International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 1 (2026)



ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

Distinguish Human Generated Text From ChatGPT Generated Text Using Machine Learning

Ms.K.Sai Durga, MCA - Assistant Professor, Department of MCA, Bapatla Engineering College, Bapatla, Andhra Pradesh

Ms.TATA VENKATA ANKA NAGA DEEPTHI, Reg No: Y25MC23085, Ms.PEESAPATI HARIKA, Reg No: Y25MC23060, Ms.PULUGUJJU SRAVANI, Reg No: Y25MC23066, Ms.VEGIRAJU HIRANMAYEE, Reg No: Y25MC23097, Department of MCA, Bapatla Engineering College, Bapatla, Andhra Pradesh.

***Abstract*—The big improvement in language models like ChatGPT has made it really hard to tell if a piece of text was written by a human or a computer. This is a problem for things like making sure students are not cheating controlling what people can post online and being honest when we talk to each other on the internet.**

This paper is about a system that uses machine learning to figure out if a piece of text was written by a human or a computer. The system looks at how the words and sentencesre put together what kind of words are used how well the text makes sense and if the meaning is consistent.

It checks the language patterns and statistics of the text to see if it was written by a human or a language model, like ChatGPT. A study is done to see how well eleven machine learning and deep learning algorithms work. These algorithms are used to classify things. The study looks at how they do, in different areas.

The study uses a dataset from Kaggle that has 10,000 pieces of text. Some of this text, 5,204 pieces was written by people. Comes from news articles and social media. The rest of the text was made by machines.

The text is cleaned up. Changed into something that the machine learning models can use to learn from. This helps the models get better at classifying things. The machine learning models are trained with this information. Additionally, the system incorporates visualization techniques to present real-time predictions, model accuracy, and performance metrics through graphical representations. The results demonstrate the effectiveness of the proposed approach in supporting reliable AI text detection and promoting responsible and transparent use of generative AI technologies.

Keywords: AI-generated text detection, Human–AI text classification, Machine learning, Natural language processing (NLP), Supervised learning, ChatGPT

I. INTRODUCTION

The fast appearance of intelligence that can create things has really changed how people talk to each other and make things. These new technologies are being used in lots of areas like news, art taking care of people, teaching and digital media. One type of intelligence that people really like is the kind that can have conversations, like ChatGPT. This is because it can make text that sounds like a human wrote it and it knows what it is talking about. Artificial intelligence like ChatGPT is very good, at making text that makes sense and is aware of what's going on. These models are really good for getting things done quickly and easily. They help a lot of people.. When we use them all the time it also causes some big problems. We have to think about the problems, the legal problems and the problems they cause for society. The use of these models raises concerns about what is right and wrong what is legal and what is not and how they affect people, in general.

The thing with intelligence that makes new things like text is that it can make things that are not true. This can help spread information around. It is hard to tell if something was written by a person or by intelligence because they look a lot alike. This can be used to trick people or organizations. That is bad for people believing what they are told and for trusting what they hear. Artificial intelligence that makes things like text is really good, at making things that look real so it is hard to know what to believe. Generative AI models are trained on big sets of data. This data may have some problems in it like biases. When we use these models to make decisions they can give us answers that're not fair or that discriminate against people. Generative AI models can be bad, at making choices because of the data they are trained on.

In school when students use much content that the computer makes it can hurt their ability to think for themselves be creative and talk to others. Also because these computer programs learn from information they might not be able to answer new or tough questions correctly. This is a problem because it shows we really need to find a way to tell if something was written by a person or a computer. We need to be able to tell the difference, between human-generated text and AI-generated text.

Natural Language Processing (NLP) plays a crucial

...

role in addressing this challenge by enabling machines to analyze and interpret human language through computational techniques. Recent advancements in machine learning and transformer-based architectures have significantly enhanced NLP capabilities, making automated text classification both feasible and effective. Motivated by these developments, this paper focuses on designing a machine learning–based approach to accurately identify and classify human-written and ChatGPT-generated text, thereby supporting responsible AI usage and digital integrity..

