International Journal of Research in Engineering, Science and Management
Volume 7, Issue 11, November 2024
https://www.ijresm.com | ISSN (Online): 2581-5792

Humanizing AI-Generated Text: Techniques and Future Directions

Riju Marwah*
Department of Computer Science Engineering, Guru Gobindh Singh Indraprastha University, Delhi, India

***Abstract*: Content creation has been revolutionized by advanced natural language generation (NLG) models but this also poses a challenge in differentiating AI-written and human-written texts. AI Humanizers have been designed to modify AI-written content into more natural, human-like words which are also contextually appropriate. This paper explores the techniques used by these humanizers to achieve content that seems human-written, often rendering AI-detectors ineffective. Methods such as lexical substitution, sentence restructuring sentiment adjustment, etc. This paper aims to give a deep understanding of how AI humanizers function.**

***Keywords*: AI-Generated text, Natural Language Processing (NLP), Language models, AI humanizers.**

1. Introduction

The advancements of natural language generation (NLG) models have led AI-generated content highly prevalent across various industries. Popular models such as ChatGPT, T5 have continuously demonstrated the ability to generate indistinguishably human-like content on demand which is also contextually accurate. However, these models lack the ability to introduce subtle features of human-communication—lack of emotional expression, tone variations, grammatical errors, etc. Certain patterns emerge—such as repetition, overly formal, unnaturally phrased content. This giveaway its non-human origin.

A. Problem Statement

Despite the advancements in NLG models, it remains a challenge to achieve human-like qualities in AI-generated text. While machine generated texts are grammatically and contextually correct, they still lack emotion expression, variation in tone, etc. **[1]** Machine generated texts exhibit patterns that can be recognized—these can be repetitiveness, overly formal tone, and rigid transitions. AI humanizers therefore have been created to bridge this gap between AI-generated and human-generated texts, where the challenges lie in using natural language processing (NLP) techniques to make AI-generated texts as indistinguishable from human-written texts as possible.

B. Objectives

The objective of this paper is to explore how various NLP and AI techniques are used to humanize AI-generated texts to bridge the gap between machine generated and human-written texts. These techniques aim to make texts more emotionally expressive, fluid, have tone variations and contextually accurate. Specifically, this paper will explore the following techniques:

- *Lexical substitution* to select context appropriate words through embedding.
- *Style transfer* to *adjust the tone* of writing style of text
- *Tokenization* & *syntactic parsing* to analyze and restructure the sentences.
- *Sentiment Analysis* & *Tone Adjustment* to simulate human emotion through text.
- *Controlled Imperfections* to simulate human error into the content.

C. Scope and Significance

It is crucial to produce quality, human-like content in various field to maintain engagement, trust and credibility. Humanizers play a vital role in performing the process of converting machine generated to human-like content ensuring users do not feel alienated by the content appearing artificial or robotic. By making machine generated content more human-like, it will be more relatable and appropriate contextually. Continued advancement of AI humanization techniques are essential as several industries rely on it more while maintaining high standards and quality of content.

2. Literature Review

Advancements in NLPs have directed the development of highly sophisticated content generation models like GPT, generating human-like content by using pre-trained datasets and BERT which excels in understanding context of words related to one another.

While models like GPT and BERT are highly effective, their machine generated output lacks subtle features of human communication. They exhibit patterns that are detectable as machine written. Certain techniques have been explored that can bridge this gap and humanize machine generated content by including varying tone, styling adjustments, controlled imperfections, etc. that can fool AI-detectors

Modern NLP require fundamental embedding techniques like BERT and GloVe. This is where humanizers introduce

*Corresponding author: marwah.riju@gmail.com

variability to these embeddings making machine generated content more fluid and have more human-like features. These AI humanizers have application in various sectors such as marketing and content creation. When the content is more human-like it is more relatable and appropriate contextually.

