Hybrid Approach for AI-Generated Text Detection Using Classical Machine Learning and Transformer-Based Models

Syed Ar Rafi
Department of Computer Science
BRAC University
Dhaka, Bangladesh
syed.ar.rafi@g.bracu.ac.bd

Adiba Islam Khan
Department of Computer Science
BRAC University
Dhaka, Bangladesh
adiba.islam.khan@g.bracu.ac.bd

MD. Atiq Mahbub
Department of Computer Science
BRAC University
Dhaka, Bangladesh
atiq.mahbub@g.bracu.ac.bd

***Abstract*—In the era of increasingly advanced large language models (LLMs), differentiating between textual content generated by AI and written by humans has become both a technological difficulty and an ethical requirement. The purpose of this study is to enhance academic integrity, AI accountability, and digital content management by examining the effectiveness of several machine learning techniques in classifying text origin. We demonstrate a complete supervised learning pipeline that consists of transformer-based models like DistilBERT and RoBERTa, data preprocessing, TF-IDF-based feature extraction, and classification using Logistic Regression and Naive Bayes. For training and evaluation, a large labeled dataset including more than 480,000 samples was utilized. Our findings show that among traditional machine learning techniques, Logistic Regression and Naive Bayes, especially Logistic Regression, provide competitive performance with better interpretability, but the transformer-based model, RoBERTa, outperformed all the other models with an accuracy of 0.9957. This work provides useful information and standards for the continued advancement of reliable AI detection systems that can be used in a variety of fields.**

I. Introduction

Recently, there has been a blurred boundary between human-generated content and AI-generated content [1]. Artificial intelligence systems are now capable of generating text-based content that closely mimics human writing style with the rapid development of large language models (LLMs) such as OpenAI's GPT series, Gemini, and Meta's LLaMA [2]. They are raising concerns about authenticity, authorship and the potential risk of misusing AI to generate content for academic, journalistic, and social media platforms. As artificial intelligence-generated content becomes increasingly prevalent in various fields, the need for a reliable method to differentiate between text created by humans and that produced by AI has emerged as both a crucial technical challenge and an ethical responsibility.

The primary motivation for this work is the growing community, educational, and regulatory concerns about the seamless integration of content generated by artificial intelligence. Undoubtedly, AI has several positive aspects for us, such as improved creativity and productivity. Nevertheless, if employed improperly, it can compromise academic credibility, manipulate public opinion with misleading data, or promote widespread plagiarism. People from a multitude of professions, including professors, publishers, professionals in law, and even social media platforms, are on the lookout for tools that can assist in identifying AI-generated content to ensure transparency and accountability. However, identifying AI and human-generated content is a great challenge. Large datasets of human text are used to train language models, which allow them to reproduce syntax accurately. When dealing with texts generated through complex algorithms, traditional detection techniques, such as keyword matching or rule-based heuristics, are often insufficient. Therefore, a more sophisticated machine learning-based technique is required to determine minor statistical and structural characteristics that distinguish AI-generated text from human-written content.

RQ1: *How effectively can ML models and transformer based models classify AI-generated text vs. human-written text?*
RQ2: *How can the combination of classical machine learning models and transformer-based models improve the accuracy of classifying AI-generated text versus human-written text?*

This study works on such approaches using machine learning models and techniques, which promotes the expanding fields of digital investigation, natural language processing, and AI ethics, as well as -

- *Quantitative Insight:* Proposes visual and statistical insights into the linguistic patterns that separate AI texts from humans.
- *Practice Application*: Offers a classifier that can be modified for practical uses in digital publishing, education, and content moderation.

II. Related Work

Powerful generative language models like GPT-3 [3], GPT-4 [4], and other transformer-based systems have emerged, generating an upsurge in academic research in the identification and categorization of text produced by artificial intelligence [5]. Differentiating between machine-generated and human-written content has become a major concern in a variety of sectors, from journalism [6] and education [7] to security and digital ethics, as these models' outputs become increasingly human-like.

