

AI-Generated Content Detection Model Using a Zero-Shot Learning Algorithm

Olukayode Aiyeniko¹, Hadiatillah Oyediran Sanusi², Abdulrahman Gado Sani³, Abubakar Mu'azu Abujiya⁴, Aminat Omotayo Adebayo⁵, Babatunde Cyrus Okunlola⁶

¹Department of Computer Science, Lagos State University, Lagos, Nigeria

²⁻⁶Department of Computer Science, Alhikmah University, Ilorin, Kwara State, Nigeria

*Corresponding Author: olukayode.aiyeniko@lasu.edu.ng

ABSTRACT: The rapid evolution of generative Artificial Intelligence (AI) models has revolutionized digital content creation, yet it has also raised significant ethical and practical concerns regarding authenticity, misinformation, and academic integrity. Existing detection methods, which largely depend on supervised machine learning with labeled datasets, struggle to keep pace with the adaptability of emerging AI models. This study proposes a novel AI-generated content detection system leveraging Zero-Shot Learning (ZSL), which eliminates the reliance on task-specific labeled data while improving adaptability and scalability. The system integrates text preprocessing, linguistic feature extraction, and a ZSL-based classification framework that categorizes content as human-written or AI-generated, supplemented with a confidence scoring mechanism for transparency. Experimental evaluations against benchmark models, including RoBERTa, DistilBERT, and the GPT-2 Output Detector, demonstrate that the ZSL model achieves highly competitive results, with accuracy rates exceeding 94% for human-written text and 99% for AI-generated text. The findings highlight ZSL's potential as a cost-effective, flexible, and future-ready approach for AI content detection across domains such as education, journalism, and social media. This work underscores the need for scalable detection frameworks that can adapt to the rapidly advancing landscape of generative AI.

KEYWORDS: AI-generated Content, Detection Mode, Machine Learning, Zero-shot Algorithm

I. INTRODUCTION

Recent advancements in Artificial Intelligence (AI), particularly in natural language processing, have transformed digital content creation, consumption, and distribution (Nasser & Erafy, 2023; Prasetya & Kurniawan, 2024). Prominent AI models, such as ChatGPT and Gemini, generate impressive results based on user prompts, producing concise and contextually relevant text (Medeiros et al., 2020; McIntosh et al., 2025). Various industries, including content marketing, customer support, education, and creative writing, have begun to adopt these automatically generated texts, underscoring the importance of AI tools in the workplace (Obinna Eziefule et al., 2022; Knoth et al., 2024). However, these developments raise ethical concerns regarding their application. A primary issue is the difficulty in distinguishing AI-generated content from human-written text (Ghiur & Popescu, 2025). The use of AI tools to create essays and assignments raises concerns regarding academic integrity. Furthermore, the proliferation of AI-generated fake news undermines trust in journalism, prompting questions about the role of journalists in an era of misinformation (Vieriu, 2025). These challenges highlight the need for effective tools and frameworks to detect AI-generated content. A key factor that plays an essential role in AI-generated content detection involves labeled data, which are datasets where each data point is paired with its corresponding label (Zhaxylykbayeva et al., 2025). These labels act as some form of objective truth for supervised machine models to learn from. For instance, in the case of text datasets, a paragraph may be labeled as either "human-generated" or "AI-generated" (Schaaff et al., 2024).

Current detection methods, such as Grover and ForensicTransfer, rely on machine learning models trained on labeled datasets, but these approaches have limitations (Elkhatat et al., 2023; Fariello et al., 2024). They often require extensive labeled data, which is costly to produce, and struggle to adapt to new AI models. As these models evolve rapidly, traditional detection tools may become ineffective (Cadet et al., 2024). Zero-shot learning (ZSL) presents a promising alternative. Unlike supervised learning, ZSL enables machine learning models to recognize and classify new concepts without labeled data, enhancing detection

capabilities for unseen AI-generated text (Kumarswamy, 2025). The capacity of artificial intelligence models to produce text that closely resembles human writing has resulted in a range of challenges and opportunities (She et al., 2020). These advancements have expedited and enhanced the efficiency of content creation; however, they have also complicated the differentiation between text authored by humans and that generated by AI. This dilemma is evident across various sectors, including media and education. In academic institutions, for example, students are now able to use AI to effortlessly generate complete essays or fulfill assignments, which raises significant concerns regarding academic integrity.

