gptcevaplar.xlsx')
Executed at 2024.03.20 14:35:21 in 44ms

df.head(10)
Executed at 2024.03.20 14:35:22 in 34ms

10 rows    10 rows × 2 columns pd.DataFrame
   No  Cevap
1   2  Birlikte oynadıkl…
2   3  İki arkadaşın bir…
3   4  Gördükleri manzar…
4   5  Ahmet'in cesaretl…
5   6  Top oynamak için …
6   7  Mahallelerini, pa…
7   8  Arkadaşlık bağlar…
8   9  İki arkadaşın bir…
```

```
1 – remove_stopwords -> Stopwords kaldırıldı.
2 – remove_numbers -> Rakamlar kaldırıldı.
3 – remove_punctuations -> Her türlü noktalama işareti kaldırıldı.
4 – lower_case -> Tüm veri küçük harfe çevrildi.

Ref: https://pypi.org/project/mintlemon-turkish-nlp/

df['Cevap'] = df['Cevap'].apply(Normalizer.remove_stopwords)
df['Cevap'] = df['Cevap'].apply(Normalizer.remove_numbers)
df['Cevap'] = df['Cevap'].apply(Normalizer.remove_punctuations)
df['Cevap'] = df['Cevap'].apply(Normalizer.lower_case)
Executed at 2024.03.20 14:35:29 in 56ms
```

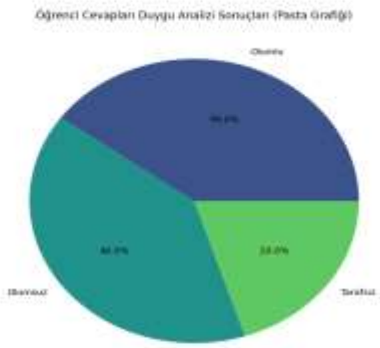
```
num_clusters = 3
kmeans = KMeans(n_clusters=num_clusters)
kmeans.fit(cevaplar_vec)

labels = kmeans.labels_

feedbacks = {
  0: "Metni tekrar okumanı ve yeni çıkarımlar yapmanı önerebilirim. Konuyu farklı yönleriyle ele almanın yorumlama yeteneğini geliştireceğini umuyorum.",
  1: "Tebrikler. Daha fazla okuma çalışması yaparak yorumlama yeteneğini artırabilirsin. Bu noktada okuman için .... kitaplarını önerebilirim.",
  2: "Keyifli bir yorum getirmişsin. Konuya hakim görünüyorsun. İnsan yaşam boyu öğrencidir. Yeni şeyler öğrenmekten hep keyif alman dileğiyle"
}
for i, label in enumerate(labels):
  cluster_feedback = feedbacks[label]
  df.at[i, 'Cluster_Label'] = f'Cluster_{label + 1}'
  df.at[i, 'Cluster_Feedback'] = cluster_feedback
```

Appendix 4.
Example of Sentiment Analysis
An example of the application and code for emotion analysis is given below.

Appendix Table 2.
Sentiment Analysis Results

Response	Sentiment Analysis Result
Matematik dersini çok seviyorum çünkü problem çözmek bana zevk veriyor.	Olumlu
Matematik dersi benim için zorlayıcı ve sıkıcı, bu yüzden sevmiyorum.	Olumsuz
Matematikte başarılı olmak benim için önemli, bu yüzden çok çalışıyorum.	Tarafsız
Matematik dersinde kendimi yetersiz hissediyorum ve bu beni üzüyor.	Olumsuz
Matematik öğretmenimiz konuları çok iyi anlatıyor, bu yüzden matematik dersini seviyorum.	Olumlu
Matematik dersinde zorlandığım için sık sık stres oluyorum.	Olumsuz
Matematik, gelecekteki kariyerim için önemli olduğundan, bu dersi dikkatle takip ediyorum.	Tarafsız
Matematik problemlerini çözdükçe kendime güvenim artıyor.	Olumlu
Matematik dersleri bana çok karmaşık geliyor ve bu da motivasyonumu düşürüyor.	Olumsuz
Matematikte yeni şeyler öğrenmek beni heyecanlandırıyor.	Olumlu

