DETECT AI GENERATED TEXTS

Nayakapu Ravitea
Scholar. Department of MCA
Vaageswari College of Engineering, Karimnagar

Tallapally Mounika
Assistant Professor
Vaageswari College of Engineering, Karimnagar

Dr. P. Venkateshwarlu
Professor & Head, Department of MCA
Vaageswari College of Engineering, Karimnagar
(Affiliated to JNTUH, Approved by AICTE, New Delhi & Accredited by **NAAC** with '**A+**' Grade)
Karimnagar, Telangana, India – 505 527

ABSTRACT

The rapid development of **artificial intelligence (AI)** and **natural language generation (NLG)** models, such as GPT and other large language models, has led to a surge in AI-generated textual content across social media, academic writing, and online publications. While these technologies enable automated content creation, they also pose challenges related to **authenticity, plagiarism, and misinformation**. This project focuses on **detecting AI-generated texts** by analyzing linguistic patterns, syntactic structures, and semantic features that differentiate machine-written text from human-authored content. Methods such as **stylometric analysis, statistical modeling, machine learning classifiers, and deep learning techniques** are employed to identify characteristics of AI-generated text, including repetitive phrasing, unnatural coherence, and lack of personal context. The proposed system provides an **automated and scalable solution** to identify AI-written content, supporting educators, publishers, and online platforms in ensuring content integrity, originality, and reliability.

Keywords: AI-Generated Text Detection, Natural Language Processing (NLP), Stylometric Analysis, Machine Learning, Deep Learning, Text Authenticity, Content Integrity.

1.INTRODUCTION

With the rise of advanced **AI language models** such as GPT, BERT, and other natural language generation systems, the creation of machine-generated text has become widespread. These AI systems can produce coherent, human-like content in a variety of contexts, including articles, essays, social media posts, and academic writing. While this technology offers efficiency and creativity, it also raises concerns about **plagiarism, misinformation, and content authenticity**.

Detecting AI-generated text is crucial to maintaining the **integrity of information** and ensuring that content is genuinely authored by humans. Traditional plagiarism detection tools are insufficient because AI-generated content is original in structure but may still lack human nuances. Researchers have found that AI-generated texts often exhibit **statistical patterns, repetitive phrasing, and certain syntactic or semantic consistencies** that differ from human writing.

This project focuses on developing a system capable of identifying AI-generated texts by leveraging **Natural Language Processing (NLP), stylometric analysis, and machine learning techniques**. By analyzing linguistic features, text patterns, and writing styles, the system can distinguish machine-authored content from human-authored material. The goal is to provide an **automated, scalable, and accurate solution** that can assist educators, publishers, and online platforms in ensuring content authenticity and preventing the spread of misleading or artificially generated information.

2.LITERATURE REVIEW

The detection of AI-generated texts has become an important research area due to the proliferation of large language models (LLMs) capable of producing human-like content. Early approaches focused on **stylometric analysis**, which examines writing style features such as sentence length, punctuation usage, word frequency, and syntactic patterns. These features help identify subtle differences between human-authored and machine-generated texts.

With the advent of advanced AI models, researchers began leveraging **machine learning and deep learning techniques** to classify texts. Supervised learning models, such as **Support Vector Machines (SVM), Random Forests, and Logistic Regression**, have been trained on labeled datasets to distinguish AI-generated content from human writing. More recently, **neural networks and transformer-based models** have demonstrated higher accuracy by learning complex linguistic and contextual patterns that are difficult for traditional methods to capture.

Studies also indicate that AI-generated texts often exhibit **repetitive phrasing, unusual word distributions, and overuse of certain structures**, which can serve as distinguishing markers. Tools like **GLTR (Giant Language model Test Room)** analyze statistical patterns in token usage to identify AI-generated content, providing a probabilistic approach to detection.

In addition, researchers emphasize the importance of **real-time and automated detection systems** capable of handling large volumes of text from social media, academic submissions, and online publications. These systems combine NLP, statistical analysis, and machine learning to provide scalable solutions for content verification.

3. EXISTING SYSTEM

Currently, many existing systems rely on **manual review or traditional plagiarism detection tools** to evaluate text authenticity. While these systems can identify copied content, they are largely ineffective at detecting AI-generated texts because such content is **original in wording** yet lacks human-like nuances.

Some approaches use **stylometric analysis**, which examines writing patterns, word usage, sentence length, punctuation, and syntactic structures. While these methods can detect basic anomalies, they struggle with **advanced AI-generated texts** produced by large language models like GPT-3 or GPT-4, which mimic human writing styles more convincingly.

Other systems, such as **GLTR (Giant Language Model Test Room)**, use statistical methods to highlight tokens and word probabilities typical of AI-generated content. However, GLTR and similar tools often require **expert interpretation** and are not fully automated or scalable for large volumes of text.