In the realm of journalism, AI often produces misleading information while reporting on events, as it strives to maintain an unbiased or neutral perspective, regardless of the subject matter. Consequently, this practice contributes to the proliferation of misinformation, making it increasingly challenging for consumers to trust the information they encounter (Aissani et al., 2023; Adjin-tetty et al., 2024). Current tools for detecting AI-generated content struggle to keep pace with these developments. A significant number of these tools depend heavily on supervised learning, which necessitates a substantial amount of labeled data for effective performance. This requirement not only demands considerable time and resources for data production but also results in inadequacies when faced with the emergence of new and more advanced AI systems. Thus, conventional methods lack the flexibility necessary to adapt to the rapid advancements in AI technology (Aram & Juliana, 2024). An alternative approach, zero-shot learning, may be introduced as it does not rely on labeled training data and has the potential to make detection tools significantly more adaptable. Despite its potential, zero-shot learning has not been sufficiently explored in the context of detecting AI-generated text. This study applied a shot model for the development of an AI-generated content detection system.

II. RELATED WORK

Ghiur and Popescu (2025) analyzed the characteristics, methodologies, and challenges associated with detecting the synthetic content across multiple modalities to safeguard digital authenticity and integrity. Major detection approaches reviewed include stylometric analysis, watermarking, pixel prediction techniques, dual-stream networks, machine learning models, blockchain, and hybrid approaches, highlighting their strengths and limitations, as well as their detection accuracy, with an independent accuracy of 80% for stylometric analysis and up to 92% using multiple modalities in hybrid approaches. The effectiveness of these techniques was explored in diverse contexts, from identifying deepfakes and synthetic media to detecting AI-generated scientific texts. Ethical concerns, such as privacy violations, algorithmic bias, false positives, and overreliance on automated systems, were also critically discussed. Furthermore, the study addressed legal and regulatory frameworks, including intellectual property challenges and emerging legislation, emphasizing the need for robust governance to mitigate misuse. Real-world examples of detection systems are analyzed to provide practical insights into implementation challenges. A review of AI detection tools (Khan Ejaj, Khan Imran, Khan Talha, 2025), the study mentioned literature revealing better accuracy with GPT-3.5 than GPT-4 4 but noting false positives with human-written text. This highlighted the need to refine these tools as AI content advances. The study aims to build a machine learning model to improve content authenticity for educators, journalists, and moderators. Using Python, Jupyter Notebook, VS Code, Transformers, and Torch, it also leverages RoBERTa for enhanced accuracy on a balanced dataset. (Akash Jha, Abhishek Rawat, 2025) designed a machine learning-based model for automatic detection of AI-generated text. The Kaggle dataset comprises different textual features, including syntactic patterns, semantic coherence, and language model perplexity scores, which were used for evaluation. By leveraging advanced feature engineering techniques and deep learning architectures, the developed model showed high accuracy in differentiating between AI-generated and human-generated text. Demanding testing and evaluation highlighted the effectiveness of the system in identifying AI-generated content, thus mitigating the spread of misinformation and ensuring the integrity of textual data.

Development of a model to ensure the authenticity of information, guarantee the reliability and credibility of information sources, and prevent the spread of false information, fabricated data, and misleading content (Zeng, 2024). In the academic field, detecting AI-generated papers, articles, and assignments helps maintain academic integrity, prevent academic fraud and plagiarism, and thus improve academic capabilities. This study summarizes the characteristics of the three selected models, which are Logistic Regression, Support Vector Machine (SVM), and Naive Bayes (NB) Classifier. And provide recommendations and directions for improvement in the choice of detection models for AI-generated content. Through comparison of three models, logistic regression, SVM, and Naive Bayes on the same dataset in terms of Accuracy, Precision, and F1-score, it is determined that logistic regression performs the best for this type of dataset. Logistic regression achieves superior performance with metrics exceeding 90%. SVM shows suitability for large