```
If num_positive ≥ num_negative:
   sentiment = "Olumlu"
elif num_negative > num_positive:
   sentiment = "Olumsuz"
else:
   sentiment = "Tarafsız"
print("Duygusal ton:", sentiment)
Duygusal ton: Tarafsız
```



Appendix 5.
Example of Item Generation
Below are examples of open-ended and multiple-choice questions generated using GPT-4 and Gemini.

Text: Küçük Prens ve Tilki

Bir zamanlar küçük prens bir gezegende yaşar. Bu gezegenin herhangi bir yerinde, özel bir çölde, altı çalı çırpıyla kaplı bir yere gömülmüş bir yıldız var. Bir gün küçük prens, çölde bir tilki ile karşılaşır. Tilki, küçük prensin onu evcilleştirmesini ister. Küçük prens, tilkinin ne demek istediğini anlamaya çalışır. Tilki, "Evcilleştirme, özlemin derecesine bağlıdır. Benimle ilgilenirsen, benimle dost olursan, benimle oyun oynarsan, senin için çok farklı bir ışık oluşur. Yıldızları seyretmek güzel olur. İnsanlar, bir yıldızı seyrettiğinde, senin yıldızında olduğun saatte gülümseyeceklerdir." der. Küçük prens tilkiyi evcilleştirmeyi kabul eder. Tilki, küçük prense insanların ne anlama geldiğini anlatır. Ona göre insanlar, birbirinden farklı olan gülüşlerdir. Tilki, küçük prense insanların onları evcilleştirenlerdir. Evcilleştirmek ise, birbirine alıştırmaktır. Küçük prens, tilkinin anlattıklarını düşünür ve onunla dost olur.

Appendix Table 3.
Open-Ended Questions Generated with GPT-4o

Question Number	Question
1	Tilki, Küçük Prens'e "evcilleştirilmek" terimini nasıl açıkladı?
2	Küçük Prens ve tilki neden her gün aynı saatte buluştu?
3	Tilki, Küçük Prens'in gezegenine geri döneceğini öğrendiğinde nasıl hissetti ve neden?
4	Küçük Prens, tilkiyi evcilleştirmenin sonunda hangi önemli dersi öğrendi?
5	Tilki'nin, "İnsan ancak yüreğiyle baktığında doğruyu görebilir. Gözler gerçeği göremez." sözü ne anlama gelir?

Appendix Table 4.
Examples of Multiple-Choice Questions Generated with GPT-4o

Question	Options	Correct Answer
KüçükPrens Dünya'da kiminle karşılaştı?	a) Tilki b) Yılan c) Kral d) İş adamı	a) Tilki
Tilki, Küçük Prens'e ne olmayı teklif etti?	a) Düşman b) Öğretmen c) Arkadaş d) Rehber	c) Arkadaş
Küçük Prens tilkiyi evcilleştirmek için ne yaptı?	a) Onu besledi b) Onunla konuştu c) Onunla oyun oynadı d) Onun yanına her gün aynı saatte gitti	d) Onun yanına her gün aynı saatte gitti
Tilki, Küçük Prens'e neyi öğretti?	a) Zamanın değerini b) Arkadaşlığın önemini c) Cesaretin gücünü d) Bilginin değerini	b) Arkadaşlığın önemini
Tilki, Küçük Prens'e veda ederken ne dedi?	a) "Seni unutmayacağım." b) "Her zaman buğday tarlalarına bakacağım ve seni hatırlayacağım." c) "Geri dön." d) "Seni seviyorum."	b) "Her zaman buğday tarlalarına bakacağım ve seni hatırlayacağım."

Appendix Table 5.
Open-Ended Questions Generated with Gemini-1.5-Flash

Question
1. "İnsanlar, yaşadıkları dünyayı anlamak için sürekli yeni keşiflere ve öğrenmelere ihtiyaç duyarlar." cümlesini kendi cümlelerinizle açıklayınız. Bu cümle size ne düşündürüyor?
2. Bir yazarın romanında, kahramanın yaşadığı zorlukları ve bunlarla başa çıkma yollarını ele aldığını düşünün. Sizce yazar bu romanı neden yazmış olabilir? Bu soruyu yanıtlamak için romanın konusu ve kahramanın özellikleri hakkında tahminlerde bulunabilirsiniz.
3. "Herkesin bir hikayesi vardır." sözüyle ne anlatılmak isteniyor olabilir? Bu sözü destekleyen kendi yaşamınızdan bir örnek veriniz.
4. "Teknoloji geliştikçe, insanlar arasındaki iletişim biçimleri de değişiyor." cümlesini ele alarak, teknolojinin iletişim üzerindeki olumlu ve olumsuz etkilerinden bahsediniz.
5. Bir şiirde, yazarın doğayı, insanları veya duyguları nasıl anlattığını düşünüyorsunuz? Bir şiir örneği vererek, şiirde kullanılan dilin ve imgelerin okuyucunun duygularını nasıl etkilediğini açıklayınız.

Appendix Table 6.

Examples of Multiple-Choice Questions Generated with Gemini-1.5-Flash

Question	Response
"Ağaçlar, doğanın akciğerleridir." cümlesi aşağıdaki hangi anlamı taşır?	a) Ağaçlar, doğada önemli bir rol oynar b) Ağaçlar, insanlara oksijen sağlar. c) Ağaçlar, doğanın güzelliğini temsil eder. d) Ağaçlar, insanlara gölge yapar.
"Yazarın romanında, kahramanın yaşadığı zorluklar ve bunlarla başa çıkma yolları ele alınıyor." cümlesine göre, romanın ana teması nedir?	a) Aşk ve romantizm b) Savaş ve kahramanlık c) Yaşamın zorlukları ve üstesinden gelme d) Teknoloji ve gelecek
"İnsanlar, yaşadıkları dünyayı anlamak için sürekli yeni keşiflere ve öğrenmelere ihtiyaç duyarlar." cümlesinde vurgulanan ana fikir nedir?	a) İnsanlar meraklıdır. b) İnsanlar öğrenmeyi sever. c) İnsanlar bilginin önemini anlar. d) İnsanlar sürekli gelişen bir dünyada yaşar.
"Teknoloji geliştikçe, insanlar arasındaki iletişim biçimleri de değişiyor." cümlesinde, teknolojinin iletişim üzerindeki etkisi nasıl gösteriliyor?	a) Teknoloji, iletişimi daha hızlı ve kolay hale getiriyor. b) Teknoloji, iletişimi daha kişisel hale getiriyor. c) Teknoloji, iletişimi daha karmaşık hale getiriyor. d) Teknoloji, iletişimi daha az etkili hale getiriyor.
"Herkesin bir hikayesi vardır." sözüyle ne anlatılmak isteniyor olabilir?	a) Her insanın yaşamında önemli olaylar vardır. b) Her insanın benzersiz bir yaşam hikayesi vardır. c) Her insanın anlatmaya değer bir hikayesi vardır. d) Her insanın başkalarına anlatabileceği bir hikayesi vardır.