

Implementation of AI vs Human Text Detection

Prashanth Kumar HM¹

Student: Dept. of Computer Science
Srinivas University, Mangalore
Karnataka, India

Dr. Subramanya Bhat S²

Professor: Dept. of Computer Science
Srinivas University, Mangalore
Karnataka, India

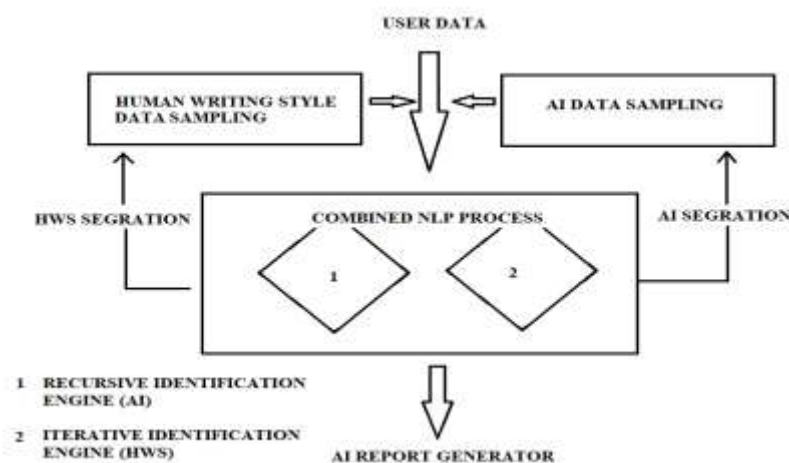
Abstract: AI text refers to text that has been generated or processed using artificial intelligence (AI) technologies. AI text can be produced using natural language processing (NLP) models. These models are trained on large datasets and can understand and generate human-like text. It is developed by expertized neural and NLP algorithm; such sort of algorithm automatically understands user input and generate sentences with systematic manner. AI text generation has both positive and negative implications. While it can automate various tasks, it also raises concerns about misinformation, bias, and ethics. As AI technology continues to advance, it's essential to use it responsibly and be aware of its potential impact on society. Recent AI language models have taken a significant step forward in producing remarkably controllable, fluent, and grammatical text based on pre-defined algorithms set. We implemented AI versus Human text detection using the features (implemented algorithm), i.e., writing style, coherence, consistency, and argument logistics. Finally, we developed tool to investigate the segregation between AI-generated scientific text and human-written scientific text by AI-generated scientific text detection models.

Keywords: AI Text, Human Text, Natural Language Processing, Algorithm, Data Sampling.

I. INTRODUCTION

In the realm of artificial intelligence, the concept of AI text has emerged as a fascinating and powerful application. AI text refers to text that is generated or processed using artificial intelligence technologies, such as natural language processing (NLP) models. These NLP models, like the one used in this interaction, are trained on massive datasets, and can comprehend and generate human-like text. One of the most remarkable uses of AI text is seen in chatbots. AI-powered chatbots have revolutionized customer support and engagement in various industries. By utilizing AI text generation capabilities, these chatbots can hold conversations with users, answering questions, and helping. This advancement has not only improved the efficiency of customer service but has also led to more personalized interactions. Another important application of AI text is in language translation. Thanks to sophisticated NLP algorithms, AI can accurately translate text from one language to another, breaking down language barriers and facilitating communication between people from different linguistic backgrounds. This has significant implications for global business, travel, and international relations. Content generation is yet another area where AI text excels. AI can be employed to generate articles, product descriptions, and other written content, saving time and effort for content creators. However, there are concerns about potential misuse, as AI-generated content could be used to spread misinformation or create fake news. AI text is also capable of analyzing sentiment in text. Sentiment analysis involves determining the emotional tone expressed in a piece of writing. This technology is utilized in social media monitoring, market research, and brand management, among other applications. In speech recognition, AI can transcribe spoken language into written text with remarkable accuracy. This advancement has made voice-controlled digital assistants like Siri and Alexa possible, enhancing the way we interact with technology. Moreover, AI text has enabled the development of text summarization tools. These tools can automatically condense large documents or articles into concise summaries, making it easier for users to grasp the main points of the content without reading the entire document. However, while AI text offers incredible possibilities, it also comes with challenges. One of the main concerns is bias. Since AI models are trained on vast datasets, they can inadvertently perpetuate societal biases present in the data. Efforts are being made to reduce bias and improve fairness in AI text generation. In conclusion, AI text represents a powerful advancement in artificial intelligence. Its applications in chatbots, language translation, content generation, sentiment analysis, speech recognition, and text summarization have transformed various industries. Nevertheless, it is crucial to use AI text responsibly and address potential ethical and societal issues. In the exciting field of artificial intelligence, a new and powerful application called AI text has emerged. AI text refers to text that is generated or processed using artificial intelligence technologies, particularly through sophisticated natural language processing (NLP) models. These models, such as the ones powering our conversation, have been trained on vast datasets, allowing them to understand and produce human-like text. One of the most impressive applications of AI text can be found in the realm of chatbots. These AI-powered virtual assistants have revolutionized customer support and engagement across various industries. By leveraging the capabilities of AI text generation, chatbots can hold meaningful conversations with users, addressing their queries, and helping. This has not only improved the efficiency of customer service but has also enabled more personalized interactions between businesses and their customers. Language translation is another area where AI text shines. With the help of advanced NLP algorithms, AI can accurately translate text from one language to another, breaking down language barriers and fostering better communication between individuals from different linguistic backgrounds. This has significant implications for international relations, business, and travel, making the world more interconnected.