datasets, with metrics around 70% in this dataset. However, Naive Bayes, typically suitable for smaller datasets, performs poorly on this dataset, achieving only 50% accuracy. Weber-Wulff et al.(2023) examined the general functionality of detection tools for AI-generated text and evaluated them based on accuracy and error type analysis. The study produced answers to research questions about whether existing detection tools can reliably differentiate between human-written text and ChatGPT-generated text, and whether machine translation and content obfuscation techniques affect the detection of AI-generated text. The work covered 12 publicly available tools and two commercial systems (Turnitin and PlagiarismCheck) that were widely used in the academic setting. Conclusion was made that the available detection tools were neither accurate nor reliable, and have a main bias towards classifying the output as human-written rather than detecting AI-generated text. Furthermore, content obfuscation techniques significantly worsen the performance of tools. Investigation of the capabilities of different AI content detection tools in discerning human and AI-authored content (Elkhatat et al., 2023). Fifteen paragraphs each from ChatGPT Models 3.5 and 4 on the topic of cooling towers in the engineering process, and five human-written control responses were produced for evaluation. AI content detection tools developed by OpenAI, Writer, Copyleaks, GPTZero, and CrossPlag were employed to evaluate these paragraphs. The results showed that the AI detection tools were more accurate in the detection of content generated by GPT 3.5 than GPT 4. However, when applied to human-written control responses, the tools exhibited inconsistencies, producing false positives and uncertain classifications. This study suggested the need for further development and refinement of AI content detection tools as AI-generated content becomes more sophisticated and harder to distinguish from human-written text. A multi-domain dataset for testing the state-of-the-art APIs and tools for detecting artificially generated information used by universities and other research institutions (Akram, 2023). A large dataset consisting of articles, abstracts, stories, news, and product reviews was created for this study. The second step is to use the newly created dataset to put six tools through their paces. Six different artificial intelligence (AI) text identification systems, including "GPTkit," "GPTZero," "Originality," "Sapling," "Writer," and "Zylalab," have accuracy rates between 55.29 and 97.0%. Although all the tools fared well in the evaluations, originality was particularly effective across the board.

III. METHODOLOGY

The developed system presents an innovative method for detecting AI-generated content by utilizing zero-shot learning (ZSL). In contrast to traditional supervised models that rely heavily on labeled datasets, this system employs pre-trained models that can generalize across various tasks without the need for specific training data. The study adopted an experimental research design that involved developing and evaluating an AI-generated content detection model. The core idea is to leverage Zero-Shot Learning (ZSL) to classify content as either AI-generated or human-written without requiring large-scale annotated training datasets.

(a). Dataset Collection

The effectiveness of an AI-generated content detection model relies heavily on the quality, diversity, and balance of the dataset used. Since zero-shot learning does not require task-specific labeled training data, careful curation of test and evaluation datasets becomes critical. The data used in this study were gathered through two sources:

- (a). **Human-Written Texts:** Collected from open-source repositories such as Wikipedia, news articles, academic papers, ArXiv, and (scientific/technical writing PubMed abstracts and blogs.
- (b). **AI-Generated Texts:** Generated using large language models (LLMs) such as GPT, LLaMA, Claude, or similar. Multiple prompts are designed to ensure diversity across domains (news, creative writing, technical content).

(b). Preprocessing of the Dataset

The preprocessing stage ensures that both human-written and AI-generated texts are in a uniform, clean, and comparable format before feeding them into the zero-shot learning algorithm. The input text undergoes a cleaning and standardization process to ensure uniform formatting. This process includes the removal of extraneous characters, case normalization, punctuation handling, text length normalization, tokenization of the text, and normalization for linguistic analysis.

(c). Feature Extraction

The goal of this phase is to transform raw text data into meaningful numerical representations that capture the semantic, syntactic, and stylistic properties of the text. These representations are then used for similarity comparison with predefined class labels (human-written vs AI-generated) in the Zero-Shot Learning

framework. The pre-trained ZSL model examines the input text to extract linguistic features such as sentence structure, coherence, vocabulary distribution, and contextual relevance. These features form the basis for comparison against patterns typically associated with AI-generated content.

(d). Zero-Shot Learning Model

A ZSL framework is used for classification purposes. The model employs a natural language understanding (NLU) approach to categorize the input text into predefined classes: "Human-Generated" or "AI-Generated." The ZSL model's capacity to infer patterns without task-specific training enhances its adaptability to new AI models and a variety of text domains.

(e). Confidence Scoring Mechanism

The system generates a confidence score that indicates the probability of the text being AI-generated. This score offers users a transparent metric for evaluating the reliability of the classification.

(f). Features of the Developed System Adaptability

The incorporation of zero-shot learning allows the system to adjust to newly released AI models without necessitating retraining or updates to labeled datasets.

