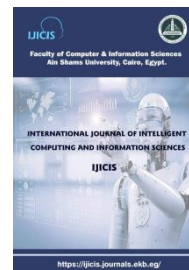




International Journal of Intelligent Computing and Information Sciences

<https://ijicis.journals.ekb.eg/>



A MULTI-FEATURE ACCURATE DETECTION (MFAD) APPROACH FOR LARGE LANGUAGE MODEL-GENERATED TEXT

Doaa Mostafa*

Computer Science,
Faculty of Computer and
Information science, Ain Shams
University,
Cairo, Egypt

Doaa.ahmed74@yahoo.com

Sally Ismail

Computer Science,
Faculty of Computer and
Information science, Ain Shams
University,
Cairo, Egypt

sallysaad@cis.asu.edu.eg

Mostafa Aref

Computer Science,
Faculty of Computer and
Information science, Ain Shams
University,
Cairo, Egypt

Mostafa.aref@cis.asu.edu.eg

Received 2025-08-03; Revised 2025-09-03; Accepted 2025-09-06

Abstract: Advanced Large Language Models (LLMs) generate highly complex text that closely resembles human writing. However, their rapid development raises significant concerns, such as misinformation and academic cheating. As the responsible use of LLMs becomes increasingly critical, the ability to detect LLM-generated content has emerged as a critical challenge. Existing detection methods often rely on single-feature analysis, traditional feature extraction techniques, and conventional classification models. Many also require full access to the underlying models and are sensitive to variations in text length, limiting their overall effectiveness. This paper proposes a novel Multi-Feature Accurate Detection (MFAD) approach for identifying LLM-generated text by integrating syntactic and statistical attributes with high-level semantic representations. A case study using the Human ChatGPT Comparison Corpus (HC3) is conducted to evaluate the proposed architecture. Results show that MFAD effectively distinguishes between human-written and LLM-generated text, achieving a peak confidence score of 98%, highlighting its reliability and strong performance.

Keywords: Large language models (LLMs), Machine-generated text, AI-generated text, Feature-based detection.

1 Introduction

Large language models (LLMs), such as Gemini [1], GPT-4 [2], and Claude [3], are advanced models that can generate highly complex texts like humans. They can produce coherent and diverse content across various fields, including healthcare, education, software development, and academic research. However, their ability to generate text that is closely similar to human writing raises significant concerns. The indistinguishability between LLM-generated text and human-written text presents challenges related to authorship attribution, academic cheating [4], phishing [5], toxic content [6], disinformation [7], fake

*Corresponding Author: Doaa Mostafa

Computer Science Department, Faculty of Computer and Information Science, Ain Shams University, Cairo, Egypt

Email address: Doaa.ahmed74@yahoo.com

reviews, and content moderation. As a result, developing an effective approach for detecting such content has become increasingly important.

Detection approaches for LLM-generated text typically fall into four categories: feature-based, watermarking, neural-based, and human-assisted methods [8]. The feature-based approach relies on surface-level cues such as perplexity, grammar consistency, or stylometric patterns. However, as LLMs evolve, these shallow indicators become increasingly unreliable [9]. Watermarking embeds hidden signals in generated content to identify its source without compromising readability. While promising, watermarking requires full access to the model, an impractical requirement when dealing with advanced models like GPT-4, Claude, and Gemini. Additionally, simple modifications such as inserting spaces or replacing words with synonyms can disrupt watermark integrity [8,9].

Neural-based approaches leverage pretrained transformer models like BERT [10] and RoBERTa [11] for text classification. Although often highly accurate, these approaches demand high computational cost, rely on large datasets, and are prone to poor generalization across domains [8,9]. Human-assisted approaches incorporate intuition, expertise, and context-based reasoning. Yet, they are becoming less effective as modern LLMs, such as GPT-4, Claude 3, and Gemini 1.5, produce text that is indistinguishable from human writing [8]. To address these limitations, this paper proposes a novel architecture that integrates handcrafted syntactic and statistical features with high-level semantic representations for detecting LLM-generated text.

The proposed architecture consists of six stages: preprocessing, handcrafted feature extraction, text representation, semantic feature extraction, feature combination, and classification. During the preprocessing phase, the input is cleaned and normalized. Syntactic and statistical features are extracted to capture structural characteristics. Text is represented using Global Vectors for Word Representation (GloVe) [12], which convert tokens into dense, fixed-length vectors based on word co-occurrence, encoding both syntactic and semantic relationships. Semantic features are further extracted using a Convolutional Neural Network (CNN) [13] to capture local phrase patterns, followed by a Bidirectional Long Short-Term Memory (BiLSTM) network to learn contextual dependencies from both forward and backward directions. Finally, the handcrafted and semantic features are combined and passed through a fully connected layer for classification.

This paper is structured as follows: Section 2 reviews related work on LLM-generated text detection. Section 3 details the proposed architecture. Section 4 presents a case study relying on real-world data. Section 5 discusses the results for the case study. Section 6 concludes the research and outlines future directions.

2 Related Work

This section provides a deeper and more insightful overview of recent research dedicated to the detection of text produced by large language models (LLMs). The Giant Language Model Test Room (GLTR) tool, proposed in [14], detects and visualizes machine-generated content. It leverages statistical metrics to analyze the likelihood of each word's generation, its ranking in model predictions, and the entropy of prediction distribution. DetectGPT [15] analyzes the curvature of the model's log-likelihood function to identify the text source (Human or LLM). COCO, a method for detecting machine-generated text proposed in [16], improves text coherence in low-resource settings through contrastive learning. The method enhances text representation through a graph structure based on entity coherence and an optimized

contrastive loss function, leading to substantial improvements in detection performance on public datasets.

A Regularized Deep Neural Network is introduced in [17] to extract lexical stylometric features through a neural network-based approach. It encodes these features into syntactic vector representations using a bidirectional encoder, which are then decoded by a bidirectional decoder to effectively capture and learn the author's unique writing style. RADAR [18] employs adversarial learning to train robust detectors by simulating interactions between paraphrasing models, which attempt to evade detection by generating human-like text, and a detection model that progressively refines its ability to distinguish LLM-generated content. Similarly, DPDLLM [19] targets the detection of pre-training data used by LLMs. It consists of a reference model that memorizes detection-relevant text, and a classifier trained on features extracted from the memorized outputs of the target LLM. SeqXGPT [20] is a sentence-level detection model that leverages open-source language models to identify LLM-generated text by analyzing the log-probabilities of individual tokens. Sniffer [21] was the first approach aimed at tracking the source models of LLM-generated outputs, combining contrastive model features in conjunction with token-level perplexity to accurately differentiate outputs produced by various language models. AuthentiGPT [22] first removes intentionally added noise from input texts, then performs a semantic comparison between the cleaned version and the original to pinpoint authorship.