AI text has also made significant strides in content generation. By harnessing the power of artificial intelligence, it can create articles, product descriptions, and other written content, streamlining the content creation process for businesses and individuals. However, the rise of AI-generated content also raises concerns about the potential misuse of such technology for spreading misinformation or generating fake news. Another impressive application of AI text is in sentiment analysis. This technology enables AI to determine the emotional tone expressed in a piece of text, making it useful for social media monitoring, market research, and brand management, among other applications. Additionally, AI text has greatly improved speech recognition capabilities. It can accurately transcribe spoken language into written text, paving the way for voice-controlled digital assistants like Siri and Alexa, which have become an integral part of our daily lives. Furthermore, AI text has facilitated the development of text summarization tools. These tools can automatically condense lengthy documents or articles into concise summaries, making it easier for users to grasp the main points without reading the entire content. However, despite its many benefits, AI text also poses challenges. One of the main concerns is bias in AI-generated content. Since AI models are trained on vast and sometimes biased datasets, they can inadvertently perpetuate societal biases present in the data. Efforts are being made to address this issue and improve the fairness of AI text generation. In conclusion, AI text is a powerful and transformative application of artificial intelligence. Its contributions to chatbots, language translation, content generation, sentiment analysis, speech recognition, and text summarization have reshaped various industries. Nevertheless, it is essential to use AI text responsibly and ethically, ensuring that the benefits it brings are maximized while mitigating potential drawbacks. Our tool comes with all levels of differentiation in the sentences which comes under input fields, those fields having a collection of more phrases. Our AI Algorithms expressed the value of phrases and make it different categories like huma text and AI generated with number of potentials. Our architecture shows the process of AI detection system below.



When a user submitted data to an AI detection model, the AI supported Combined Natural Language Processing (CNLP) will separate AI text with Human text by using 2 different levels called Recursive Identification for AI text and Iterative identification for Human Writing Style. The CNPL is purely AI process its segregate AI and HWS combined phrases and store related sampling data base for next-generation process, here a greater number of AI detection process will generate more accuracy, ex during first 100 files of AI detection may get up to 10 % accuracy, next second times 100 files of AI detection may get up to 20 % accuracy. Till if we run up to 5000 files for AI detection may increase to 90-95 % accuracy, because every file processing time CNLP engine is released AI and Human Text unique data sampling and stored to sampling data base with avoiding duplicate sample phrases. Before sampling databases required minimum data samples, the developer should collect and store manually and it's a onetime operation.

II. OBJECTIVES

Data Sampling: Data sampling is a technique used in statistics and data analysis to select a subset of data points from a larger data set for the purpose of analysis or modeling. The process of data sampling involves choosing a representative sample that can accurately reflect the characteristics and distribution of the entire dataset. This is often done to save computational resources, reduce processing time, or obtain insights from a manageable amount of data. Here it is helping us to separate AI data sampling and Human Text Data Sampling with different layers.