The general stylistic and linguistic characteristics of sentence length, vocabulary richness, syntactic variation, and punctuation use were the main focus of early attempts to identify the authorship of texts [8]. AI models with limited stylistic variation, such as GPT-2 [9], allowed this rule-based system to detect them, but quickly lost ground as it struggled with more sophisticated AI models [10]. The researchers were able to create classifiers based on engineering features such as TF-IDF scores, POS tags, and lexical data due to the move towards supervised learning [11]. Studies by Eilifsen et al. (2020) [12] and Yan et al. [13] demonstrated that while trained on high-quality labeled datasets logistic regression can potentially identify AI-generated texts with reasonable accuracy. Zeng et al. [14] have also mentioned that logistic regression and Naive Bayes can perform well in text classification. Although these conventional models provided interpretability and transparency, they did not resist deceptive or adverse data [15]. A recent study by Fariello et al. [16] demonstrates various detection methods, including watermarking, feature-based, neural-based, hybrid, and human-aided approaches, as well as several significant shortcomings in previous models, such as their heavy reliance on surface-level inputs and lack of domain flexibility. Their study also emphasizes the necessity of continual enhancement to keep up with the evolution of LLM capabilities [17].

Deep learning architectures such as RoBERTa, DeBERTa, and DistilBERT, which are frequently optimized for binary classification, have been used in more recent approaches [18], [19]. These methods frequently outperform traditional classifiers and can identify subtle contextual signals [20]. In the past, tools like OpenAI's AI Text Classifier, which is no longer in use, along with research models like DetectGPT, aimed to provide generalized solutions for detection [21]. But dependability issues, particularly when dealing with brief texts or translated content, are leading to a critical reevaluation of these techniques. In their integrative hybrid review, Chaka et al. [22] evaluated the efficiency of current AI detection technologies and highlighted their inconsistent results across a range of datasets and writing styles. The 2024 SemEval competition includes a shared challenge to differentiate between writing produced using the RoBERTa Base model and fine-tune it on a broad dataset including various subjects, languages, and sources [23]. The challenge's results confirmed the problem's complexity, especially when considering prompt-engineered and instruction-tuned AI models. Furthermore, multimodal and metadata-aware detection frameworks have been developed by authors such as Elkhatat et al. to increase adaptability in real-world contexts, indicating that language features alone may no longer be sufficient [24].

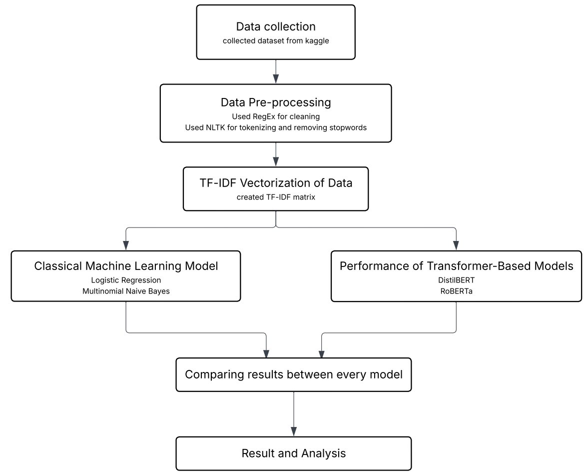

Fig. 1: Theoretical Framework of the proposed system.

III. METHODOLOGY

This study adopts a supervised machine learning approach to differentiate human- and AI-generated text using textual characteristics. It consists of core stages, such as data collection and handling, Exploratory Data Analysis (EDA), text preprocessing, feature extraction and vectorization, and model training. Each step is sincerely structured to ensure the reproducibility, transparency, and solidity of the classification pipeline. The full method of our research can be found in Fig 1.

A. Data Collection and Handling

A labeled dataset of 487,235 unique texts (305,797 human, 181,438 AI) was selected from a public repository, *Kaggle* [25]. 0 for human and 1 for AI; such binary labels have been designated to associate with each text instance. Pandas processed the data, assuring its integrity by examining it for formatting errors, missing values, and null values.

B. Exploratory Data Analysis (EDA)

An extensive exploratory data analysis has been conducted to develop the characteristics of the dataset. This included visualization tools such as kernel density plots (KDE), histograms, and word clouds, as well as class distribution checks using bar charts and text length comparisons Fig 4 5 using word clouds. Fig 2 and Fig 3 respectfully display the commonly used words in AI generated texts and human written texts. In addition to highlighting each class's frequently used terms, these visualizations assisted in identifying unique patterns. Thus, this analysis helped us to make preprocessing decisions.

