AI-generated Text detection

Shivam Shrikant Zample.
B.Tech in Mathematics and Computing.
IIT Ropar, Class of 2026
2022mcb1280@iitrpr.ac.in

Ashu Kaushik
PhD scholar,Department of computer science and engineering
ashu.kaushik@iitrpr.ac.in

1 July,2024

Abstract

Ongoing development in gpt models have raised the concern about the misuse of generated texts like causing disruptions in education system,cheating writers etc.Many detection models are proposed .The major challenges they are facing include detecting paraphrased text and high false positive.This study deals with possible method to tackle these issues.

1 Introduction

Large Language Models (LLMs) represent a significant breakthrough in the area of natural language processing (NLP).The ongoing advancements in the generative models keeps on reducing the gap between human and ai generated text making it more difficult to distinguish between them.

This study first examines the statistical differences between human and AI-generated text. Factors considered include entropy distribution, part-of-speech and dependency analysis, and sentiment analysis. Furthermore, we will look at how the entropy distribution can be used to detect whether the text is human or AI generated, followed by using GAN-BERT to tackle paraphrasing attacks.

In addition to the above methods, we explore the potential of transfer learning in enhancing the detection capabilities of our models. Transfer learning, which involves pre-training a model on a large corpus of data before fine-tuning it on a specific task, has shown promising results in various NLP applications. By leveraging pre-trained models like BERT, RoBERTa, and GPT-3, we can improve the sensitivity of our detection systems to subtle nuances in text that distinguish human writing from AI-generated content. This approach not only accelerates the training process but also enhances the model's ability to generalize across different types of text and contexts, making it a robust tool in the ongoing effort to maintain the integrity of written communication.

2 Statistical Contrasts in Human vs AI Text

NOTE: The observations are made on the basis of this: data.

2.1 POS and dependency analysis

Figure 1 illustrates two significant observations about AI-generated texts: High noun count indicates more informativeness and objectivity, which can often be observed in AI-generated texts. AI models like GPT are trained on vast datasets, allowing them to produce text rich in factual content and specific details. This characteristic is reflected in the higher frequency of nouns, which serve as the primary carriers of information. For instance, an AI-generated text about a scientific topic might densely pack terms related to the subject matter, contributing to a more informative and objective style.
Less punctuation count indicates that GPT prefers to