AI-Data: AI data sampling is a sub process of machine learning. In the context of AI and machine learning, AI data sampling plays a vital role in training, validating, and evaluating machine learning models. AI-data sampling involves selecting a subset of data from a larger AI dataset to create representative and balanced training and testing sets for machine learning algorithms. This process is crucial for building accurate and effective AI models. It's useful for data sampling process to make separate data sets for two levels of data handling for detection tool.

Human writing style: Human text data sampling is also a sub process of machine learning. In the context of Human text and machine learning, it plays a vital role in training, validating, and evaluating machine learning models. This level involves selecting a subset of data from a larger internet or student paper dataset to create representative and balanced training and testing sets for machine learning algorithms. This process is crucial for building accurate and effective AI models. It's useful for data sampling process to make separate data sets for two levels of data handling for detection tool.

Coherence Style: Coherence in writing refers to the logical and smooth flow of ideas throughout a piece of text. It is a crucial aspect of effective communication as it helps readers understand the intended message without confusion or ambiguity. When a piece of writing is coherent, each sentence, paragraph, and section connect logically and naturally to the next, forming a cohesive and unified whole. Ideas should be organized in a logical sequence, with one idea naturally leading to the next. This could be done chronologically, by cause and effect, by importance, or in any other coherent order. It's created a writing enables readers to follow the author's thoughts and arguments effortlessly, leading to a more engaging and effective piece of communication.

Sentence Clarity: Sentence clarity is the ability of a sentence to convey its intended meaning in a straightforward and understandable manner. Clear sentences are free from ambiguity, jargon, and convoluted phrasing. Achieving clarity involves using simple and precise language, avoiding wordiness, and organizing thoughts logically. Proper sentence structure, correct grammar, and punctuation also play crucial roles. Additionally, considering the context and audience when constructing sentences helps ensure comprehension. Well-crafted, clear sentences enhance communication, making reading easier and more enjoyable. Authors, speakers, and communicators should prioritize sentence clarity to effectively convey their ideas and messages to their intended recipients.

Intercommunication: While writing sentences refers to the exchange of information and ideas between the sentences and the reader through the written text. Effective intercommunication is achieved when the author can convey their intended message clearly, and the reader can understand and interpret the message accurately at every level, most of AI generated text is followed same while generation sentences. The author should use clear and unambiguous language, avoiding jargon or overly complex phrasing that may confuse or overrule the reader. But it did not happen in AI text generator.

Grammar Clarity: Grammar is a set of rules that govern the structure and composition of a language. It is a fundamental aspect of writing, as it helps convey ideas clearly, accurately, and effectively. Proper grammar ensures that sentences are well-formed, coherent, and understandable to readers. Maintaining proper grammar enhances the quality and credibility of your writing, ensuring that your ideas are effectively communicated to your readers. Most AI generated text followed the grammar rule to generate phrases. Here we are assuming 90% of AI text have not any issue.

Consistency: Consistency in writing refers to maintaining uniformity and coherence throughout a piece of text. It involves using the same conventions, style, and formatting choices consistently to create a smooth and harmonious reading experience. Consistency in writing helps readers focus on the content rather than being distracted by variations in style or formatting. It instills confidence in the author's credibility and professionalism. Maintaining consistency is especially important in longer pieces, such as academic papers, reports, or novels, where it ensures that the writing feels cohesive and polished. Proofreading and editing play a vital role in checking for and rectifying any inconsistencies that may have inadvertently crept into the text.

III. IMPLEMENTATION

Combined Natural Language Processing: Combined Natural Language Processing (CNLP) is a subfield of artificial intelligence (AI) and computational linguistics that focuses on the interaction between humans and computers using natural language. The primary goal of CNLP is to enable software to understand and segregate AI and Human Text unique data sampling and stored to sampling data. CNLP involves a wide range of tasks and techniques that deal with the processing and analysis of AI and human language, including:

Recursive AI Identification: It's a process between user data and AI sampling techniques, the main goal is to summarize the collective AI models using user defined complicated algorithms. This algorithm is implemented using arithmetic recursive 'Quadratic' equations called $[2(n)2+1]$ terms. This formula can be fixed with different looping stages, a single stage having multiple functional operations which support identified and segregation purposes in the AI detection process. The below algorithm shows basic steps of the recursive AI identification process.