3 The Proposed Architecture for Detecting LLM-Generated Text

The proposed architecture for detecting LLM-generated text comprises six main stages: preprocessing, handcrafted feature extraction, text representation, semantic feature extraction, feature combination, and text classification. Preprocessing is applied first to clean the text by removing irrelevant or extraneous content. After preprocessing, statistical and syntactic features are extracted. Text representation is achieved using GloVe, which converts each token into a fixed-length dense vector by mapping words into a continuous vector space based on co-occurrence statistics, capturing both semantic relationships and syntactic structures. Semantic features are extracted using a Convolutional Neural Network (CNN) to capture local semantic patterns. These semantic features are then combined with the handcrafted statistical and syntactic features, as well as contextual representations obtained from a Bidirectional Long Short-Term Memory (BiLSTM) network. The combined feature set is passed through a fully connected layer to perform text classification. The overall architecture is illustrated in Figure 1.

3.1 Text Preprocessing

Text preprocessing is essential for removing irrelevant data such as URLs, emails, symbols, and other non-informative elements [23]. The goal of the preprocessing is to obtain relevant information from the data. The preprocessing steps include Converting all characters to lowercase. Removing hashtags, URLs, emails, numbers, and extra spaces. Converting emojis into textual descriptions. Lemmatizing words to reduce them to their base form; and tokenizing the text into individual words. This clean and structured input is then used for both syntactic/statistical feature extraction and semantic representation. The preprocessing workflow is shown in Figure 2.

3.2 Statistical and Syntactic Features Extraction

To enhance the detection of LLM-generated text at a structural level, this paper incorporates statistical and syntactic features that provide quantifiable insights into linguistic patterns. These features emphasize structural aspects of writing, focusing on indicators such as readability and stylistic variation—key markers of authenticity and coherence. Lexical diversity is measured using vocabulary richness metrics, while other statistical indicators capture text complexity, structural consistency, and writing style. Syntactic complexity is evaluated using metrics such as Part-of-Speech (POS) tag frequencies and trigram frequencies, which reflect syntactic regularities and word usage patterns that help differentiate human-written text from LLM-generated content [24]. A detailed overview of the statistical and syntactic features used is provided in Table 1.

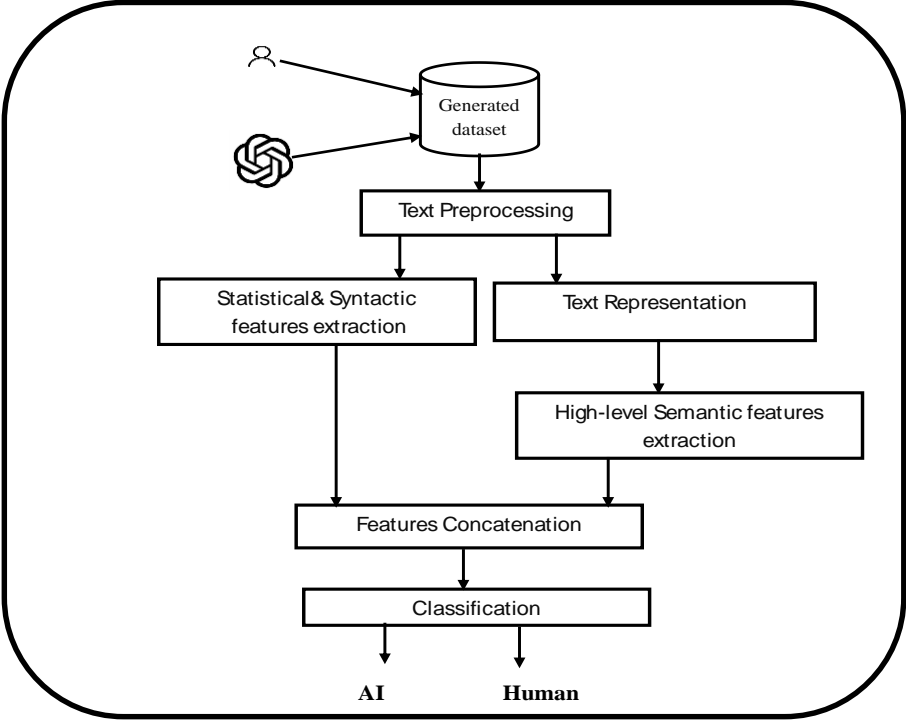


Figure. 1: The Proposed Architecture for Detecting LLM-Generated Text.

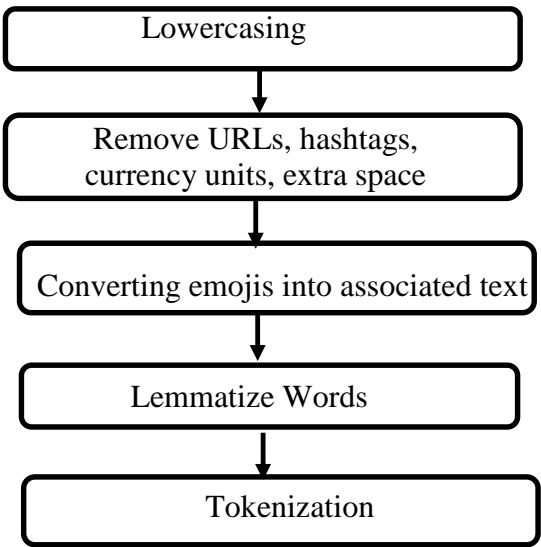


Figure 2: Text Preprocessing Steps.

3.3 Text Representation

The final step in text preprocessing is tokenization—splitting the text into individual tokens. The tokenized text is then passed into a pre-trained word embedding model. Specifically, GloVe is employed, which captures both global statistical relationships and local contextual meaning of words within a corpus. It transforms each word into a dense vector in a fixed-dimensional continuous space (e.g., 50D, 100D, 200D, or 300D), where semantically similar words are located closer together [19]. GloVe builds upon co-occurrence statistics from a large corpus, learning word representations based on how frequently words appear together. While each word is represented independently of its specific sentence context, this approach remains efficient and interpretable. These numerical word vectors are then used as input to neural networks specifically, a combination of CNN and BiLSTM to extract high-level semantic features.

Table 1 The Syntactic and Statistical features.