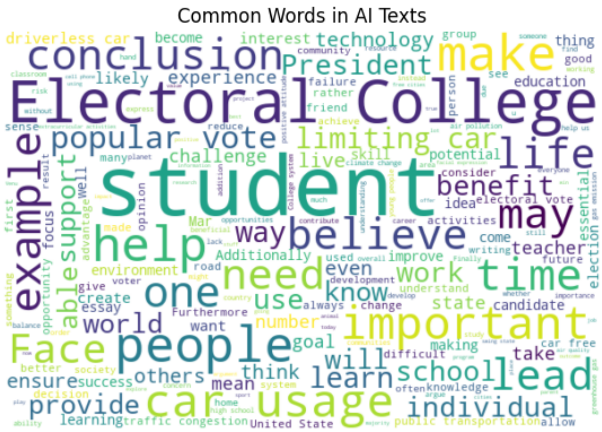

Fig. 2: Common Used Words in AI Generated texts

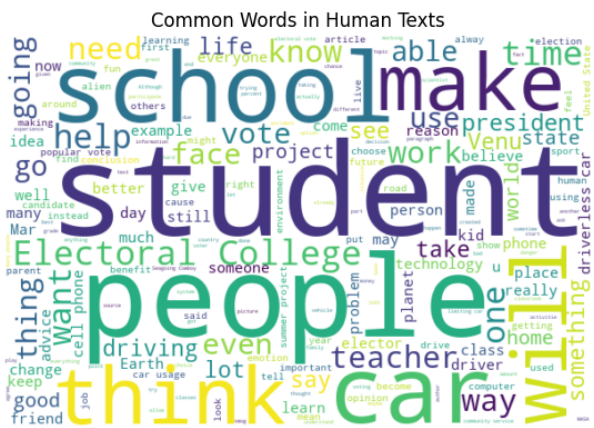

Fig. 3: Common Used Words in human written texts

C. Text Preprocessing and Cleaning

In this text preprocessing phase, the textual inputs are converted into logical numerical representations, essential for natural language processing (NLP) tasks. Texts were normalized through lowercase conversion and RegEx removal of HTML, hyperlinks, numerals, and punctuation. NLTK was used for tokenization and stopword elimination: 'the, 'is,' 'and'. A new column called "clean text" was created by joining cleaned tokens to single-line text. Thus, the reduction of redundancy and noise has been ensured systematically.

D. Feature Extraction and Vectorization

We have implemented the Term Frequency-Inverse Document Frequency (TF-IDF) [26] vectorization method to transform cleaned textual data into a numeric form to improve computational learning. Here, the vocabulary was confined to the top 20000 terms by frequency, and both unigrams (single words) and bigrams (two-word phrases) were included to capture more contextual information. To mitigate the impact of frequent terms, stopwords were removed, and sublinear term frequency scaling was applied. The generated TF-IDF matrix was applied as the model training input.

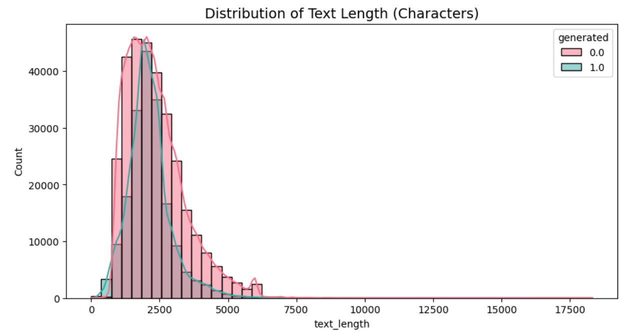

Fig. 4: Distribution of character length in text

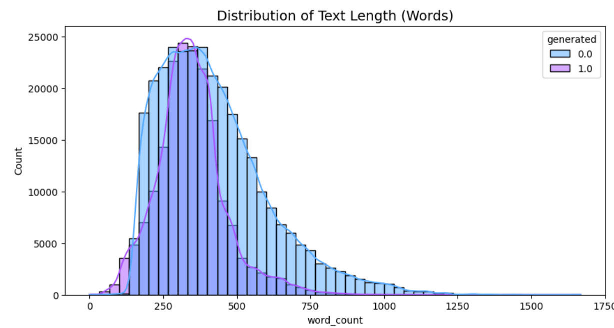

Fig. 5: Distribution of word length in text

E. Train-Test Splitting and Preparation for Modeling

The dataset has been separated into training and testing subsets using an 80/20 stratified split as the final action in the methodological pipeline. The binary class label (produced as 0 for humans and 1 for AI) was the intended outcome variable, while the TF-IDF matrix was the input feature set. The test set was set aside for evaluation of performance on unseen data, whereas the training set was utilized to adapt the machine learning models.