Step 1: Analysis conigcutive phrases

Step 2: Tokanizing the phrases in loop where

```
for i=1; to 10 (where 1 is min. and 10 is max. phrases)
then Tokenization end, goto step 3.
```

Step 3: Tagging phrases (e.g., noun, verb, adjective)

```
for i=0; to N (where N is total number of words)
{
for j=0 to M (where M is total part of speech words)
{
combination: value of j with value of i;
} end of loop i
} end of loop j then goto step 4
```

Step 4: Recursive Method: Named_Entity_Recognition_NET()

```
{
initilize: array people[]=AI_Samples;
organizations[]=AI_Samples
locations[]=AI_Samples
dates[]=AI_Samples
//call method untill end of phrases
Recursive Method: Named_Entity_Recognition_NET()
```

```

} end of NET() then goto step 5
Step 5: Recursive Method: Sentiment_Analysis_SA()
{
  initialize: array +ve[]=AI_Samples;
             -ve[]=AI_Samples
             natural[]=AI_Samples
             //call method untill end of phrases
  Recursive Method: Sentiment_Analysis_SA()
} end of SA() then goto Step 6
Step 6: Recursive Method: Objectives()
{
  initialize: array obj_1[]=AI_Samples;
             obj_2=AI_Samples
             obj_n=AI_Samples
             //call method untill end of phrases
  Recursive Method: Sentiment_Analysis_SA()
} end of SA() then goto Step 7
Step 7: Recursive Method: Sampling()
{
  //Combination model 1
  Combination_1 []=Named_Entity_Recognition_NET();
  Samples_1[]=Sentiment_Analysis_SA()+Objectives();
  //Combination model 2
  Combination_2 []=Sentiment_Analysis_SA();
  Samples_2[]=Named_Entity_Recognition_NET()+Objectives();
  //Combination model 3
  Combination_3 []=Objectives();
  Samples_3[]=Named_Entity_Recognition_NET()+Sentiment_Analysis_SA();
  //Recursive Sampling
  Sampling();
} end of SA() then goto Step 8
Step 8: Stop

```

Algorithm Iterative HWS Identification: It's a process between user data and Human text sampling techniques, the main goal is to summarize the collective Human writing style models using user defined complicated algorithms. This algorithm is implemented using arithmetic 'Runge-Kutta' equations using an iterative method called $[y_1 = y_0 + (\frac{1}{2})(k_1 + k_2)]$ terms. This formula can be fixed with different looping stages, a single stage having multiple functional operations which support identified and segregation purposes in the human writing style detection process. The below algorithm shows basic steps of the iterative Human text identification process.

```

Step 1: Analysis conigcutive phrases
Step 2: Tokanizing the phrases in loop where
       for i=1; to 10 (where 1 is min. and 10 is max. phrases)
       then Tokenization end, goto step 3.
Step 3: Tagging phrases (e.g., noun, verb, adjective)
       for i=0; to N (where N is total number of words)
       {
         for j=0 to M (where M is total part of speech words)
         {
           combination: value of j with value of i;
         } end of loop i
       } end of loop j then goto step 4
Step 4: Iterative Method: Named_Entity_Recognition_NET()
       {
         for i=0; to N (where N is total number of samples)
         {
           initialize: array people[]=HWS_Samples;
                     organizations[]=HWS _Samples
                     locations[]=HWS _Samples
                     dates[]=HWS _Samples
         } end of loop
       } end of NET() then goto step 5
Step 5: Iterative Method: Sentiment_Analysis_SA()
       {
         for i=0; to N (where N is total number of samples)
         {