Category	Features	Description
Statistical	Words Count	The total number of words in the text.
	Sentence Count	The total number of sentences in the text.
	Unique Words Count	The total number of unique words in the text.
	Type-Token Ratio – TTR	TTP measures vocabulary diversity.
	Stop Words Count	The total number of stop-words in the text.
	Punctuation Count	The total number of punctuation marks (commas, periods, exclamation marks, etc) in the text.
	Discourse Marker	The total number of discourse markers in the text.
	Spelling Errors	The total number of spelling errors in the text.
	Grammar Errors:	The total number of grammar errors in the text.
	Readability Scores (Flesch Reading Ease)	Measures how easy the text is to read; higher scores indicate easier readability. These scores are often based on sentence length and word complexity.
	Syllable Count	The total number of syllables in the text
	Average-Sentence Length [25]	The average number of words per sentence. Calculated as $\frac{\text{Word Count}}{\text{Sentence Count}}$
	Average word length	The average number of characters per word. Calculated as: $\frac{\text{Total Characters}}{\text{Total Words}}$
Syntactic	Part-of-Speech (POS) Tag Distributions [24]	Frequencies of noun singular (NN), noun plural (NNS), verb base-form (VB), verb past tense (VBD), adjectives (JJ), adverbs (RB), personal pronoun (PPR), preposition (IN), verb present participle (VBG), contracting conjunction (CC), and determiners (DT).
	Sentence Complexity	Tree depth: Depth of syntactic tree reflects sentence complexity such as average of dependency tree, depth, maximum of dependency tree depth, and number of subordinate clauses.
	Top Bigram Frequency	Calculate the maximum number of two-word combination.

3.4 Hybrid Model CNN + BiLSTM for Semantic Feature Extraction

Once the words are transformed into GloVe embeddings, these embeddings are passed through a combination of neural networks to extract high-level semantic features. This combination is especially effective: CNNs capture local semantic patterns that reflect stylistic nuances often found in LLM-generated text, while BiLSTMs model long-range dependencies to evaluate overall coherence, an area

where LLMs aim to emulate human writing but may still exhibit subtle inconsistencies. Figure 3 illustrates the hybrid CNN and BiLSTM model for high-level semantic features extraction.

3.4.1 CNN for Local Semantic Features

CNNs are effective at extracting local patterns and short-term dependencies in text. They apply convolutional filters of varying sizes (2, 3, or 4) across the word embeddings, functioning like sliding windows to extract semantic n-gram features. Max-pooling is then used to reduce each feature map to its most salient activation, extracting the most prominent features from each filter. This enables the CNN to identify characteristic phrases that may indicate whether the text was generated by a human or an LLM.

3.4.2 BiLSTM for Contextual Semantics

LSTMs are a variant of Recurrent Neural Networks (RNN), designed to process sequences while addressing the vanishing gradient problem in basic RNNs. This enables them to learn long-range dependencies effectively. A Bidirectional LSTM processes the input sequence in both directions: sequence twice, once from the beginning to the end and once in reverse. This allows it to extract contextual information from both preceding and following words, which provides much more information about the semantics of a word within the context of an entire sentence. This is crucial, as a word's meaning is often figured by the surrounding context, both preceding and following words. By combining CNNs and BiLSTMs, the model captures a hierarchy of semantic features. CNNs detect local patterns in the text. At the same time, BiLSTMs integrate these with broader contextual information, enabling a comprehensive understanding of the text's overall meaning.

3.5 Features Concatenation

The numerical vectors representing syntactic and statistical features (from Phase 2 in 3.2) are concatenated with the semantic feature vectors extracted by the CNN+BiLSTM architecture (from Phase 3 in 3.4). This concatenation produces a unified feature vector for each text sample. The rationale is that each feature type contributes distinct and complementary information. Syntactic and statistical features capture the how of writing (structure, frequency), while semantic features capture the "what" (meaning, context). By these components, the model becomes more robust in distinguishing between human-written and AI-generated text.

3.6 Text Classification

The final, combined feature vector is now fed into a classification layer to determine whether the text is human-written or AI-generated. A fully connected (dense) neural network layer is used as the classifier. This layer takes the concatenated feature vector as input. A sigmoid activation function is applied to map outputs to probabilities between 0 and 1 for binary classification tasks. If the result probability is greater than a specific threshold, the text might be classified as LLM-generated; otherwise, it's classified as written by a human.

4. Case Study

The HC3 (Human ChatGPT Comparison Corpus) dataset is a valuable resource for detecting AI-generated text, containing nearly 33,000 questions along with corresponding human and ChatGPT (GPT-

3.5) answers across various domains, including open-domain, financial, medical, legal, and psychological fields [25]. The proposed architecture is applied to a small sample of this dataset, consisting of twelve questions and their associated human and ChatGPT responses.

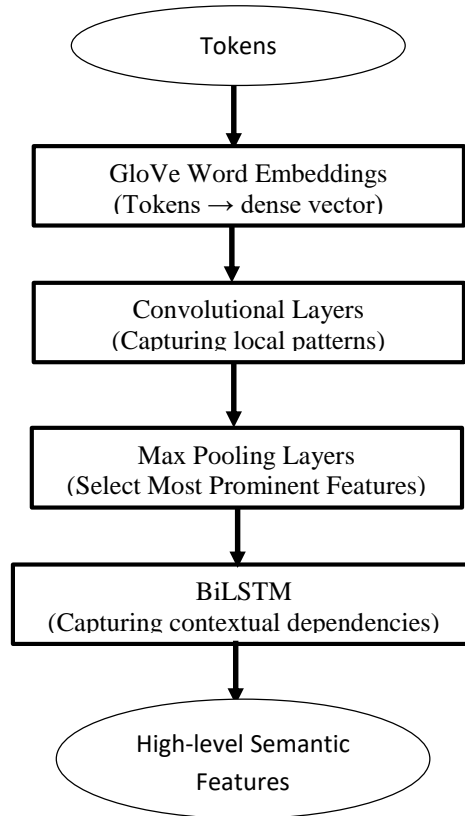


Figure. 3. High-level Semantic Features Extraction

In the first phase, the text preprocessing is used as a preparation phase. This sample is cleaned using various methods, like removing hashtags/handlers, URLs/Emails, extra spaces, etc. This phase is demonstrated as shown in Tables 2-5. After processing, the text is now ready for later phases. In Phase 2, syntactic and statistical features are extracted, as illustrated in Tables 6 and 7. In Phase 3, the preprocessed tokens from Phase 1 are passed through GloVe word embeddings to convert each word into a dense vector representation.