- (a). Scalability: The lightweight architecture of the ZSL model enables the system to efficiently process large volumes of text, rendering it suitable for a variety of applications, including education, journalism, and social media moderation.
- (b). Language-Agnostic Design: The system is designed to be extendable to support multiple languages, thereby enhancing its applicability.
- (c). Ethical Integration: By providing confidence scores, the system promotes transparency, enabling users to make informed decisions based on the detection outcomes. The developed system endeavors to address the limitations of current methodologies by introducing a flexible and future-oriented solution. This design guarantees that the system remains relevant and effective in tackling the challenges presented by the rapid advancement of AI-generated content. The use-case diagram is shown in Figure 1.

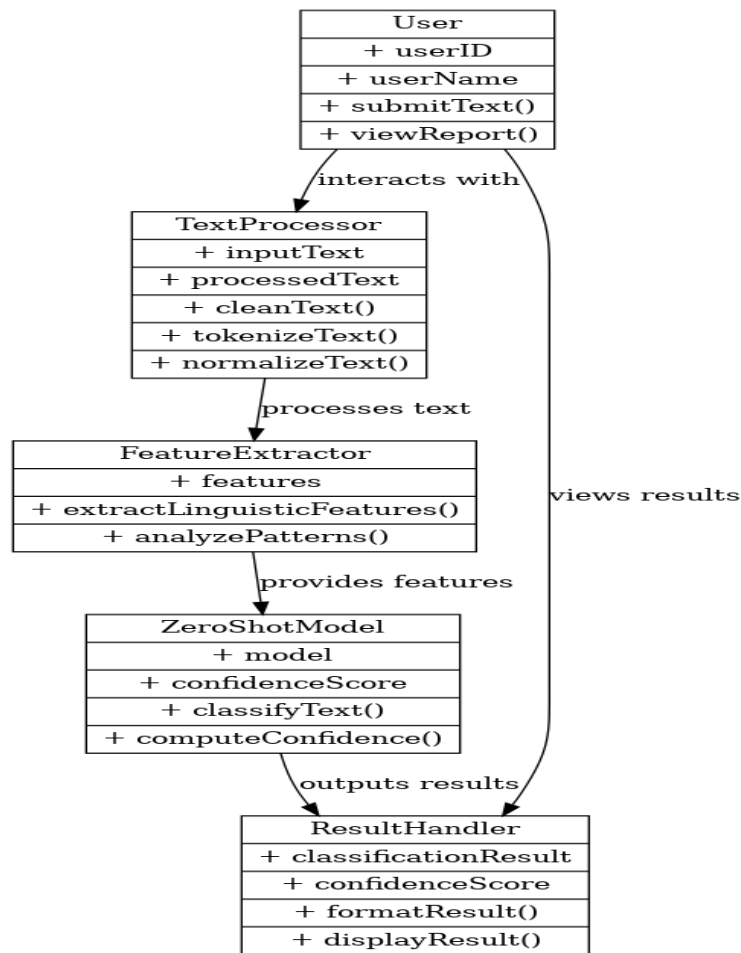


Figure 1: Class Diagram

These diagrams provide a clear conceptual blueprint of the developed system, laying the foundation for its detailed implementation.

(g). Algorithm of the AI-Generated Content Detection

Step 1: Start: Initialize the system.

Step 2: Text Cleaning:

(a). Remove noise (special characters, excess whitespace).

(b). Normalize text (lowercase conversion, consistent formatting).

Step 3: Feature Extraction: Tokenize and analyze linguistic structures (e.g., grammar, syntax, vocabulary diversity).

Step 4: Run Zero-Shot Model:

(a). Input the extracted features into the zero-shot learning model.

(b). Predict the category of the text (AI-Generated or Human-Written).

(c). Calculate the confidence score for the prediction.

Step 5: Result Compilation:

(a). Format the classification result with the confidence score.

(b). Store the result for reporting.

Step 6: Output Results: Display the final classification and confidence score to the user.

Step 7: End.

IV. RESULTS AND DISCUSSION

The results of the developed AI-generated content detection model are presented in a stepwise approach. These are represented in the following sub-sections

(a). Main Page

The main page of the AI Content Detection system serves as the initial interface where users interact with the tool to submit text for classification. Designed with simplicity and clarity in mind, this page seamlessly blends functionality with a user-friendly layout, ensuring an intuitive experience for both technical and non-technical users. At the top, a concise header introduces the purpose of the system: enabling users to evaluate whether a given text is AI-generated or human-written using multiple advanced models. A short flavor text follows, briefly outlining the supported models and their capabilities, such as Zero-Shot Learning (ZSL), RoBERTa, DistilBERT, and GPT-2 Output Detector. This contextual information prepares users for the nature of the analysis they are about to receive. The central feature of the page is the text input section, which leverages a rich text editor built using the Tip-Tap framework. This editor offers a smooth typing experience and encourages users to input any block of text they wish to analyze. Once submitted, the system runs the text through all available models simultaneously and returns the results in real-time, as indicated in Figure 2.