F. Machine Learning Models

In this study, we employed supervised machine learning models to classify text as either human-generated or AI-generated. The models used in this research were Logistic Regression and Multinomial Naive Bayes (MNB), both of which are well-established techniques in text classification tasks. These models were trained using previously generated TF-IDF data.

*1) **Logistic Regression**:* Logistic regression is a statistical method widely used for binary classification problems [27]. It is a linear model that works well with high-dimensional data, making it suitable for text classification tasks, especially when combined with TF-IDF vectorization [28]. In this study, logistic regression was trained using the TF-IDF matrix as input, where the features represent the importance of words and word pairs (bigrams) in distinguishing between human and AI texts. The target labels were binary, with 0 for human-generated texts and 1 for AI-generated texts.

The Logistic Regression model was trained using the sparse matrix obtained from TF-IDF. A maximum iteration of 1000 was set to ensure convergence.

*2) **Multinomial Naive Bayes (MNB)**:* Multinomial Naive Bayes is a probabilistic classifier based on applying Bayes' theorem with the "naive" assumption of conditional independence between every pair of features given the class label. It models the distribution of word frequencies in the documents and is known for its simplicity, efficiency, and strong performance in natural language processing tasks where feature vectors are counts or TF-IDF scores. This vectorization helps capture the importance of specific words and word pairs in distinguishing between human and AI-generated texts. TF-IDF feature matrix extracted from the training dataset, which was split from the original dataset using an 80/20 stratified train-test split to maintain class balance.

G. Transformer-based Models

This study also explored the use of transformer-based models, specifically DistilBERT and RoBERTa, for text classification. Transformer models, which have demonstrated exceptional performance in a variety of NLP tasks, were chosen to leverage their ability to understand contextual relationships within the text, which is critical in distinguishing between human-generated and AI-generated text. These models, based on the BERT (Bidirectional Encoder Representations from Transformers) architecture, are pre-trained on large corpora and fine-tuned on specific tasks like classification.

*1) **DistilBERT**:* DistilBERT [29] is a smaller, more efficient version of the BERT model, designed to retain 97% of BERT's language understanding capabilities while being faster and more memory-efficient. In this study, DistilBERT was fine-tuned for binary text classification. The DistilBERT model was fine-tuned on the pre-trained weights, specifically adapted for text classification with the output layer modified to predict a binary classification task. We utilized the Hugging Face transformers library to load the pre-trained model and tokenizer, enabling us to efficiently fine-tune it on our specific task. The training was performed using a learning rate of 2e-5, which is standard for fine-tuning transformer models. We fine-tuned DistilBERT for 4 epochs to allow it to learn the nuances of the task.

*2) **RoBERTa**:* RoBERTa [30] is another transformer-based model, built upon the BERT architecture, but with improvements like dynamic masking and removal of the Next Sentence Prediction task, which were originally included in BERT's pre-training. Like DistilBERT, the RoBERTa model was loaded from the Hugging Face transformers library, and the output layer was modified for binary classification. RoBERTa was chosen over BERT because of its higher accuracy in many NLP tasks. RoBERTa is a larger model compared to DistilBERT but we trained it for 3 epochs to avoid overfitting.

H. Model Evaluation

To evaluate the performance of the models, we used several standard metrics that are commonly applied in text classification tasks:

Accuracy: The proportion of correctly classified instances out of the total number of instances.

Precision: The proportion of positive predictions that were correct, i.e., how many of the predicted AI texts were actually AI.

Recall: The proportion of actual positive instances that were correctly predicted, i.e., how many of the AI texts were correctly identified.

F1 Score: The harmonic mean of precision and recall, providing a balance between the two metrics, particularly when dealing with imbalanced datasets.

Confusion Matrix: A matrix that shows the number of correct and incorrect classifications, offering a more detailed view of model performance. The confusion matrix helps to visualize misclassifications, where the model incorrectly classifies human-generated text as AI-generated, and vice versa.