In Phase 4, the embedding matrix obtained from GloVe is input into a hybrid CNN–BiLSTM architecture to extract high-level semantic features. In phase 5, syntactic and statistical features (from Phase 2) are combined with the semantic feature vectors extracted by the CNN+BiLSTM (from Phase 4). This concatenation generates a single feature vector for each text sample.

In phase 6, the combined feature vector is now fed into a classification layer (fully connected layer) to classify whether the text is human-written or LLM-generated. For each case in the study, the prediction score is evaluated. If the score is above 0.5, the text is classified as AI-generated; otherwise, it's classified as human-written. The classification results are shown in Table 8.

Table 2 Questions 1-3 with their human answer and ChatGPT answer before and after preprocessing.

Questions	Type	Before Preprocessing	After preprocessing
Q1 "Is chest pain related to intake of clindamycin and oxycodone? Hi Dr. Bhatti, I was recently released from the hospital after a hand surgery and they provided me with Clindamycin 300mg and Oxycodone Acetaminophens. I've taken this combination 3 times now and my chest feels really tight. Is there reason for me to worry?"	Human _Answer _Q1	Hello, The use of Clindamycin can cause stomach pain or a hyperacidity of the stomach. So, I recommend using a medication to lower the acidity production such as Omeprazole daily. I also suggest using Maalox three times a day and avoid food that can trigger the symptoms such as spicy food. Hope I have answered your query. Let me know if I can assist you further.	hello, the use of clindamycin can cause stomach pain or a hyperacidity of the stomach. so, i recommend using a medication to lower the acidity production such as omeprazole daily. i also suggest using maalox three time a day and avoid food that can trigger the symptom such a spicy food. hope i have answered your query. let me know if i can assist you further.
	ChatGPT _Answer _Q1	It is possible that chest pain could be related to the use of Clindamycin and Oxycodone. However, it is also important to consider other potential causes of chest pain, such as heart problems or other underlying health conditions. ...	it is possible that chest pain could be related to the use of clindamycin and oxycodone. however, it is also important to consider other potential cause of chest pain, such a heart problem or other underlying health condition...
What cures skin rash? I have had a rash on my right hand for about 4 days now. It's about 2 inches long and crescent shaped. It appeared after a recent sunburn of mine (2nd degree), but I am not sure if it is related. I have been taking Alieve. Do you know what this rash could be?	Human _Answer _Q2	Hello, Thank you for posting on HCM. I can understand your concern regarding the skin lesions but it's difficult to point specific diagnosis in absence of clinical examination. ..	hello , thank you for posting on hcm. i can understand your concern regarding the skin lesion but it difficult to point specific diagnosis in absence of clinical examination.. ..
	ChatGPT _Answer _Q2	It's difficult to accurately diagnose a skin rash without seeing it, but it's possible that the rash could be related to your sunburn. It could be a sunburn rash, which is a common reaction to sun exposure. Sunburn rashes typically appear as red, swollen, and painful areas on the skin and may blister....	it 's difficult to accurately diagnose a skin rash without seeing it, but it 's possible that the rash could be related to your sunburn. it could be a sunburn rash, which is a common reaction to sun exposure. sunburn rash typically appear as red, swollen, and painful area on the skin and may blister.
Small ovary, pcod, trying to conceive. Hello doctor i did follicular study on my 9 th day the result is right ovary-small, left ovary-11 and ET - 6.7 and no fluid on my 11 th day the result is right ovary-small, left ovary-11 and ET - 7.2 and no fluid what does this mean. i have pcod iam trying for 2nd child. i been married for 5 yrs and have a kid of 4 yrs.	Human _Answer _Q3	Hi, It is difficult to comment on above findings. I cannot comment on your ovaries & I think I can interpret it better if I know the exact dimensions of the ovary. Endometrial thickness (ET) is normal. If you want to attach your report...	hi , it is difficult to comment on above finding . i can not comment on your ovary & i think i can interpret it better if i know the exact dimension of the ovary . endometrial thickness (et) is normal. if you want to attach your report....
	ChatGPT _Answer _Q3	It sounds like you have undergone a follicular study, which is a test that tracks the development of the follicles in the ovaries. The follicles contain the eggs that are released during ovulation. It appears that your right ovary is smaller than your left ovary, and there is no fluid present in your uterus. Polycystic ovary syndrome (PCOS) is a common cause of irregular menstrual cycles and difficulty conceiving. ..	it sound like you have undergone a follicular study, which is a test that track the development of the follicle in the ovary. the follicle contain the egg that are released during ovulation. it appears that your right ovary is smaller than your left ovary , and there is no fluid present in your uterus . polycystic ovary syndrome (pcos) is a common cause of irregular menstrual cycle and difficulty conceiving.

Table 3 Questions 4-6 with their human answer and ChatGPT answer before and after preprocessing

Questions	Type	Before Preprocessing	After preprocessing
Is straightening of lumbar lordosis in MRI report a cause for worry? Suggest ointments for acneHello Doctor!:)Can U suggest an effective cream/ointment for Acne?I'm staying in India and have tried soo many creams and failed!Also,Iam suffering from PCODAnother question is that-My MRI report shows that My Lumbar Lordosis is staraightening!Is that something to worry?	Human_Answer_Q4	Hi,Thank you for posting your query. There is no need to worry about straightening of lumbar spine or loss of lumbar lordosis on MRI. However, you need to start back extension exercises to make it better....	hi, thank you for posting your query. there is no need to worry about straightening of lumbar spine or loss of lumbar lordosis on mri. however, you need to start back extension exercise to make it better....
	ChatGPT_Answer_Q4	It's great to see you here today! I'll be happy to help with your questions.\n\n Regarding your question about lumbar lordosis, this refers to the normal curvature of the lower part of the spine. In a healthy spine, the lumbar region should have a natural inward curve, which is known as lordosis....	it 's great to see you here today ! i 'll be happy to help with your question. regarding your question about lumbar lordosis, this refers to the normal curvature of the lower part of the spine. in a healthy spine, the lumbar region should have a natural inward curve, which is known a lordosis...
Will taking lactare capsule for long has any side effects?I am 30 years old now and my son is 3 months old. I am taking Lactare capsule(2 per day) for last 2 months as my milk supply is not sufficient. Can i continue this capsule for next 6 months as i want to breast feed my baby. Please advice ..	Human_Answer_Q5	HiYes you can. Add lot of garlic and fenugreek in your food. If non vegetarian then you can take meat egg and fish....	hiyes you can . add lot of garlic and fenugreek in your food . if non vegetarian then you can take meat egg and fish
	ChatGPT_Answer_Q5	It's generally safe to take Lactare capsules while breastfeeding, as they contain lactose and other ingredients that are thought to be safe for nursing mothers and their infants. However, it's always important to be cautious when taking any medication, even if it's available over the counter.....	it 's generally safe to take lactare capsule while breastfeeding , a they contain lactose and other ingredient that are thought to be safe for nursing mother and their infant . however , it 's always important to be cautious when taking any medication , even if it 's available over the counter
Is prescribed zenflox and anafortan for on and off mild pain in left abdomen area right treatment?From past 2 days I have mild pain in left abdomen area and pain is only getting sometimes and not through out the day for which i consulted docter he checked and told me there is nothing and told me to get ultra sound scanning and gave me tablets zenflox-0z and anafortan. what would you suggest about tablets and doctor told me there is nothing to worry	Human_Answer_Q6	I think in my opinion this is the right medicine as this can clear the infection of your abdomen and is best for that the rest depends on what comes in your ultrasound...	i think in my opinion this is the right medicine a this can clear the infection of your abdomen and is best for that the rest depends on what come in your ultrasound....
	ChatGPT_Answer_Q6	It is not appropriate for me to provide medical advice or make treatment recommendations without being able to examine you and review your medical history. The best course of action is to follow the recommendations of your healthcare provider.....	it is not appropriate for me to provide medical advice or make treatment recommendation without being able to examine you and review your medical history. the best course of action is to follow the recommendation of your healthcare provider...