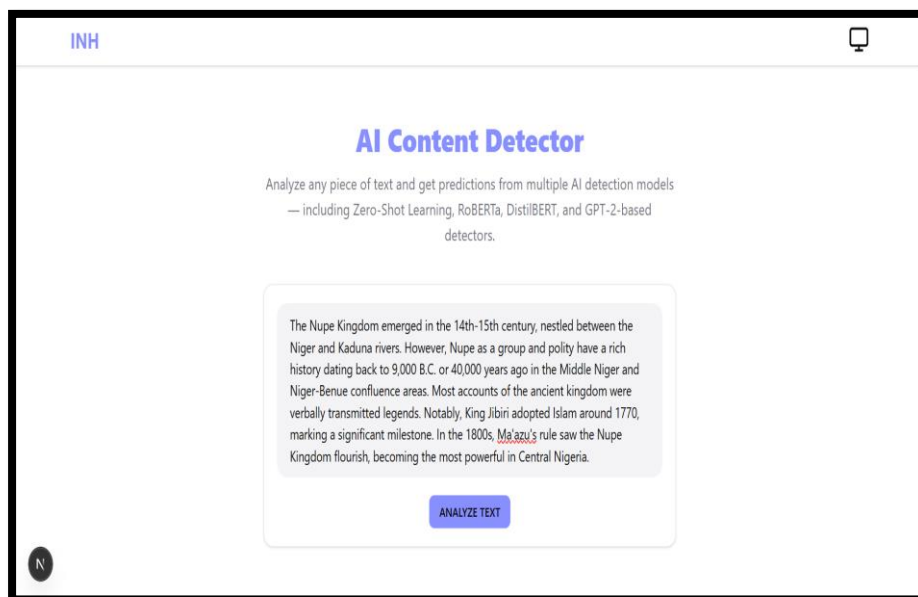


Figure 2: AI Content Detector

Figure 2 shows the interface of the AI Content detector that indicates or flags up that particular content is generated through AI.

(b). Result of Text Editor

At the core of user interaction within the AI Content Detection platform is the custom-built Text Editor, which plays a pivotal role in the input design and user experience. The editor is implemented using Tip-Tap, a powerful headless framework built on top of Prose Mirror. This choice was deliberate, allowing for extensive flexibility and control over the editing experience while maintaining a lightweight and modern interface. Functionally, the editor captures and processes plain text input, which is then passed to the classification engine upon submission. It restricts excessive formatting to maintain data uniformity and ensure that the underlying models receive clean, unstructured text. To prevent accidental empty submissions or misfires, the “ANALYZE TEXT” button is intelligently disabled unless the editor contains valid input. After successful classification, the editor clears its content, allowing users to seamlessly input new samples without page refreshes or additional clicks, as indicated in Figure 3.