Hyperparameter Tuning and Fine-tuning: For the transformer models, fine-tuning was performed to adapt the models to the specific task. The models were trained using learning rates, batch sizes, and epochs that were chosen through experimentation. In the case of RoBERTa, fewer epochs were used due to its larger size and greater computational expense. Hyperparameter tuning was also applied to Logistic Regression and MNB models to optimize their performance. Although these models are less complex compared to transformers, tuning the regularization parameter can significantly improve their ability to generalize to unseen data.

IV. Model Testing and Results

Four different classification models—logistic regression, MNB, DistilBert, and Roberta—have been tested on the unseen test data that constituted 20% of the original dataset to evaluate the performance.

A. *Classical Machine Learning Model*

1) Logistic Regression: A classical string baseline has been served by logistic regression as the model demonstrated solid performance, achieving an accuracy of 0.989. This high accuracy make logistic regression the best performing classical machine learning model. The confusion matrix Fig 6 showed that 61,038 human labeled and 35,956 AI-labeled samples were correctly classified by the logistic regression. However, it had difficulties with borderline cases when the tone, grammar, and vocabulary of AI-generated content were very similar to those of human writing.

2) Multinomial Naive Bayes (MNB): MNB slightly underperform logistic regression because it achieved 0.9738 accuracy, with a precision of 0.98. Also, the confusion matrix Fig 6 demonstrated improved classification across the classes by correctly identifying 60,135 human-written and 34,767

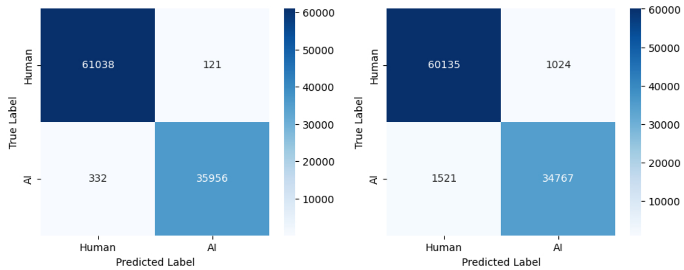

Fig. 6: Confusion Matrix of LR and MNB

AI-generated samples. While MNB may not capture the nuanced contextual relationships present in language as well as transformer-based models, its efficiency and simplicity make it a valuable baseline model in this classification task. These findings encourage the use of MNB in scenarios requiring fast inference and lower computational requirements.

B. Performance of Transformer-Based Models

1) DistilBERT: With an accuracy of 0.9943 and an F1-score of 0.9923 after being fine-tuned across four epochs with a learning rate of 2e-5 and batch size of 32. In it's third epoch DistilBERT provided the best accuracy with a training loss of 0.000100 and validation loss of 0.030400. DistilBERT's architectural strengths are the reason for its exceptional results. It uses bidirectional attention [31] to capture context from both previous and subsequent tokens, and although it is a lighter version of BERT, it nonetheless retains over 97% of BERT's language understanding capabilities. This is essential for distinguishing the more homogeneous structure of AI-generated text from the complex stylistic patterns typical of human writing. Its effectiveness also makes it suitable for real-world deployment without compromising functionality.

2) RoBERTa: RoBERTa, an optimized and robust variant of BERT, was also fine-tuned on the same classification task. With fine-tuning for three epochs due to its larger size and computational demands, RoBERTa delivered 0.9957 accuracy, and an F1 score of 0.9899. To improve its generalization, RoBERTa, an optimized version of BERT, uses techniques including dynamic masking [30] and the elimination of the next sentence prediction target. Despite having fewer training epochs than DistilBERT, it nonetheless performed reasonably well and is especially well-suited for applications requiring accuracy. Here we have used a learning rate of 2e-5 and a batch size of 32. Similar to DistilBERT, RoBERTa also give it's best accuracy in epoch 3 with a training loss of 0.00300 and a validation loss of 0.0209070.

C. Comparative Analysis of Model Performance

The performance of four classification models—Logistic Regression, Multinomial Naive Bayes, DistilBERT, and RoBERTa—was assessed for distinguishing between human-generated and AI-generated text. As detailed in Table 1, transformer-based models (DistilBERT and RoBERTa) achieved superior accuracy compared to classical machine learning models. Specifically, RoBERTa attained the highest accuracy at 99.57%, followed closely by DistilBERT with 99.43% accuracy. These results underscore the efficacy of transformer architectures in capturing the nuanced semantic and contextual cues required for this classification task.