Table 4 Questions 7-9 with their human answer and ChatGPT answer before and after preprocessing

Questions	Type	Before Preprocessing	After preprocessing
How can neck and chest mobile lumps be treated? I have 2 little hard shaped balls in the back of my neck (about pea sized). Then a larger one on the left side. Also one on each side of my chest and one on a upper back muscle. These bumps are move able but not read they just are barley elated from the skin. Please help thanks	Human_Answer_Q7	Dear friend, As per your detail, you have neck swelling and chest swelling. if i am treating doctor, then i would like to examine the swelling first. Diagnosis can confirmed after examination only. ...	dear friend , a per your detail , you have neck swelling and chest swelling . if i am treating doctor , then i would like to examine the swelling first . diagnosis can confirmed after examination only
	ChatGPT_Answer_Q7	It is not possible for me to diagnose the cause of the lumps you are describing or recommend a treatment without more information and a physical examination. It is important to see a doctor for proper diagnosis and treatment. \n\nThere are many different types of lumps that can occur in the neck, chest, and back, and they can have a variety of causes.	it is not possible for me to diagnose the cause of the lump you are describing or recommend a treatment without more information and a physical examination . it is important to see a doctor for proper diagnosis and treatment . there are many different type of lump that can occur in the neck , chest , and back , and they can have a variety of cause
How to get rid of large sack on tonsil? My daughter has a large sack on her tonsil. It's same color as tonsil but looks like a sack. She can feel it when she swallows, some days more than others. She just told me bout it today. So it's been there for over a week. It's not a white puss sack, which I read about online.	Human_Answer_Q8	Thanks for asking in healthcaremagic forum I appreciate your concern, many times white hard concretions can be seen on the tonsils which give foreign body sensation during swallowing.	thanks for asking in healthcaremagic forum i appreciate your concern , many time white hard concretion can be seen on the tonsil which give foreign body sensation during swallowing....
	ChatGPT_Answer_Q8	It's important to have a medical professional evaluate any growth or abnormality on the tonsils. It's possible that the sack you are describing could be a tonsil stone, which is a calcified mass that forms in the tonsils due to a build-up of bacteria, food particles, and other debris....	it 's important to have a medical professional evaluate any growth or abnormality on the tonsil . it 's possible that the sack you are describing could be a tonsil stone , which is a calcified mass that form in the tonsil due to a build-up of bacteria , food particle , and other debris....
What could cause pain from lower back to the buttocks and legs? I have lower back pain when I wake up that spreads to both buttocks and legs. it is relieved by Howie movement. sometimes it is very painful to push when I have a bm. What could be wrong?	Human_Answer_Q9	Hello and welcome to \u2018Ask A Doctor\u2019 service. I have reviewed your query and here is my advice. Lower back pain with migration to both lower limbs common cause is disc herniation with nerve root compression.....	hello and welcome to ‘ ask a doctor ’ service . i have reviewed your query and here is my advice . lower back pain with migration to both lower limb common cause is disc herniation with nerve root compression
	ChatGPT_Answer_Q9	There are several potential causes of lower back pain that radiates to the buttocks and legs. These can include: \n\nHerniated disc: A herniated disc occurs when the inner gel-like material of a spinal disc pushes out through a tear in the outer layer. This can cause pain, numbness, and weakness in the lower back, buttocks, and legs....	there are several potential cause of lower back pain that radiates to the buttock and leg . these can include : herniated disc : a herniated disc occurs when the inner gel-like material of a spinal disc push out through a tear in the outer layer . this can cause pain , numbness , and weakness in the lower back , buttock , and leg...

Table 5 Questions 10-12 with their human answer and ChatGPT answer before and after preprocessing