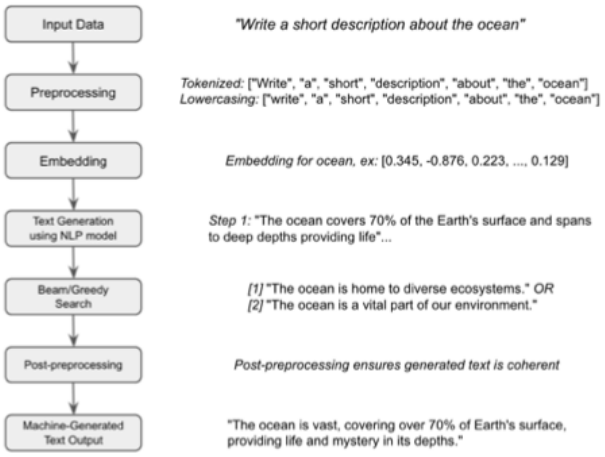

Fig. 1. Artificial Intelligence text generation using NLP model

3. Techniques and Methods

To understand the humanization of text generated by artificial intelligence, it has to be understood how using NLP models this text is generated in the first place. This process can be visualized as put in *figure. 1.* The process of text generation using NLP models begins at the input of raw data, which can be a simple text prompt or any structured data. The preprocessing of this data is the first task at hand that is performed by tokenization separating the text into words or subwords compressing all letters to small cases. Concerning the tokens, after this step of the processes, they become embeddings that are known vector representations of words considered to be in a continuous vector space using GloVe, Word2Vec, or BERT. These embeddings are then provided to the language models such as GPT and BERT that are based on neural networks for predicting and generating the next word on a provided sequence considering the context of the previous word. In this way, analyzing relationships in the context of the pre-trained model and in the context of the task, generating sentences is also context-sensitive. Then, appropriate decoding techniques such as beam search and greedy search are employed which are used to obtain the most relevant sequence of internets. This has been done in order to improve the quality of the output with regards to the structure and meaning of the passage. The output is discharged with the post-preprocessing stage where the output is edited to remove invalid formatting, grammatical and punctuation mistakes. De-tokenization comes in handy and reassembles the words will as earlier outputs that are of little importance are no more as the final output is in a format acceptable for people. However, the generated content contains certain detectable markers which still mark it as being "machine written".

To counter this pattern detection, a variety of techniques are applied to the machine generated content in order to humanize it. The following are the techniques.

A. Tokenization and Syntactic Parsing

The text is broken down into several subwords and characters, or individual tokens that will be processed. The text may go through subword tokenization e.g. with Byte Pair Encoding (BPE) to handle uncommon words. *Constituency* or *dependency parsers* are used in analyzing the structure of sentences to gain knowledge of the grammatical relations—identifying verbs, objects, etc. To map the structure of sentences, dependency trees and constituency trees are constructed to make intelligent edits to the sentence. Dependency trees store words at each node, in contrast, constituency trees only store words in the leaves, and the nodes are marked with part-of-speech tags.

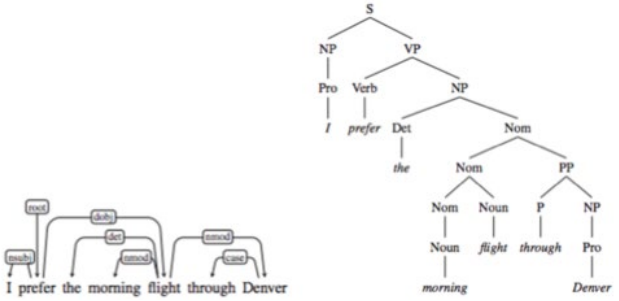

Fig. 2. Comparison between constituency and dependency tree [4]

B. Synonym Replacement and Lexical Substitution

Word2Vec, GloVe, transformer-based embeddings (e.g. BERT) are used to identify contextually correct synonyms. The models capture word meanings in multi-dimensional space, allowing for humanizers to perform context-aware lexical substitution. Models like BERT are used to replace certain words with synonyms that would fit the specific context of the original sentence. Substitution is done while ensuring meaning of the actual sentence isn't affected. Before substitution, Part-of-Speech tagging (POS) is performed to ensure that words with appropriate grammatical category are replaced to maintain grammatical integrity.

C. Sentence Restructuring and Paraphrasing

To rephrase or paraphrase sentences, humanizers often rely on *sequence-to-sequence* (Seq2Seq) models which have attention mechanisms (transformers like BART, T5). These models learn mapping in between AI generated sentences and their human-like paraphrases. Humanizers often deploy *syntactic transformations* like passivization, clefting, fronting to add variability to AI-generated texts lack variation in sentence structure. "He cop chased the criminal" can be transformed into "The criminal was the one who was chased by the cop" to add complexity to the sentence. *Beam search* or *temperature-controlled* sampling is performed to explore multiple sentence structures, to avoid repeating patterns in the output which are often tagged as AI-generated [5].

D. Sentiment and Tone Adjustment

Sentiment analysis models (e.g. BERT-based classifiers) may be used by humanizers to assess emotional tone of the of

Table 1
AI-generated vs Humanized text and scores

AI-generated Text	Humanized Text	AI Detection (Written by AI?)
The coffee machine whirred softly as the aroma of freshly brewed coffee filled the air.	The coffee machine let out a soft swoosh as the wonderful smell of rich coffee permeated the room.	0%
He picked up the guitar, strumming a familiar tune that echoed through the quiet room.	He grabbed the guitar and began to play a song that would once again oversaturate the stillness in the air.	0%
A rainbow appeared after the storm, its vibrant colors stretching across the sky.	Post storm there was a rainbow, and the colors radiated wide.	0%
The cat leapt onto the windowsill, basking in the golden sunlight of the afternoon.	The feline sprang to the windowsill envy of an afternoon's golden rays.	0%
A gentle breeze rustled the leaves, carrying with it the scent of fresh rain.	There were soft sounds of the leaves dancing carried by the breeze which also bore other fresh scents the smell of rain.	0%

machine generated text and edit it to adjust a targeted tone—neutral, positive, negative. Humanizers also fine-tune GPT3 or T5 on sentiment-tagged data, allowing for conditional text generation with desired emotional or politeness tone. Empathy and Enthusiasm can be adjusted into the text to make it sound more engaging using affective computing models.