II. LITERATURE SURVEY

The rapid growth of large generative AI models has introduced significant regulatory, ethical, and societal challenges. Hacker et al. [1] discuss the urgent need for regulating large language models such as ChatGPT, focusing on risks related to misinformation, bias propagation, lack of transparency, and accountability. The authors propose a risk-based regulatory framework inspired by European Union governance models and emphasize the importance of mandatory AI impact assessments, human oversight, and public disclosure of AI-generated content to mitigate societal harm.

The security threats posed by AI-generated misinformation are further examined by Botha and Pieterse [2], who analyze the impact of fake news and deepfake technologies on global information security. Their study highlights how AI-driven content generation tools facilitate the rapid spread of disinformation, particularly through social media platforms. The authors review several detection approaches, including forensic analysis and machine learning-based classifiers, and stress the importance of integrated defense strategies to preserve public trust.

In the context of education, Qadir [3] explores the benefits and limitations of ChatGPT in engineering education. While generative AI can enhance learning through personalized tutoring and automated assistance, excessive dependence may lead to academic dishonesty and reduced critical thinking skills. The study advocates a hybrid educational model where AI supports educators rather than replacing them, emphasizing ethical AI literacy and curriculum redesign.

Cao et al. [4] present a comprehensive survey of AI-generated content, tracing the evolution of generative models from GANs to transformer-based architectures like ChatGPT. The authors analyze applications across multiple domains and address ethical challenges such as misinformation, identity spoofing, and intellectual property violations. The survey underscores the necessity of reliable detection mechanisms for synthetic content.

Finally, Khurana et al. [5] provide an extensive overview of the current state of Natural Language Processing, covering foundational techniques, deep learning models, and emerging trends. The authors highlight challenges such as language ambiguity, bias in datasets, and ethical concerns, reinforcing the relevance of machine learning-based approaches for distinguishing AI-generated text from human-written content.

III. ALGORITHM

Human vs ChatGPT Generated Text Classification Input: Text dataset containing human-written and ChatGPT-generated documents

Output: Predicted label indicating Human-generated or AI-generated text

1. Load the labelled dataset D D
2. For each text sample $d \in D$ do

2.1 Remove special characters, punctuation, and stop words

2.2 Convert text to lowercase

2.3 Tokenise the cleaned text

3. Extract numerical features from the processed text using TF-IDF or Bag-of-Words

4. Split the dataset into training set D train
D train and testing set D test
D test

5. Train multiple machine learning classifiers using D train

6. Evaluate each trained model on D test using accuracy, precision, recall, and F1-score

7. Select the best-performing model based on evaluation metrics

8. Use the selected model to classify unseen text samples

9. Visualize classification results and model performance metrics

IV. METHODOLOGY

The proposed methodology aims to accurately distinguish between human-generated text and ChatGPT-generated text using supervised machine learning techniques. The overall workflow consists of data collection, preprocessing, feature extraction, model training, evaluation, and visualization.

Data Collection : A publicly available Kaggle dataset is used for experimentation. The dataset consists of 10,000 text samples, including human-written content collected from news articles and social media platforms, along with

AI-generated text produced by ChatGPT. The dataset is balanced to ensure fair and unbiased model training.

Data Preprocessing : Preprocessing is performed to enhance data quality and consistency. This step includes converting text to lowercase, removing punctuation, special characters, stop words, and extra whitespace. Tokenization is applied to break text into meaningful units suitable for feature extraction.

Feature Extraction : The cleaned textual data is transformed into numerical representations using Natural Language Processing techniques such as Term Frequency–Inverse Document Frequency (TF-IDF) and Bag-of-Words. These features capture important linguistic patterns, vocabulary usage, and word distributions that help differentiate human and AI-generated text.

Model Training : Multiple machine learning and deep learning classifiers are trained using the extracted features. The dataset is split into training and testing sets using an 80:20 ratio. Supervised learning algorithms are employed to learn distinguishing patterns between the two classes.

Model Evaluation : The trained models are evaluated on unseen test data using standard performance metrics, including accuracy, precision, recall, and F1-score. A comparative analysis is conducted to identify the best- performing model.

Visualization : Graphical representations are used to visualize model performance, classification accuracy, and prediction results, enabling better interpretation and real-time analysis.