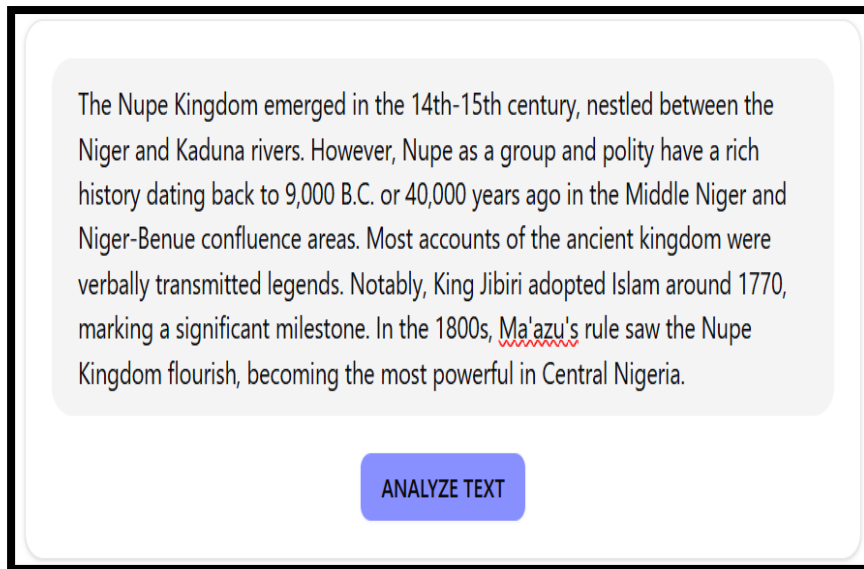


Figure 3. Text Editor

(c). Result of Model Prediction

A key highlight of the AI Content Detection platform is the comprehensive presentation of Model Predictions, where outputs from four distinct AI detection models—Zero-Shot Learning (ZSL), RoBERTa, DistilBERT, and GPT-2 Output Detector—are displayed in a clean and interactive format. Each model operates independently and provides a confidence-based assessment of whether the input text is AI-generated or human-written. These outputs empower users to understand not only the final verdict of each model but also the nuanced probabilities behind their decisions. Upon submission of text via the editor, the backend processes the input through all four models simultaneously. Each model returns its prediction as a set of labels (e.g., "AI-Generated", "Human-Generated") along with associated confidence scores. These predictions are mapped and standardized on the front end to ensure consistent user interpretation, especially since model labels vary (e.g., "Real", "Fake", or "LABEL_0"). The modular architecture ensures that each model's result is encapsulated and can be expanded upon or replaced in future upgrades. This layered visualization of model performance allows users to cross-reference outputs, notice inconsistencies, and better understand the reliability of each model. It also sets the stage for more advanced future functionality, such as highlighting specific parts of the input text flagged as AI-generated. The resulting panel gives comprehensive information as depicted in Figure 4. Table 1 displays the entire result of the accuracy obtained from experimental analysis.

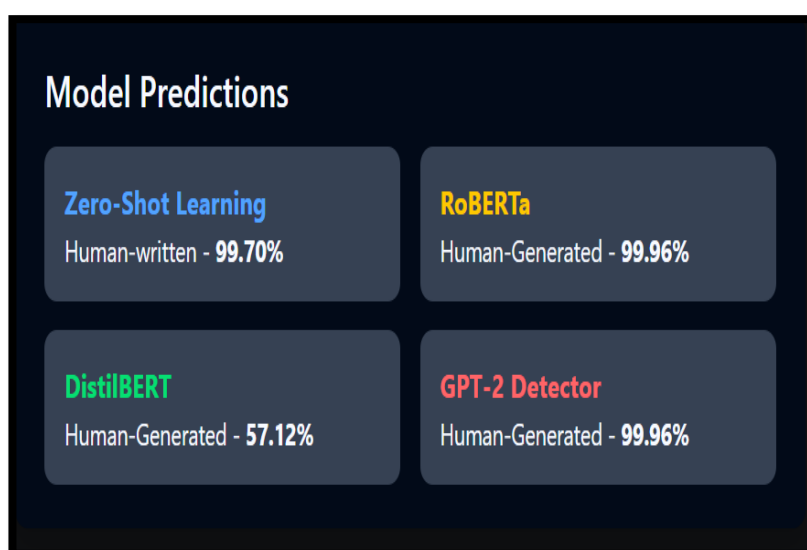


Figure 5: Result-Panel

Figure 5 mentions different results obtained from some selected AI-content generator software. The comparative comparison indicates that the highest results in Zero-shot learning (Human-written) recorded 99.70%, 0.9970, 0.9605, RoBERTa (Human-Generated) gave 99.97%, 0.9990, 0.9900, DistilBERT (Human-Generated) gave 57.12%, 0.5712, 0.5700 and GPT-2 Detector (Human-Generated) recorded 99.97%, 0.9970, 0.9960.

Table 1: Result of Recognition Rate (Accuracy, Precision, and Recall)

Test Text	Ground Truth	ZSL Prediction	RoBERTa Prediction	DistilBERT Prediction	GPT-2 Prediction
Wikipedia expert articles	Human-written	Human-Generated 99.70%, 0.9970, 0.9605	Human-Generated 99.97%, 0.9990, 0.9900	AI-Generated 57.12%, 0.5712, 0.5700	Human-Generated 99.97%, 0.9970, 0.9960
GPT-generated response	AI-generated	AI-Generated 98.22%, 0.9820, 0.9700	AI-Generated 97.50%, 0.9700, 0.9705	AI-Generated 84.75%, 0.8470, 0.8450	AI-Generated 97.90%, 0.9700, 0.9780
Human opinion blog post	Human-written	Human-Generated 94.63%, 0.9463, 0.9470	Human-Generated 92.88%, 0.9290, 0.9300	Human-Generated 86.10%, 0.8600, 0.8620	Human-Generated 94.02%, 0.9400, 0.9301
AI-generated academic report	AI-generated	AI-Generated 99.95%, 0.9905, 0.9914	Human-Generated 99.81%, 0.9990, 0.9900	Uncertain 46.51%, 0.4050, 0.4001	Human-Generated 99.81%, 0.9990, 0.9915
News Article	Human-written	Human-Generated 96.20%, 0.9700, 0.9630	Human-Generated 99.93%, 0.9994, 0.9900	Human-Generated 76.10%, 0.7610, 0.7700	Human-Generated 97.45%, 0.9700, 0.9750