```

```

initilize: array +ve[]=HWS _Samples;
          -ve[]=HWS _Samples
          natural[]=HWS _Samples
          //call method untill end of phrases
          Iterative Method: Sentiment_Analysis_SA()
        } end of loop
      } end of SA() then goto Step 6
Step 6: Iterative Method: Objectives()
      {
        initilize: array obj_1[]=AI_Samples;
                obj_2= HWS _Samples
                obj_n= HWS _Samples
                //call method untill end of phrases
                Iterative Method: Sentiment_Analysis_SA()
      } end of SA() then goto Step 7
Step 7: Iterative Method: Sampling()
      {
        //Combination model 1
        Combination_1 []=Named_Entity_Recognition_NET();
        Samples_1[]=Sentiment_Analysis_SA()+Objectives();
        //Combination model 2
        Combination_2 []=Sentiment_Analysis_SA();
        Samples_2[]=Named_Entity_Recognition_NET()+Objectives();
        //Combination model 3
        Combination_3 []=Objectives();
        Samples_3[]=Named_Entity_Recognition_NET()+Sentiment_Analysis_SA();
      } end of SA() then goto Step 8
Step 8: Stop

```

Using the above two major algorithms we are achieve these tasks, such as machine learning algorithms and linguistic rules (e.g., grammar, syntax and other way the NLP models and have led to breakthroughs in many NLP tasks fro resolving text complication.

IV. RESULT AND COMBINATION

Our accuracy level is quite good compared with others global AI detector tools. We are providing up to 99 % accuracy for AI text detector and below 1 % accuracy for false positive detector. The analysis page of AI detection page is shown below.

The screenshot displays the DriftBit AI detector interface. At the top, it shows a blue header with the DriftBit logo and a large blue box indicating an AI detection score of 86%. Below this, there are three bullet points describing the tool's features:

- The DriftBit AI detector model has been developed to identify AI generated text from tools like CHATGPT.
- It is designed to provide a percentage estimator of AI generated content and highlight specific sections in a document.
- In the absence of specific guidelines from the UGC or academic institutions regarding the utilization of AI tools in academic writing, DriftBit AI model serves as a preliminary indicator of AI generated text within a document.

On the right side of the interface, there is a text snippet about the Sikh Empire and the British East India Company, which is the content being analyzed by the detector.

We are expanding our result combination in 3 ways, those are.

1. Complete AI

It means the data should be completely copied from any AI text generator tool like ChatGPT3, 4 or Broad etc... the data can be altered or modification up to 30% only, data contains at least minimum 200 words and there is no maximum words limit. We will give up to 95 - 99 percent accuracy while data copying from AI tool. Below image showing AI 100 percent similarity and all data copied from chatGPT3 AI text generator tool.

AI - 100%

- The DrillBit AI detector model has been developed to identify AI generated text from tools like CHATGPT.
- It is designed to provide a percentage estimation of AI generated content and highlight specific sections in a document.

Sentence clarity is the ability of a sentence to convey its intended meaning in a straightforward and understandable manner. Clear sentences are free from ambiguity, jargon, and convoluted phrasing. Achieving clarity involves using simple and precise language, avoiding wordiness, and organizing thoughts logically. Proper sentence structure, correct grammar, and punctuation also play crucial roles. Additionally, considering the context and audience when constructing sentences helps ensure comprehension. Well-crafted, clear sentences enhance communication, making reading easier and more enjoyable. Authors, speakers, and communicators should prioritize sentence clarity to effectively convey their ideas and messages to their intended recipients. Sentence clarity is the ability of a sentence to convey its intended meaning in a straightforward and understandable manner. Clear sentences are free from ambiguity, jargon, and convoluted phrasing. Achieving clarity involves using simple and precise language, avoiding wordiness, and organizing thoughts logically. Proper sentence structure, correct grammar, and punctuation also play crucial roles. Additionally, considering the context and audience when constructing sentences

A sample of 10 papers is checked and mentioned below report, here we observed overall accuracy of AI detection is 97 %, all data completely copied from ChatGPT-3 and there is no alteration is made inside report. The column AI report DrillBit showing percentage of AI detection, and this is integrated with DrillBit Plagiarism software.