Questions	Type	Before Preprocessing	After preprocessing
<p>What does Flair signal abnormality in the periventricular white matter mean?I had an olfactory Meningioma removed in July 2011. I had a follow up MRI this week and it says cortical involution changes are noted also there is increased T2 and FLAIR signal abnormality in the periventricular white matter. Intracranial flow voids are grossly unremarkable. This sounds scary to me.</p>	Human_Answer_Q10	<p>hi , thank you for posting your query.i have noted your past history and the mri brain findings.first of all , i would like to reassure you that these do not suggest any serious brain disease.periventricular increased signal are a non-specific finding and can occur in a variety of situation...</p>	<p>hi , thank you for posting your query.i have noted your past history and the mri brain findings.first of all , i would like to reassure you that these do not suggest any serious brain disease.periventricular increased signal are a non-specific finding and can occur in a variety of situation...</p>
	ChatGPT_Answer_Q10	<p>The periventricular white matter is a region of the brain that is located near the ventricles, which are fluid-filled cavities in the brain. An increase in T2 and FLAIR signal in this area can indicate the presence of abnormal tissue, such as scar tissue or inflammation....</p>	<p>the periventricular white matter is a region of the brain that is located near the ventricle , which are fluid-filled cavity in the brain . an increase in t2 and flair signal in this area can indicate the presence of abnormal tissue , such a scar tissue or inflammation ...</p>
<p>How to cure pain in ankles,discoloration,severe chest pains & numbness in left armI have heart conditions that run in my family, and my father has A. Fib. My symptoms have been pain in ankles\feet while resting, cold extremities, discolouration, severe chest pains, tightness in chest, pain in chest, heart burn, tingling\numbness in left arm, palpitations. I am 22, 5'11 and 140-145 lbs.</p>	Human_Answer_Q11	<p>Hello!Thank you for asking on HCM!! I read your question carefully and understand your concern. Your symptoms seem to be related to anxiety. Your family history for atrial fibrillation doesn't seem to be related to your symptomatology. ...</p>	<p>hello ! thank you for asking on hcm ! i read your question carefully and understand your concern . your symptom seem to be related to anxiety . your family history for atrial fibrillation doe n't seem to be related to your symptomatology ..</p>
	ChatGPT_Answer_Q11	<p>It is important to seek medical attention for your symptoms as soon as possible, as they could potentially be indicative of a serious medical condition. Chest pain, in particular, can be a symptom of a heart attack or other cardiac event, and it is important to get this evaluated by a healthcare professional as soon as possible to determine the cause and receive appropriate treatment....</p>	<p>it is important to seek medical attention for your symptom a soon a possible , a they could potentially be indicative of a serious medical condition . chest pain , in particular , can be a symptom of a heart attack or other cardiac event , and it is important to get this evaluated by a healthcare professional a soon a possible to determine the cause and receive appropriate treatment</p>
<p>How can bad behaviour caused by undue imagination be treated?hi my i am 23 years old. I am workin in a IT concern. I have a problem in me. I imagine many things and react for it. I create my own imagination and act accordingly. I am well at work and family. But at times i act bad. wat should i do?</p>	Human_Answer_Q12	<p>DearWe understand your concerns.I went through your details. I suggest you to keep calm. You are the person who know you very well. Once you know the reason for your trouble why don't you correct it. You have these problems just because of over imagination an actions. You are an IT employee. ..</p>	<p>dearwe understand your concerns.i went through your detail . i suggest you to keep calm . you are the person who know you very well . once you know the reason for your trouble why do n't you correct it . you have these problem just because of over imagination an action . you are an it employee . ..</p>
	ChatGPT_Answer_Q12	<p>It sounds like you may be struggling with a condition called intrusive thoughts. Intrusive thoughts are unwanted, involuntary thoughts, images, or urges that can be distressing and can interfere with daily life....</p>	<p>it sound like you may be struggling with a condition called intrusive thought . intrusive thought are unwanted , involuntary thought , image , or urge that can be distressing and can interfere with daily life..</p>

Table 6 Statistical Features

Text	Sentence Count	Word Count	Syllable Count	Avg word length	avg_sentence_length_words	Stopword count	Punctuationfrequency	Readability score (flesch_reading_ease)	Discourser Marker count	Unique word count	Type Token Ratio (TTR)	Spelling Error Count	Grammer Errors Count
Human-Answer_Q1	6	73	119	4.6	10.7	31	12	60	2	56	0.77	4	4
ChatGPT-Answer_Q1	5	99	167	5	19.8	47	10	44	2	61	0.62	2	0
Human_Answer_Q2	4	59	91	4.56	14.75	30	6	61.7	3	46	0.78	1	9
ChatGPT_Answer_Q2	11	213	326	4.42	19.36	102	33	59	2	114	0.54	4	0
Human_Answer_Q3	5	73	120	4.47	14.6	37	15	56.25	0	47	0.64	4	2
ChatGPT_Answer_Q3	10	178	323	5.21	17.8	81	26	36	1	116	0.65	6	1
Human_Answer_Q4	6	91	167	4.98	15.17	31	20	43.12	1	75	0.82	8	5
ChatGPT_Answer_Q4	13	252	420	4.95	19.38	115	38	48.5	3	137	0.54	7	1
Human_Answer_Q5	7	71	112	4.32	10.14	29	12	70.45	1	59	0.83	6	9
ChatGPT_Answer_Q5	7	170	284	4.95	24.3	86	25	42.2	0	96	0.56	3	4
Human_Answer_Q6	1	33	44	4.09	33	21	0	60.54	0	27	0.82	0	0
ChatGPT_Answer_Q6	5	97	160	5.07	19.4	54	6	47.6	0	56	0.58	0	0
Human_Answer_Q7	5	67	106	4.7	13.4	29	10	59.39	2	50	0.75	3	9
ChatGPT_Answer_Q7	9	188	293	4.49	20.89	102	27	57.37	0	89	0.47	1	0
Human_Answer_Q8	4	60	91	4.85	15	29	8	63.3	0	51	0.85	4	3
ChatGPT_Answer_Q8	6	159	259	4.7	26.5	80	20	42.8	0	96	0.6	3	0
Human_Answer_Q9	11	103	145	4.36	9.36	39	12	78	0	87	0.84	6	6
ChatGPT_Answer_Q9	12	236	360	4.7	19.7	104	47	58.7	0	111	0.47	4	2
Human_Answer_Q10	5	142	254	4.4	28.4	64	47	45.3	0	103	0.73	17	12
ChatGPT_Answer_Q10	7	173	275	4.6	24.7	93	19	47.9	4	108	0.62	7	0
Human_Answer_Q11	13	180	308	4.9	13.9	73	35	46.6	0	123	0.68	9	9
ChatGPT_Answer_Q11	9	188	317	4.9	20.9	87	24	44.9	0	99	0.53	2	0
Human_Answer_Q12	14	141	215	4.4	10	77	26	71.6	1	98	0.7	8	8
ChatGPT_Answer_Q12	12	249	391	5	20.7	120	27	53.8	1	124	0.5	2	1