The results obtained from our multi-model evaluation strongly support the core objectives of our project to demonstrate that Zero-Shot Learning (ZSL) models can serve as highly effective and flexible alternatives to traditional AI content detection models. Our ZSL model, powered by the cross-encoder/nli-deberta-v3-large, consistently produced predictions that aligned closely with ground truth classifications across various types of input, from AI-generated text to human-written historical and academic content. Notably, ZSL achieved a 99.95% confidence in identifying AI-generated content and maintained over 94% accuracy in detecting human-written content. While traditional models like RoBERTa and the GPT-2 Detector also showed high performance, they required task-specific fine-tuning and were more rigid in their output interpretation. DistilBERT's occasional uncertainty further highlighted the limitations of relying solely on pre-labeled training datasets. In contrast, ZSL demonstrated strong adaptability without the need for retraining, making it a cheaper, scalable, and generalizable solution. These findings validate our hypothesis that Zero-Shot Learning offers a powerful method for AI content detection, capable of performing at par with and in some cases outperforming traditional detectors. This positions ZSL as a viable foundation for future research and real-world applications in AI-generated content monitoring.

V. CONCLUSION

This study developed and evaluated an AI-generated content detection system that leverages the power of Zero-Shot Learning (ZSL) to address the limitations of conventional supervised models. Unlike existing detectors that rely heavily on extensive labeled datasets, the proposed approach demonstrates adaptability, scalability, and cost-effectiveness by classifying unseen text without task-specific training. Comparative analysis with established models such as RoBERTa, DistilBERT, and the GPT-2 Output Detector revealed that the ZSL model not only performs competitively but also offers superior flexibility in adapting to newly emerging AI models. With accuracy levels exceeding 94% for human-written text and 99% for AI-generated text, the system validates the potential of ZSL as a robust alternative for future-oriented content authenticity verification. Beyond technical contributions, this work highlights the broader implications of

reliable AI-content detection in safeguarding academic integrity, mitigating misinformation, and enhancing trust in digital communication. Future research can expand this framework by integrating multilingual support, domain-specific adaptations, and hybrid architectures to further strengthen its applicability across diverse contexts.