This structured methodology ensures reliable detection of AI-generated text and supports transparency and responsible use of generative AI systems.

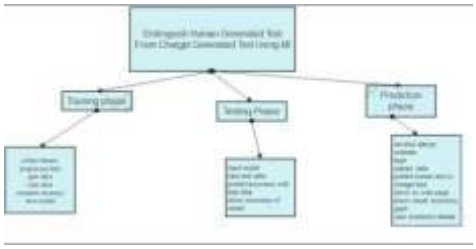


Fig. 1. Architecture Overview

V. SYSTEM ANALYSIS

Existing System :

When we try to figure out if a piece of text was written by a human or a computer we usually use methods. These methods are like recipes that people have used before. We use things like Bayes and Decision Trees because they are easy to understand and do not take a lot of computer power. Some systems also use rules that people have made like how often words are used what the writing style is, like how easy it is to read or tools that check if someone copied something from somewhere else. We use these rules to help us tell human-written text from AI-generated text and human-written text is what we are trying to find.

The thing, with Bayes is that it thinks features are independent which does not work well for natural language data. Naive Bayes has this problem.

KNN classifiers have a time handling big sets of text data.

They are slow.

Decision Trees are not great either. They. Do a bad job of generalizing.

Heuristic-based methods are not very good because AI-generated text can easily get around the rules by saying things a way.

Plagiarism detection tools do not work when the AI content is new and made up.

Readability-based metrics are not good enough because they do not really understand what the text means or the context.

Naive Bayes and these other methods have a lot of problems. The systems we have now are not very good, at dealing with things that change. They also have a time handling new language patterns that are made by artificial intelligence. These language patterns are always. The systems we have now are just not able to keep up with them. Existing systems lack the strength and flexibility they need to handle this. They need to be able to adapt to things and handle the changing language patterns that artificial intelligence is making.

5.1.1 Disadvantages of Existing System

1. Low classification accuracy due to shallow linguistic feature modelling
2. Poor scalability for large-scale or real-time text analysis
3. Limited adaptability to newer and more advanced generative AI models
4. Inadequate semantic and contextual understanding
5. Vulnerability to paraphrasing and stylistic manipulation

5.2 Proposed System

The new system is a way to tell if a piece of writing was written by a person or by a computer program like ChatGPT. It starts by gathering a lot of texts that people have written and texts that ChatGPT has written. Then it makes all these texts clean and simple so the system can understand them. This means it gets rid of any stuff breaks the text into small

parts and makes everything the same. After that it looks for things, about the texts like what words are used and how they are used to figure out what makes them different. The system uses things called TF-IDF and word embeddings to do this.

The things we find out about the text capture what the words mean how they sound and how they go together in a sentence. We look at things like how words are used how complicated the sentences are and how the words relate to each other. We try out different models to see which one works best including Logistic Regression, Support Vector Machine, Random Forest Classifier and Convolutional Neural Networks. We. Compare many models, like Logistic Regression, Support Vector Machine, Random Forest Classifier and Convolutional Neural Networks to see which one is the best.

Logistic Regression is a way to classify things into groups. It is the starting point for complicated methods. On the hand SVM is good at finding complicated patterns in data by using special helper functions.

Random Forest is another method that makes the results more reliable by combining different models. It also helps us understand which features are the important.

CNN is used to learn a lot about text by looking at the relationships between words and phrases. This helps us find differences in style and context that might be hard to see otherwise.

We use accuracy and other measures like precision and recall and F1-score to see which classifier is the best at its job. Logistic Regression and the other methods, like SVM and Random Forest and CNN are all compared to find the one that works the best.

5.2.1 Advantages of Proposed System

1. Higher detection accuracy through advanced feature representation

2. Effective handling of large and diverse text datasets

3. Scalability for real-time and multi-source applications

4. Improved adaptability to evolving AI-generated text patterns

5. Enhanced semantic and contextual understanding using deep learning

6. Supports digital content authenticity, transparency, and accountability

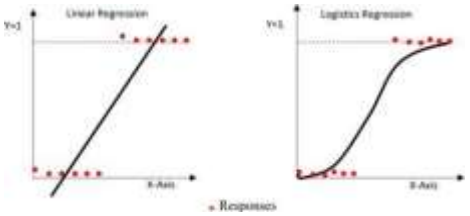


Fig. 2. Logistic Regression

VI. SYSTEM TESTING

Testing and Validation: Testing is a critical phase in the development of the proposed system, ensuring correctness, reliability, and performance before deployment. The primary objective of testing is to identify errors, validate system functionality, and verify that the implemented model meets the specified requirements efficiently. Quality of test data is prioritized over quantity, and representative sample datasets are used for evaluation.