Model	Accuracy	Computational Time
Logistic Regression	0.9890	4.6 s
Multinomial Naive Bayes	0.9738	0.3 s
DistilBERT (4 epochs)	0.9943	18.56 min/epoch
RoBERTa (3 epochs)	0.9957	37.23 min/epoch

TABLE 1: Performance metrics of four models on classification task

Classical models such as Logistic Regression and Multinomial Naive Bayes demonstrated respectable accuracy scores of 98.90% and 97.38%, respectively. While these models provide faster computational performance—training in seconds—they show slightly lower predictive accuracy compared to the transformer models. The Naive Bayes model, in particular, achieved the shortest computational time of just 0.3 seconds, reflecting its simplicity and efficiency.

However, the improved accuracy of transformer models comes at a higher computational cost. Training times for DistilBERT and RoBERTa were approximately 18.56 and 37.23 minutes per epoch, respectively, reflecting the increased complexity and resource requirements of these architectures. This trade-off highlights an important consideration for deployment scenarios where computational efficiency is critical.

V. Limitations and Future Work

This study has some limitations. The dataset, though large, may not fully capture the diversity of real-world human and AI-generated texts, which could impact generalizability. Classical models rely heavily on feature engineering and may miss deeper language nuances, while transformer models require significant computational resources and longer training times. Additionally, the evaluation did not consider adversarial robustness or continuous adaptation to evolving AI text generators.

For future work, several directions are recommended. Expanding the dataset to include diverse languages, genres, and emerging AI-generated content would improve model robustness and applicability. Incorporating techniques such as transfer learning and domain adaptation could further enhance transformer model performance on specialized corpora. Exploration of ensemble learning methods that combine classical and transformer-based models may provide a balance between accuracy and computational efficiency. Additionally, integrating explainability frameworks would offer insights into model decision-making, crucial for ethical and transparent AI deployment. Finally, addressing the challenge of detecting evolving AI-generated text through continual learning and real-time adaptation remains an important avenue for future research.

VI. Conclusion

This research systematically evaluated the performance of four distinct text classification approaches—Logistic Regression, Support Vector Machine, and Transformer-based models such as DistilBERT and RoBERTa—in the task of distinguishing human-written text from AI-generated content. Across a large and balanced dataset, all models demonstrated high classification performance, with the RoBERTa model achieving the best results. The findings suggest that traditional machine learning models, particularly linear regression, remain highly effective in this task when combined with well-engineered feature representations. However, transformer-based models provide a more refined classification capability, particularly in edge cases where stylistic distinctions are subtle. This study contributes to the growing body of work on AI detection and content authenticity by providing empirical benchmarks for model performance. In a digital age where synthetic text generation is increasingly prevalent, accurate classification methods are critical for maintaining trust, combating misinformation, and preserving authorship integrity.

References

[1] M. Jung, "Responsible computational text generation: Ai content classification and policy framework," Ph.D. dissertation, Massachusetts Institute of Technology, 2025.

[2] M. U. Hadi, R. Qureshi, A. Shah, M. Irfan, A. Zafar, M. B. Shaikh, and Mirjalili, "Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects," *Authorea Preprints*, vol. 1, pp. 1–26, 2023.

[3] L. Floridi and M. Chiriatti, "Gpt-3: Its nature, scope, limits, and consequences," *Minds and Machines*, vol. 30, pp. 681–694, 2020.

[4] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat *et al.*, "Gpt-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023.

[5] G. Yenduri, M. Ramalingam, G. C. Selvi, Y. Supriya, G. Srivastava, P. K. R. Maddikunta, G. D. Raj, R. H. Jhaveri, B. Prabadevi, W. Wang *et al.*, "Gpt (generative pre-trained transformer)–a comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions," *IEEE Access*, 2024.

[6] S. Mahony and Q. Chen, "Concerns about the role of artificial intelligence in journalism, and media manipulation," *Journalism*, p. 14648849241263293, 2024.

[7] F. Pedro, M. Subosa, A. Rivas, and P. Valverde, "Artificial intelligence in education: Challenges and opportunities for sustainable development," 2019.

[8] G. U. Yule, "On sentence-length as a statistical characteristic of style in prose: With application to two cases of disputed authorship," *Biometrika*, vol. 30, no. 3/4, pp. 363–390, 1939.