REFERENCES

- Adjin-tettey, T. D., Muringa, T., & Danso, S. (2024). The Role of Artificial Intelligence in Contemporary Journalism Practice in Two African Countries. *Journalism and Media*, 5, 846–860.
- Aissani, R., Abdallah, R. A., Taha, S., Noor, M., & Adwan, A. (2023). Artificial Intelligence Tools in Media and Journalism: Roles and Concerns. *2023 International Conference on Multimedia Computing, Networking and Applications (MCNA)*, 19–26. <https://doi.org/10.1109/MCNA59361.2023.10185738>
- Akash Jha, Abhishek Rawat, M. S. G. (2025). Automatic Detection of an AI-Generated Text Model. *International Journal of Research Publication and Reviews*, 5(5), 8760–8768. <https://doi.org/10.2139/ssrn.5134717>
- Akram, A. (2023). An Empirical Study of AI-Generated Text Detection Tools An Empirical Study of AI-Generated Text Detection Tools. *Advances in Machine Learning & Artificial Intelligence*, 4(2), 44–55.
- Aram, I. A., & Juliana, E. A. (2024). Digitisation and artificial intelligence in the world of media. *Journal of Media And*, 16(2), 31–37. <https://doi.org/10.5897/JMCS2024.0823>
- Cadet, E., Osundare, O. S., Ekpobimi, H. O., Samira, Z., & Wondaferew, Y. (2024). AI-powered threat detection in surveillance systems: A real-time data processing framework AI-powered threat detection in surveillance systems: A real-time data processing framework. *Journal of Engineering and Technology*, 7(2), 1–16. <https://doi.org/10.53022/oarjet.2024.7.2.0057>
- Elkhatat, A. M., Elsaid, K., & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and generated text. *International Journal for Educational Integrity*, 5, 1–16. <https://doi.org/10.1007/s40979-023-00140-5>
- Fariello, S., Fenza, G., Forte, F., Gallo, M., & Marotta, M. (2024). Distinguishing Human From Machine: A Review of Advances and Challenges in AI-Generated Text Detection. *International Journal of Interactive Multimedia and Artificial Intelligence*, 1–14. <https://doi.org/10.9781/ijimai.2024.12.002>
- Ghiur, D., & Popescu, D. E. (2025). Distinguishing Reality from AI: Approaches for Detecting Synthetic Content. *Computers*, 14(1), 1–33.
- Hua, H., & Yao, C. (2024). Investigating generative AI models and detection techniques: impacts of tokenization and dataset size on identification of AI-generated text. *Frontiers*, 1–10.
- Khan Ejaj, Khan Imran, Khan Talha, Y. J. (2025). AI Content Detector. *International Journal of Advanced Research in Science, Communication and Technology International*, 5(10), 505–510. <https://doi.org/10.48175/568>
- Knoth, N., Tolzin, A., Janson, A., & Marco, J. (2024). Computers and Education: Artificial Intelligence, AI literacy, and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence*, 6, 100225. <https://doi.org/10.1016/j.caeai.2024.100225>
- Kumarswamy, S. (2025). Zero-Shot Learning in AI: Enabling Machines to Understand the Unseen Zero-Shot Learning in AI: Enabling Machines to Understand the Unseen. *International Journal of Scientific Research & Engineering Trends*, 10(2), 1–6.
- Mcintosh, T. R., Susnjak, T., Liu, T., Watters, P., Xu, D., Liu, D., & Halgamuge, M. N. (2025). From Google Gemini to OpenAI Q * (Q-Star): A Survey on Reshaping the Generative Artificial Intelligence (AI) Research Landscape. *Technologies*, 13(51), 1–56.
- Medeiros, N., Ivaki, N., Costa, P., & Vieira, M. (2020). Software Metrics and Machine Learning. *IEEE Access*, 4, 1–25. <https://doi.org/10.1109/ACCESS.2020.3041181>
- Nasser, A., & Erafy, E. (2023). Applications of Artificial Intelligence in the field of media Keywords. *International Journal of Artificial Intelligence and Emerging Technology*, 2(2), 19–41.
- Obinna Eziefule, A., Oyinkansola Adelakun, B., Okoye, I. N., & Attieku, J. S. (2022). The Role of AI in Automating Routine Accounting Tasks: Efficiency Gains and Workforce Implications. *European Journal of Accounting, Auditing and Finance Research*, 10(12), 109–134. <https://doi.org/10.37745/ejafr.2013/vol10n12109134>
- Prasetya, A., & Kurniawan, F. (2024). Telematics and Informatics Reports Advancements in natural language processing: Implications, challenges, and future directions. *Telematics and Informatics Reports*, 16, 100173. <https://doi.org/10.1016/j.teler.2024.100173>

- Schaaff, K., Schlippe, T., & Mindner, L. (2024). Classification of human - and generated texts for different languages and domains. *International Journal of Speech Technology*, 27(4), 935–956. <https://doi.org/10.1007/s10772-024-10143-3>
- She, D., Chen, Y., Shah, A., Ray, B., & Jana, S. (2020). Neutaint: Efficient dynamic taint analysis with neural networks. *Proceedings - IEEE Symposium on Security and Privacy*, 1527–1543. <https://doi.org/10.1109/SP40000.2020.00022>
- Vieriu, A. M. (2025). The Impact of Artificial Intelligence (AI) on Students ' Academic Development. *Education Sciences*, 15(343), 1–12.
- Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, 19(1), 1–39. <https://doi.org/10.1007/s40979-023-00146-z>
- Zeng, M. (2024). Research on AI-Generated Text Detection Based on Machine Learning Models. *Transactions on Computer Science and Intelligent Systems Research*, 7, 229–233. <https://doi.org/10.62051/8k1jga32>
- Zhaxylykbayeva, R., Burkitbayeva, A., Zhakhyp, B., Kabylgazina, K., & Ashirbekova, G. (2025). *Artificial Intelligence and Journalistic Ethics : A Comparative Analysis of AI-Generated Content and Traditional Journalism*. 1–22.