Testing Objectives: The main objectives of testing are to systematically detect errors with minimal effort and time, validate expected outputs, and ensure that the system conforms to reliability and quality standards. A successful test case is one that uncovers previously undiscovered errors, while an effective testing process confirms system robustness and correctness.

Levels of Testing:

1. Unit Testing: Unit testing is performed on individual modules such as dataset loading, preprocessing, data splitting, and model training. Each module is tested independently to verify correct functionality and expected outputs. Validation checks are applied to ensure correct handling of user inputs and dataset configurations. Test results confirmed that all modules performed as expected.

2. Black Box Testing: Black box testing evaluates the system based on functional requirements without considering internal implementation details. The focus is on validating system behavior by providing valid and invalid inputs and comparing actual outputs with expected results. Functional testing, non-functional testing, and regression testing were conducted to ensure accuracy, performance, and stability of the system.

3. White Box Testing : White box testing focuses on the internal logic, control flow, and structure of the system. This testing verifies conditional loops, data flow, and algorithm execution paths to identify logical errors and security vulnerabilities. Test cases were designed to validate correct flow of inputs and outputs across all processing stages.

4. Integration Testing : Integration testing is conducted to verify the interaction between combined modules,

including feature extraction, model training, and performance evaluation. Both bottom-up and top-down integration strategies are applied to ensure smooth data flow and correct interaction among system components. The results confirm successful integration with no functional conflicts.

5. System Testing : System testing evaluates the complete, integrated system in an environment similar to real-world deployment. The system is tested for functional correctness, performance, and operating system compatibility. Experimental results show consistent and reliable performance across different operating systems, with optimal performance observed on Windows 7 and later versions.

Overall, the testing results demonstrate that the proposed system is stable, accurate, and reliable, satisfying both functional and technical requirements for effective detection of AI-generated text.

VII. RESULT ANALYSIS

This part shows what we found out when we tested the system that uses machine learning to tell if a piece of text was written by a human or by ChatGPT. We tried out a lot of models to see which one was best, at figuring out if something was written by a machine. The goal was to find out how well these models could detect text that was generated by intelligence specifically text that was generated by ChatGPT.

The dataset was split into two parts the training dataset and the testing dataset with 80 percent of the data used for training and 20 percent used for testing. We then looked at how the models performed using simple measures, like accuracy, precision, recall and the F1-score of the dataset to get a complete picture of how the models compared to each other and to really understand the performance of the models and the dataset.

I think Logistic Regression is a model that works well. It is a baseline to compare other models to. However it has a problem. It can only make decisions. This means it is not very good at understanding language patterns in text that was generated by Artificial Intelligence.

The Support Vector Machine model is actually better at this. It can understand relationships between different features even when there are a lot of them. This makes it very good, at classifying text. Support Vector Machine is really effective when it comes to classifying text because it can handle these relationships.

The Random Forest Classifier did a job than most traditional models because it used ensemble learning. This helped to reduce overfitting and made the model work better with data. The Random Forest Classifier can also look at which features are important. This gave us some ideas about what linguistic attributes make text different from AI-generated text. The Random Forest Classifier is really good, at showing us which attributes are influential.

The Convolutional Neural Network achieved the results overall. It did this by learning from text in a way that understood how things are connected and what they mean. The Convolutional Neural Network was really good at finding differences in how words are used and what they mean in text made by artificial intelligence. This meant the Convolutional Neural Network was more accurate and got scores, than other models. The Convolutional Neural Network was able to do this because it learned from text embeddings in a way that let it see patterns that other models could not see.