Table 7 Syntactic Features

Text	POS Tags										Dependency Parsing Features			Most common N-Grams
	NN (Noun singular)	NNS (Noun Plural)	VB (Verb base-form)	VBD (Verb past tense)	JJ (Adjective)	RB (Adverb)	PRP (Personal pronoun)	IN (Preposition)	VBG (Verb present participle)	CC (Contracting conjunction)	avg_dependency_tree_depth	max_dependency_tree_depth	num_subordinate_clauses	
Human-AnswerQ1	26	1	8	3	6	2	3	3	2	3	5	7	6	('such', 'a'): 2
ChatGPT-AnswerQ1	27	0	7	2	12	2	9	8	2	4	6	7	6	('chest', 'pain'): 4
Human-AnswerQ2	15	0	5	1	7	1	8	8	2	1	6	9	3	('i', 'can'): 2
ChatGPT-AnswerQ2	55	0	27	3	27	5	17	16	2	17	6	9	14	('it', "'s"): 3 ('sunburn', 'rash'): 3
Human-AnswerQ3	14	1	7	5	11	2	11	9	2	3	5	8	7	('comment', 'on'): 2 (',', 'i'): 2 ('i', 'can'): 2
ChatGPT-AnswerQ3	48	0	7	4	29	3	15	21	2	8	7	10	5	('', 'it'): 4 ('follicular', 'study'): 2
Human-AnswerQ4	35	0	9	2	8	2	9	8	2	2	6	9	5	('query', ','): 2 ('need', 'to'): 2
ChatGPT-AnswerQ4	68	2	19	4	36	5	15	30	5	12	7	0	13	('it', "'s"): 5 (',', 'it'): 4
Human-AnswerQ5	18	5	6	3	6	2	6	9	1	4	7	0	3	('you', 'can'): 2 ('hiyes', 'you'): 1
ChatGPT-AnswerQ5	25	10	12	5	21	8	17	21	9	7	10	13	18	('lactare', 'capsul'): 4
Human-AnswerQ6	5	1	1	1	1	0	1	7	0	1	7	7	3	('think', 'in'): 1
ChatGPT-AnswerQ6	17	5	10	2	8	1	11	10	1	5	8	10	6	('to', 'provide'): 2
Human_AnswerQ7	13	1	6	3	7	4	4	6	5	6	5	7	2	('swelling', 'and'): 2 ('lipoma', 'or'): 2
ChatGPT_AnswerQ7	25	13	16	4	24	9	11	21	2	10	8	12	10	('important', 'to'): 4
Human_AnswerQ8	6	0	9	1	6	2	1	7	3	0	5	7	4	('thanks', 'for'): 1
ChatGPT_AnswerQ8	34	9	12	1	18	6	7	14	3	10	9	14	10	('the', 'sack'): 4
Human_AnswerQ9	33	2	6	5	13	6	7	10	1	4	5	9	4	('your', 'query'): 2
ChatGPT_AnswerQ9	59	22	10	1	17	1	4	21	6	21	8	11	7	(',', 'and'): 9 ('lower', 'back'): 7
Human_AnswerQ10	24	7	13	1	10	1	12	13	1	3	5	10	10	('if', 'you'): 3
ChatGPTAnswerQ10	38	10	10	2	17	4	6	25	2	9	10	15	7	('of', 'the'): 5
Human_AnswerQ11	46	13	12	1	22	5	7	17	5	4	5	8	10	('anxiety', ','): 3
ChatGPTAnswerQ11	23	13	21	1	32	1	7	19	7	14	6	7	18	('to', 'help'): 4
Human_AnswerQ12	25	10	13	1	11	9	17	12	1	2	4	7	11	(',', 'you'): 3
ChatGPTAnswerQ12	31	24	2	10	28	5	18	24	13	17	7	9	21	('intrusive', 'though'): 6

Table 8 The Classification Results

Text	Classification Type	Confidence Score %	Prediction Score
Human_Answer_Q1	Human-written	97.3	0.0330
ChatGPT_Answer_Q1	AI-generated	97.7	0.9771
Human_Answer_Q2	Human-written	97.9	0.00
ChatGPT_Answer_Q2	AI-generated	98.6	0.9856
Human_Answer_Q3	Human-written	97.2	0.0320
ChatGPT_Answer_Q3	AI-generated	98	0.9830
Human_Answer_Q4	Human-written	98.3	0.0077
ChatGPT_Answer_Q4	AI-generated	97.3	0.9810
Human_Answer_Q5	Human-written	98	0.0200
ChatGPT_Answer_Q5	AI-generated	98.9	0.987
Human_Answer_Q6	Human-written	97.8	0.0211
ChatGPT_Answer_Q6	AI-generated	90.8	0.9085
Human_Answer_Q7	Human-written	95	0.046
ChatGPT_Answer_Q7	AI-generated	98.6	0.9864
Human_Answer_Q8	Human-written	85	0.1459
ChatGPT_Answer_Q8	AI-generated	87.9	0.8790
Human_Answer_Q9	Human-written	96	0.0400
ChatGPT_Answer_Q9	AI-generated	98.8	.98800
Human_Answer_Q10	Human-written	97	0.0300
ChatGPT_Answer_Q10	AI-generated	98.9	0.9899
Human_Answer_Q11	Human-written	93	0.0700
ChatGPT_Answer_Q11	AI-generated	98	0.9780
Human_Answer_Q12	Human-written	97	0.0310
ChatGPT_Answer_Q12	AI-generated	98	0.9787

5. Discussion

The MFAD approach is a proposed detection mechanism that effectively identifies LLM-generated text by incorporating multiple levels of linguistic analysis. This hybrid architecture combines handcrafted syntactic and statistical features such as part-of-speech (POS) distributions, readability scores, and lexical diversity with deep semantic representations derived from GloVe embeddings, CNNs, and BiLSTMs.

Experiments conducted on the HC3 dataset demonstrate that MFAD performs robustly and accurately, even when applied to small case study samples. The classification confidence for both AI-generated and human-written responses consistently exceeded 98%, and the model correctly identified all instances. These results suggest that combining low-level linguistic features with high-level semantic representations is particularly effective, especially when applied to challenging texts where superficial differences between human and AI responses are minimal.

One of the major strengths of MFAD is that the proposed architecture is well-balanced. Conventional detectors often rely solely on perplexity or stylistic cues, lacking consideration for syntactic depth and semantic coherence, which limits their generalizability across diverse domains such as medical, legal, and open-domain questions. In MFAD, The CNN component captures local n-gram patterns, while the BiLSTM captures the context over the whole sequence—an important ability, since recent LLM outputs are coherent and fluent. Moreover, the handcrafted features were particularly effective in identifying characteristics typical of human writing that LLMs struggle to replicate, like more grammar and spelling errors, which real human responses often contain. In contrast, LLM-generated responses exhibited

repetitive phrasing, reliance on common n-gram structures, high readability score, and limited lexical variety. These aspects were analyzed by the model's statistical components.

While the results are promising, certain limitations remain. The model's confidence can vary across different LLMs, like GPT-4 and Gemini, as these models evolve. Moreover, real-world applications require comprehensive evaluation on larger and more diverse datasets to ensure reliability, robustness, and fairness.