Paper title	Description	Paper ID	AI report DrillBit
Paper1	complete chatgpt data	811513	92%
Paper2	complete chatgpt data	811524	100%
Paper3	complete chatgpt data	811535	83%
Paper4	complete chatgpt data	811537	100%
Paper5	complete chatgpt data	811538	97%
Paper6	complete chatgpt data	811539	100%
Paper7	complete chatgpt data	811540	98%
Paper8	complete chatgpt data	811492	100%
Paper9	complete chatgpt data	811493	100%
Paper10	complete chatgpt data	811494	100%

2 Complete non - AI data

It means the data should be completely copied from any internet site or student paper. Here the data can be altered, or modification won't affect AI similarity and data contains at least a minimum of 200 words and there is no maximum words limit. We will give only 0 percent similarity while data copying from any internet site called source is : <https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals> . Below image showing AI 0 percent similarity, because its Humna written text.

AI - 0%

- The DrillBit AI detector model has been developed to identify AI generated text from tools like CHATGPT.
- It is designed to provide a percentage estimation of AI generated content and highlight specific sections in a document.

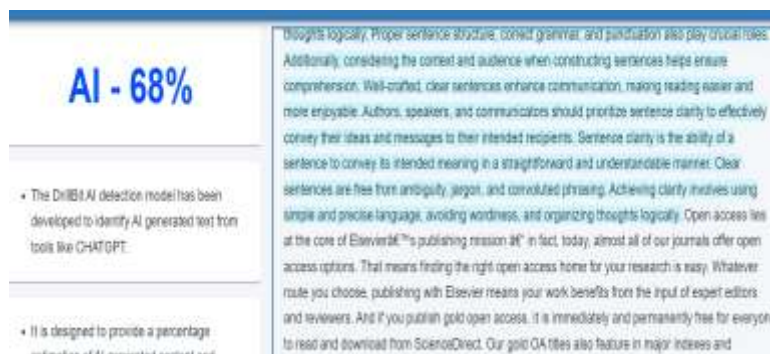
Open access lies at the core of Elsevier's publishing mission - in fact, today, almost all of our journals offer open access options. That means finding the right open access home for your research is easy. Whatever route you choose, publishing with Elsevier means your work benefits from the input of expert editors and reviewers. And if you publish gold open access, it is immediately and permanently free for everyone to read and download from ScienceDirect. Our gold OA titles also feature in major indexes and databases. Every year, we accept and publish more than 470,000 journal articles. Publishing in an Elsevier journal starts with finding the right journal for your paper. If you already know the journal to which you want to submit, you can enter the title directly in the find a journal search box. Alternatively, you can match the abstract of your article to a journal. On an annual basis we donate to many initiatives to support the global computer science community through conference sponsorships, prizes and funding. Many of our journals include young researcher awards. Reproducibility Badge Initiative (RBI) is a collaboration with the Code Ocean (CO), a cloud based computational reproducibility platform that helps the community by enabling sharing of code and data as a resource for non-commercial use. CO verifies the submitted code (and data) and certifies its reproducibility. Code submission will be verified by the Code Ocean team for computational reproducibility by making sure it runs, delivers results and it

A sample of 10 papers is checked and mentioned in the below report, here we observed overall accuracy of AI detection is 2.2 %, then the other 97.8 % contains Internet data or non-AI content. Here all data is completely copied from different internet domains and there is no alteration made inside the report, and alteration is not necessary. The column AI % DrillBit shows the percentage of AI detection.

	Paper ID	Source	AI % DrillBit
File 1	838649	Internet Source	0
File 2	838651	Internet Source	0
File 3	838653	Internet Source	0
File 4	838655	Internet Source	12
File 5	838657	Internet Source	0
File 6	838659	Internet Source	0
File 7	838662	Internet Source	0
File 8	838664	Internet Source	0
File 9	838666	Internet Source	10
File 10	838667	Internet Source	0

3. Mixed data

It means the data should be some of data copied from any AI text generator tool like ChatGPT3, 4 or Broad etc., and other data copied from any internet data, the AI data can be altered or modification up to 30%, if necessary, only, the data contains at least minimum 200 words and there is no maximum words limit. We will give accuracy while data is merged with two ends. Below image showing AI 68 percent similarity and remaining nonmatching content contains from human written style.



A sample of 10 papers is checked and mentioned below report, here we observed overall accuracy of AI detection is almost like expected percentage, all data copied from ChatGPT-3 and different internet domains. There is no alteration made inside the report. The column AI report DrillBit showing percentage of AI detection and showing best accuracy with mixed content.