Overall, the results indicate that advanced machine learning and deep learning models significantly improve detection accuracy. The proposed system demonstrates robustness, scalability, and adaptability, making it suitable for real-world applications such as academic integrity verification, content moderation, and digital media authentication.



Fig : 3. Main Page



Fig: 4. Login page



Fig : 5. Checking



Fig 6. Result

VIII. CONCLUSION

This paper presented a machine learning–based approach for distinguishing human-generated text from ChatGPT-generated text, addressing a growing challenge posed by the rapid adoption of generative AI systems. By leveraging Natural Language Processing techniques and supervised learning models, the proposed system effectively captures linguistic, syntactic, and semantic patterns that differentiate human-authored content from AI-generated text.

A comparative analysis of multiple machine learning and deep learning models demonstrated that advanced classifiers significantly outperform traditional approaches. In particular, deep learning models such as Convolutional Neural Networks showed superior performance due to their ability to learn contextual and hierarchical representations from text data. The experimental results confirm that the proposed system is accurate, robust, and scalable for real-world applications.

The findings highlight the importance of automated AI-text detection in maintaining academic integrity, combating misinformation, and ensuring transparency in digital communication. Future work may focus on extending the system to detect content generated by newer language models, incorporating multilingual datasets, and exploring transformer- based architectures for further performance improvement. Overall, the proposed approach contributes toward responsible and trustworthy use of generative AI technologies.

REFERENCES

[1]P. Hacker, A. Engel, and M. Mauer, “Regulating chatgpt and other large generative ai models,” arXiv preprint arXiv:2302.02337, 2023.

[2] J. Botha and H. Pieterse, “Fake news and deepfakes: A dangerous threat for 21st century information security,” in ICCWS 2020 15th International Conference on Cyber Warfare and Security. Academic Conferences and publishing limited, 2020, p. 57.

[3] J. Qadir, “Engineering education in the era of chatgpt: Promise and pitfalls of generative ai for education,” 2022.

[4] Y. Cao, S. Li, Y. Liu, Z. Yan, Y. Dai, P. S. Yu, and L. Sun, “A comprehensive survey of AI-generated content (aigc): A history of generative ai from gan to chatgpt,” arXiv preprint arXiv:2303.04226, 2023.

[5] Rongali, L. P. (2025). Green DevOps Metrics for Utility Operations. https://doi.org/10.36227/techrxiv.175433211.13655773/v1.

[6] Rongali, L. P. (2025). The Trinity of Cybersecurity: DevSecOps, Cloud Security and Cryptography in The Digital Age. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5229585

.

[7] Nandigama, N. C. (2023). Enhanced Fingerprint Recognition system using hybrid Feature Fusion with Deep learning and Machine learning Optimization. Research Journal of Nanoscience and Engineering, 6(1), 9–15. https://doi.org/10.22259/2637-5591.0601003

[8] Nandigama, N. C. (2021). Advancing Underwater Image Segmentation through Pix2Pix Generative Adversarial Networks. Research Journal of Nanoscience and Engineering, 5(1), 20–25. https://doi.org/10.22259/2637-5591.0501004

[9] T. Zhang, A. M. Schoene, S. Ji, and S. Ananiadou, “Natural language processing applied to mental illness detection: a narrative review,” NPJ digital medicine, vol. 5, no. 1, p. 46, 2022.

[10] Todupunuri, A. (2025). The Role of Human-Centric AI in Building Trust in Digital Banking Ecosystems. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5120605

[11] C.-F. R. Chen, Q. Fan, and R. Panda, “Crossvit: Cross- attention multi scale vision transformer for image classification,” in Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 357 366.

[12] A. Gillioz, J. Casas, E. Mugellini, and O. Abou Khaled, "Overview of the transformer-based models for nlp tasks," in 2020 15th Conference on Computer Science and Information Systems (FedCSIS). IEEE, 2020, pp. 179–183.

[13] Todupunuri, A. (2025). IMPROVING CUSTOMER EXPERIENCE WITH MODERN BANKING SOLUTIONS. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5120615

[14] S. Singh and A. Mahmood, "The nlp cookbook: modern recipes for transformer based deep learning architectures," IEEE Access, vol. 9, pp. 68675–68702, 2