6. Conclusion

This paper proposes an efficient and novel approach for LLM-generated text called MFAD, which adopts a multi-feature fusion approach. The model integrates handcrafted syntactic and statistical features with deep semantic representations learned through a neural architecture composed of GloVe embeddings, CNN, and BiLSTM. MFAD demonstrates strong performance, achieving confidence scores of up to 98% on the HC3 dataset. Unlike many existing approaches that rely on a single type of feature or use simpler, more transparent models, MFAD combines the interpretability of linguistic features with the strength of deep learning for a more robust and flexible solution. Future work will focus on enlarging the dataset, exploring cross-domain generalization, and incorporating more advanced large language models (such as GPT-4, Gemini, and Claude). Addressing adversarial attacks and paraphrased content is a very important point that should be considered. Overall, MFAD presents a promising foundation for reliable and explainable detection of LLM-generated text in an era of rapidly evolving generative AI.

References

1. G.Team, R.Anil, J.Yu, R.Soricut, S.Borgeaud, J.B.Alayrac, and J.Schalkwyk. "Gemini: a family of highly capable multimodal models." arXiv preprint; Xiv:2312.11805, 2023.
2. J.Achiam, S.Adler, S.Agarwal, , F.L.Aleman, L.Ahmad, I.Akkaya, and D.Almeida et al. "Gpt-4 technical report." arXiv preprint; arXiv:2303.08774, 2023.
3. A.Priyanshu, Z.Hong, and Y.Maurya. "AI Governance and Accountability: An Analysis of Anthropic's Claude." arXiv preprint; arXiv:2407.01557, 2024.
4. N. Dehouche, "Plagiarism in the age of massive generative pre-trained transformers (GPT-3)," *Ethics Sci. Environ. Politics*, vol. 21, pp. 17–23, 2021.
5. A. Giaretta and N. Dragoni, "Community targeted phishing," in *Proc. 6th Int. Conf. Softw. Eng. Defence Appl.*, P. Ciancarini, M. Mazzara, A. Messina, A. Sillitti, and G. Succi, Eds. Cham, Switzerland: Springer, 2020, pp. 86–93.
6. A.Kurenkov. "Lessons from the GPT-4Chan Controversy." *The Gradient* (2022).
7. K. Shu, S. Wang, D. Lee, and H. Liu, "Mining disinformation and fake news: Concepts, methods, and recent advancements," in *Disinformation, Misinformation, and Fake News in Social Media*. Cham, Switzerland: Springer, pp.1–19, 2020.
8. J.Wu, L.S.Chao, S.Yang, Y.Yuan, and D.Fai Wong. "A survey on LLM-generated text detection: Necessity, methods, and future directions." *Computational Linguistics*; pp. 1-65, 2025.
9. K.C.Fraser, H.Dawkins, and S.Kiritchenko. "Detecting ai-generated text: Factors influencing detectability with current methods." *Journal of Artificial Intelligence Research* 82 , pp. 2232-2278, 2025.

10. J.Devlin, M.W.Chang, K.Lee, and K.Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint; arXiv:1810.04805, 2018.
11. Y.Liu, M. Joshi, D.Chen, M.Ott, N.Goyal, J.Du, O.Levy, and M.Lewis, L. Zettlemoyer, and V.Stoyanov. "Roberta: A robustly optimized bert pretraining approach". arXiv preprint; arXiv:1907.11692 , 2019.
12. J.Pennington, R.Socher, and C.D. Manning. "Glove: Global vectors for word representation." In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp. 1532 – 1543, 2014.
13. H.Zhou. "Research of text classification based on TF-IDF and CNN-LSTM." In journal of physics: conference series; vol. 2171, no. 1, p. 012021. IOP Publishing, 2022.
14. S.Gehrmann,, H.Strobel, and A.M. Rush . "Gltr: Statistical detection and visualization of generated text." arXiv preprint; arXiv: 1906.04043, 2019.
15. E.Mitchell, Ch.D. Manning, Y.Lee, A.Khazatsky, and Ch.Finn. "Detectgpt: Zero-shot machine-generated text detection using probability curvature." In International Conference on Machine Learning; pp. 24950-24962. PMLR, 2023.
16. X.Liu, Z.Zhang, Y.Lan, Y.Wang, H.Pu, and C.Shen. "Coco: Coherence-enhanced machine-generated text detection under low Investigating Features for ChatGPT, Gemini and BingAI." European Journal of Education 60, no. 1, 2025.
17. A. Modupe, V. Marivate, T. Celik, and O. Olugbara, "Post- authorship attribution using regularized deep neural network," Applied Sciences; vol. 12, p. 7518, 07, 2022.
18. X.Hu, P.Y.Chen, and T.Y.Ho. "Radar: Robust ai -text detection via adversarial learning." Advances in Neural Information Processing Systems 36; pp. 15077-15099, 2023.
19. B.Zhou, Z.Wang, L.Wang, X.Sui, H.Wang, Y.Zhang, K.Song ,and K.Wong. "DPDLLM: A Black-box Framework for Detecting Pretraining Data from Large Language Models." In Findings of the Association for Computational Linguistics ACL, pp. 644- 653. 2024.
20. P.Wang, L.Li, B.Jiang, D.Zhang, K.Ren, and X.Qiu. "SeqXGPT: Sentence-level AI-generated text detection." arXiv preprint ; arXiv: 2310.08903 , 2023.
21. L.Li, T.Sun, P.Wang, K.Ren ,and X.Qiu. "Origin tracing and detecting of llms." arXiv preprint ; arXiv: 2304.14072 , 2023.
22. Z.Guo, and S.Yu. "AuthentiGPT: Detecting machine-generated text via black-box language models denoising." arXiv preprint ; arXiv: 2311.07700 , 2023.
23. M. D. N. Arusada, N. A. S. Putri, and A. Alamsyah, "Training data optimization strategy for multiclass text classification", in 5th International Conference on Information and Communication Technology (ICoIC7); IEEE, pp. 1–5, 2017.
24. Y. Jeong, J. Yang, J.J. Song, and S. Kang, Advancing tinnitus therapeutics: Gpt-2 driven clustering analysis of cognitive behavioral therapy sessions and google t5-based predictive modeling for thi score assessment; IEEE Access, 2024.
25. B.Guo, X.Zhang, J.Yue, Z.Wang, M.Jiang, J.Nie, Y.Ding, and Y.Wu. "How close is chatgpt to human experts? comparison corpus, evaluation, and detection." arXiv preprint; arXiv:2301.07597, 2023.