[9] M. Hanna, O. Liu, and A. Variengien, "How does gpt-2 compute greater-than?: Interpreting mathematical abilities in a pre-trained language model," *Advances in Neural Information Processing Systems*, vol. 36, pp. 76 033–76 060, 2023.

[10] M. Johnsen, *Developing AI Applications With Large Language Models*. Maria Johnsen, 2025.

[11] S. Siddiqui, M. A. Rehman, S. M. Daudpota, and A. Waqas, "Opinion mining: An approach to feature engineering," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 3, pp. 159–165, 2019.

[12] T. N. Eilifsen, "Mimicking the student: A feature-based approach for detecting ai-generated texts and verifying authorship in educational settings," Master's thesis, NTNU, 2024.

[13] D. Yan, M. Fauss, J. Hao, and W. Cui, "Detection of ai-generated essays in writing assessments," *Psychological Test and Assessment Modeling*, vol. 65, no. 1, pp. 125–144, 2023.

[14] M. Zeng, "Research on ai-generated text detection based on machine learning models," *Transactions on Computer Science and Intelligent Systems Research*, vol. 7, pp. 229–233, 11 2024.

[15] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable ai: A review of machine learning interpretability methods," *Entropy*, vol. 23, no. 1, p. 18, 2020.

[16] S. Fariello, G. Fenza, F. Forte, M. Gallo, and M. Marotta, "Distinguishing human from machine: a review of advances and challenges in ai-generated text detection," *International Journal of Interactive Multimedia and Artificial Intelligence*, 2024.

[17] Y. Ge, W. Hua, K. Mei, J. Tan, S. Xu, Z. Li, Y. Zhang *et al.*, "Openagi: When llm meets domain experts," *Advances in Neural Information Processing Systems*, vol. 36, pp. 5539–5568, 2023.

[18] M. Prytula, "Fine-tuning bert, distilbert, xlm-roberta and ukr-roberta models for sentiment analysis of ukrainian language reviews," *Machine learning*, vol. 3, no. 4, 2024.

[19] A. Areshey and H. Mathkour, "Exploring transformer models for sentiment classification: A comparison of bert, roberta, albert, distilbert, and xlnet," *Expert Systems*, vol. 41, no. 11, p. e13701, 2024.

[20] S. González-Carvajal and E. C. Garrido-Merchán, "Comparing bert against traditional machine learning text classification," *arXiv preprint arXiv:2005.13012*, 2020.

[21] A. Akram, "An empirical study of ai generated text detection tools," *arXiv preprint arXiv:2310.01423*, 2023.

[22] C. Chaka, "Reviewing the performance of ai detection tools in differentiating between ai-generated and human-written texts: A literature and integrative hybrid review," *Journal of Applied Learning and Teaching*, vol. 7, no. 1, pp. 115–126, 2024.

[23] S. Bahad, Y. Bhaskar, and P. Krishnamurthy, "Fine-tuning language models for ai vs human generated text detection," in *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, 2024, pp. 918–921.

[24] A. M. Elkhatat, K. Elsaid, and S. Almeer, "Evaluating the efficacy of ai content detection tools in differentiating between human and ai-generated text," *International Journal for Educational Integrity*, vol. 19, no. 1, p. 17, 2023.

[25] K. Banachewicz and L. Massaron, *The Kaggle Workbook*. Packt Publishing, 2023.

[26] H. Christian, M. P. Agus, and D. Suhartono, "Single document automatic text summarization using term frequency-inverse document frequency (tf-idf)," *ComTech: Computer, Mathematics and Engineering Applications*, vol. 7, no. 4, pp. 285–294, 2016.

[27] M. Tranmer and M. Elliot, "Binary logistic regression," *Cathie Marsh for census and survey research, paper*, vol. 20, pp. 90 033–9, 2008.

[28] X. Zou, Y. Hu, Z. Tian, and K. Shen, "Logistic regression model optimization and case analysis," in *2019 IEEE 7th international conference on computer science and network technology (ICCSNT)*. IEEE, 2019, pp. 135–139.

[29] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter," *arXiv preprint arXiv:1910.01108*, 2019.

[30] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[31] A. R. Murthy, A. K. KM, and N. Vedavathi, "Hierarchical attention mechanism with distil bidirectional encoding representations from transformers based depression detection on social media posts." *International Journal of Intelligent Engineering & Systems*, vol. 17, no. 6, 2024.