Title	Paper ID	Sources	Expected AI%	AI % DrillBit
Sample_01	870914	ChatGPT+Internet	23.80%	20%
Sample_02	870915	ChatGPT+Internet	23.40%	24%
Sample_03	870916	ChatGPT+Internet	20.50%	22%
Sample_04	870917	ChatGPT+Internet	20.60%	21%
Sample_05	870918	ChatGPT+Internet	21.10%	22%
Sample_06	870919	ChatGPT+Internet	23.60%	22%
Sample_07	870920	ChatGPT+Internet	20.60%	18%
Sample_08	870921	ChatGPT+Internet	23.30%	24%
Sample_09	870922	ChatGPT+Internet	20.60%	18%
Sample_10	870923	ChatGPT+Internet	21.10%	21%

CONCLUSION

Recent AI language models have taken a significant step forward in producing remarkably controllable, fluent, and grammatical text based on predefined algorithms set. We designed and implemented AI versus Human text detection using the features (implemented algorithm), i.e., writing style, coherence, consistency, and argument logistics. Finally, we developed a tool to investigate the segregation between AI-generated text and human-written text by AI-generated text detection models. We are providing best accuracy in three levels called direct text using ChatGPT, internet and mixed content, those we discussed above, finally we have concluded 95 – 99 % of AI accuracy and less than 1 % of false detection. We have developed and integrated AI detection in 'DrillBit Plagiarism' checker. DrillBit plagiarism is a global checker that uses the most advanced technology to catch the most sophisticated forms of plagiarism.

REFERENCE

- Agrawal, A., Gans, J., & Goldfarb, A. (2017, July 26). How AI will change the way we make decisions. Harvard Business Review. <https://hbr.org/2017/07/how-ai-will-change-theway-we-make-decisions>
- Ball, M. (2022). The metaverse and how it will revolutionize everything. W.W. Norton & Company, Inc.
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1).
- Benoit, K., Watanabe, K., Wang, H., Lua, J. W., Kuha, J., & European Research Council. (2021). quanteda. text stats: Textual statistics for the quantitative analysis of textual data(0.95).
- Nolan B. This man used AI to write and illustrate a children's book in one weekend. he wasn't prepared for the backlash. [internet]. business insider. 2023.
- DePeau-Wilson M. AI passes U.S. Medical Licensing Exam [Internet]. Medical News. Medpage Today. 2023.
- Stokel-Walker C. ChatGPT listed as author on research papers: many scientists disapprove. Nature 2023;613:620–1. 10.1038/d41586-023-00107-z
- Getahun H. After an AI bot wrote a scientific paper on itself, the researcher behind the experiment says she hopes she didn't open a "pandora's box" [internet]. insider. insider. 2022.
- Design education forum of SA [Internet]. What is an Academic Paper? Design Education Forum of SA. 2023.
- Christian J. CNET's AI journalist appears to have committed extensive plagiarism [Internet]. Futurism. 2023.
- Piltch A. I asked ChatGPT how to build a PC. It told me to break my CPU. [Internet]. Tom's Hardware. 2023.
- Else H. Abstracts written by chatgpt fool scientists. Nature 2023;613:423. 10.1038/d41586-023-00056-7

13. Campagnola C. Perplexity in language models [Internet]. Medium. Towards Data Science. 2022.
14. Memon AR. Similarity and Plagiarism in scholarly Journal submissions: bringing clarity to the concept for authors, reviewers and editors. *J Korean Med Sci* 2020;35:e217. 10.3346/jkms.2020.35.e217
15. National Academies of Sciences Engineering, Policy and Global Affairs, Committee on Science, Engineering, Medicine, and Public Policy, Committee on Responsible Science Incidence and consequences Fostering Integrity in Research, National Academies Press (2017)
16. O. Zimba, A.Y. Gasparyan Plagiarism detection and prevention: a primer for researchers *Reumatologia*, 59 (2021), pp. 132-137, 10.5114/reum.2021.105974
17. E. Brynjolfsson, D. Rock, C. Syverson Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics National Bureau of Economic Research (2017), Article w24001, 10.3386/w24001