Moreover, most existing solutions **lack real-time detection capabilities** and cannot be integrated seamlessly into content creation platforms or educational tools. They are often limited to batch processing, making them unsuitable for **high-speed, large-scale monitoring** of AI-generated content on social media or online publications.

These limitations highlight the need for a **more advanced, automated, and accurate system** that can detect AI-generated texts in real-time, using **machine learning and deep learning techniques** to analyze linguistic, semantic, and syntactic features effectively.

4.PROPOSED SYSTEM

The proposed system aims to develop an **automated, real-time framework** for detecting AI-generated texts with high accuracy and scalability. Unlike traditional systems that rely solely on manual review or stylometric analysis, this system leverages **Natural Language Processing (NLP)**, **machine learning**, and **deep learning techniques** to identify subtle differences between human-written and AI-generated content.

The system works in several stages. First, it performs **text preprocessing**, including tokenization, lemmatization, and removal of stop words, to standardize input data. Next, **feature extraction** is applied to capture linguistic, syntactic, and semantic characteristics of the text. Features may include **sentence structure, word distribution, punctuation patterns, semantic coherence, and contextual embeddings**.

After feature extraction, a **machine learning or deep learning classifier**—such as **Random Forest, Support Vector Machine (SVM), or transformer-based models like BERT**—is trained on a labeled dataset containing human-authored and AI-generated texts. The model learns patterns unique to AI-generated content, enabling accurate classification.

Finally, the system provides **real-time detection and reporting**, flagging content as AI-generated or human-written. The output can be integrated into educational platforms, publishing tools, and content verification systems, providing a **scalable and automated solution** for maintaining content integrity, preventing plagiarism, and mitigating the spread of misinformation.

5.METHODOLOGY

The proposed system follows a systematic methodology to detect AI-generated texts effectively. The process begins with **data collection**, where a diverse dataset containing both human-authored and AI-generated texts is gathered from sources such as articles, social media posts, and AI text generators. This dataset is then subjected to **preprocessing**, which includes **tokenization, lemmatization, removal of stop words, and normalization** to clean and standardize the text for analysis.

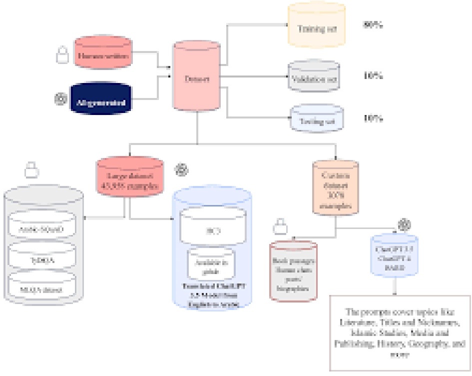
Next, **feature extraction** is performed to capture distinguishing characteristics of AI-generated content. These features include **sentence length, word frequency, punctuation usage, semantic coherence, syntactic patterns, and contextual embeddings**. Advanced embeddings from transformer models like **BERT** or **RoBERTa** can also be used to represent text in a high-dimensional semantic space.

The extracted features are then fed into **machine learning or deep learning classifiers** such as **Support Vector Machines (SVM), Random Forests, or neural networks**. These models are trained to differentiate AI-generated content from human-written text based on learned patterns and statistical anomalies.

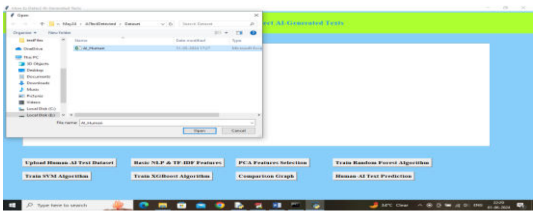
Finally, the system performs **real-time or batch classification**, flagging AI-generated texts and providing confidence scores. Visualization dashboards or reports can be used to summarize detection results. This methodology ensures **accuracy, scalability, and automation**, making it suitable for academic, publishing, and online content verification purposes.