E. Error and Human-Like Imperfections

Adding controlling noise to emulate human imperfection is a key tactic for humanizers. This involves addition of random perturbation in grammar or spellings, while ensuring it does not affect readability and clarity but also reduces perfection to the text which is easily detectible. Similar to regularization techniques in machine learning, certain humanizers apply a variation of dropout in the text. Small elements of grammar and punctuation are ignored to mimic human-like errors and hesitation.

F. N-gram and Language Model Soothing

N-gram models are used to compare word sequences in AI-generated with human-written text. Humanizers adjust the text to match n-gram distributions similar to humans. Kneser-Ney smoothing or interpolation ensure smoothness in transition between words or sentences in machine generated text to align them more closely with human language.

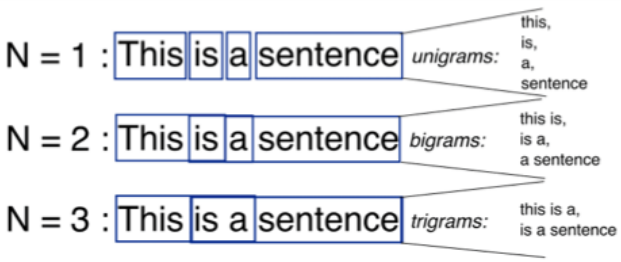

Fig. 3. N-gram model distribution [6]

G. Data Augmentation and Preprocessing

Back-translation is performed where text is translated into a second language and translate it back into the immediate text in order to create different variations of AI text and achieve natural paraphrasing. Humanizers leverage data augmentation techniques to extend its training dataset by slight modifications utterance, data augmentation is frequently applied.

H. Evaluation Metrics and Refinement

Metrics like Perplexity—how unpredictable given text is and burstiness—variation in sentence length and complexity to ensure prevention of repetitive, predictable patterns. Humanizers may use metrics like BLEU (Bilingual Evaluation Understudy) and ROGUE (Recall-Oriented Understanding) to assess the similarities between machine-generated and human-like revisions.

4. Effectiveness, Challenges & Future

A. Effectiveness of Humanization Techniques

The effectiveness of humanization can be assessed by judging if AI detectors can recognize patterns that AI-generated text possess. The following shows a comparison of GPT generated texts and it processed through a humanizer, and the score given by a well-known AI detector

B. Challenges and Limitations

Certain challenges remain in completely humanizing and removing all traces of AI-generated text, particularly in complex texts, remaining still detectable by certain detectors post humanizing. Emotional expression, nuanced tone, cultural understanding, etc. still remain areas where humanizers lack behind.

C. Future Directions

A certain technique that is often not taken leverage of by humanizers is *Style Transfer* which could enhance humanizer outputs significantly. Models can be trained to mimic the writing style of a specific author; this could make the AI-generated text much more unique and less machine generated. Most humanizers do not deeply focus on consistent tone through the entirety of a text. Style transfer could improve text by ensuring a consistent tone and emotion throughout a piece.

Certain challenges present themselves while utilizing this significant technique—such as the large data requirements needed to train models in order to mimic a certain writing style while also being highly complex to maintain semantic fidelity rather than directly restructuring sentences. Despite these challenges and gaps, Style Transfer could lead to much more natural and low AI-detectable texts. Combined with presently used techniques, text indistinguishable from human-written text can be generated.

5. Conclusion

AI humanization techniques have emerged as a vital tool for processing machine generated text to mimic human-written text, improving its clarity, readability, making it coherent and human-application acceptable. However, certain challenges

still remain as humanizers often lack in completely humanizing pieces especially in nuance, emotional depth and contextual understanding. As AI will continue to evolve, so will the techniques used in humanization, allowing for humanized text to play a vital role in communication, marketing and more. This technology has a promising future and promises to soon bridge the gap between human and machine generated text, allowing for many opportunities in various industries.

References

[1] Omar, M., Choi, S., Nyang, D., & Mohaisen, D. (2022). Robust natural language processing: Recent advances, challenges, and future directions.

[2] Narayan, S., Khetarpal, K., Khashabi, D., Deshpande, A., Hase, P., Khot, T., Wang, S., Hamilton, W., & Sabharwal, A. (2023). Who wrote this? Detecting artificial intelligence-generated text from human-written text.

[3] Smith, A., Johnson, B., Lee, C., & Patel, D. (2022). Humanizing AI-generated text: Techniques and applications.

[4] Ogden, A., & Lai, L. (2020). Computational approaches to the syntax-prosody interface: Using prosody to improve parsing.

[5] Gupta, A., Agarwal, A., Singh, P., & Rai, P. (2018). A deep generative framework for paraphrase generation. *Proceedings of the AAAI Conference on Artificial Intelligence*. https://www.cse.iitk.ac.in/users/piyush/papers/deep-paraphrase-aaai2018.pdf

[6] Brownlee, J. (2022, June 21). N-gram language modeling in natural language processing. *KDnuggets*. https://www.kdnuggets.com/2022/06/ngram-language-modeling-natural-language-processing.html