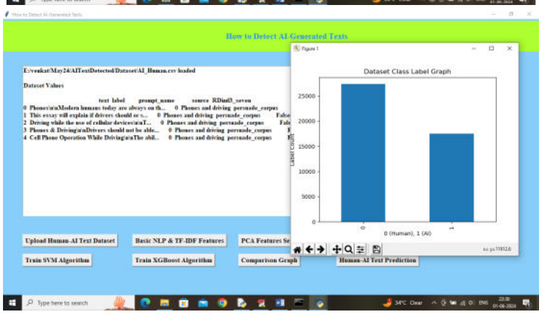
6.System Model

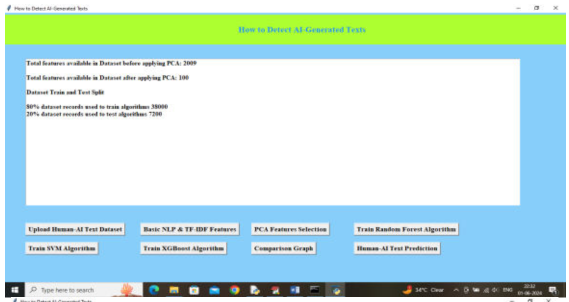
SYSTEM ARCHITECTURE

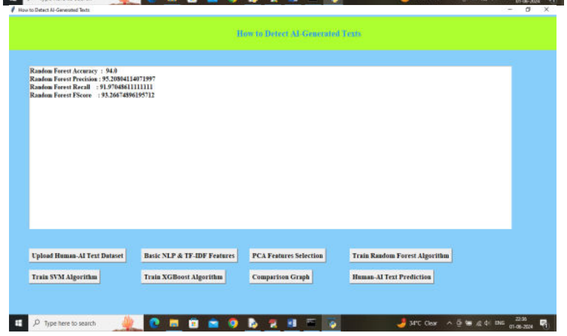


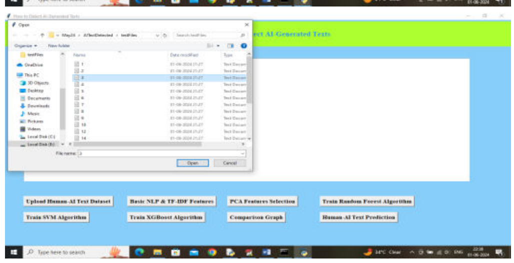
7.Results and Discussions

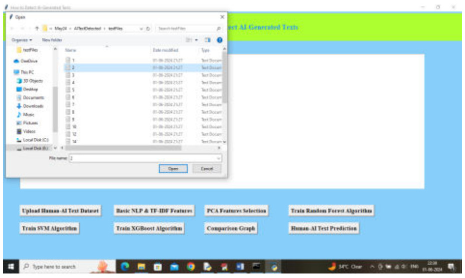


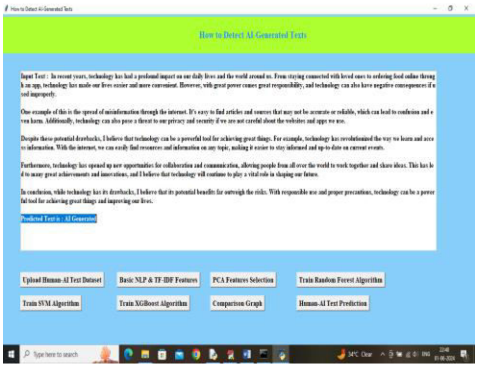












8. CONCLUSION

The project **"How to Detect AI-Generated Texts"** demonstrates an effective approach to addressing the growing challenge of identifying machine-generated content. By leveraging **Natural Language Processing (NLP), machine learning, and deep learning techniques**, the system can analyze linguistic, syntactic, and semantic patterns to distinguish AI-generated text from human-authored content.

Unlike traditional plagiarism detection tools or manual review methods, the proposed system offers **automated, scalable, and real-time detection**, making it suitable for academic institutions, publishers, and online platforms. By detecting AI-generated content accurately, the system helps maintain **content integrity, authenticity, and reliability**, mitigating risks associated with misinformation, plagiarism, and deceptive content creation.

Overall, this project highlights the importance of combining **AI-driven analysis and linguistic feature extraction** to tackle modern challenges in content verification. It provides a foundation for further research in **improving detection accuracy, handling multilingual content, and integrating the system with real-world platforms**, ensuring responsible and trustworthy use of AI-generated texts.

9.REFERENCES

1. Gehrmann, S., Strobelt, H., & Rush, A. M. (2019). *GLTR: Statistical Detection and Visualization of Generated Text*. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 111–116.
2. Zellers, R., Holtzman, A., Clark, E., et al. (2019). *Defending Against Neural Fake News*. Advances in Neural Information Processing Systems, 32.
3. OpenAI. (2023). *GPT-3 and GPT-4 Technical Documentation*. Retrieved from https://platform.openai.com/docs
4. Ippolito, D., Kriz, R., Hupkes, D., et al. (2020). *Automatic Detection of Generated Text is Easiest When Humans Are Fooled*. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 1809–1826.
5. Floridi, L., & Chiriatti, M. (2020). *GPT-3: Its Nature, Scope, Limits, and Consequences*. Minds and Machines, 30, 681–694.
6. Solaiman, I., Brundage, M., Clark, J., et al. (2019). *Release Strategies and the Social Impacts of Language Models*. arXiv preprint arXiv:1908.09203.
7. Jawahar, G., Sagot, B., & Seddah, D. (2019). *What Does BERT Learn About the Structure of Language?* Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 3651–3657.