TextVerify: A Better AI Detection Model
Dev Bhatia

Introduction

AI is used very prominently in today's day, especially in schools. Students tend to use it as a way to get assignments done. Students also tend to take AI text content, and paraphrase it in order to evade detectors. This leads to difficulty telling original work from AI generated text, especially for teachers. Current AI detectors accuracy isn't that high. Also, the use of AI is a problem because it can lead to students having a lack of knowledge in key areas. This study aims to create a better AI detector to solve these issues.

Related Work and Novelty

Some commonly used AI Detectors are GPTZero, CrossPlag, and Copyleaks. They detect well on older GPT generations, but do not work as well with GPT 4 or AI that was paraphrased. Unlike conventional AI detectors, TextVerify employs advanced retrieval-based methods—a sophisticated technique that fetches and compares relevant information from a vast database. By leveraging cosine similarity scoring and precise database matching, it ensures highly accurate text verification.

Methods

TextVerify begins by converting the input text into a numerical vector representation using Doc2Vec, a technique that captures semantic meaning. The system maintains a pre-processed database of AI-generated content, where each entry is similarly encoded into vectors using the same Doc2Vec model. During verification, the algorithm calculates cosine similarity scores between the input vector and all vectors in the reference database. This metric quantifies how closely the input text resembles known AI-generated content. If the similarity score exceeds a predetermined confidence threshold, the system classifies the text as likely AI-generated. The entire process is illustrated in Figure 1.
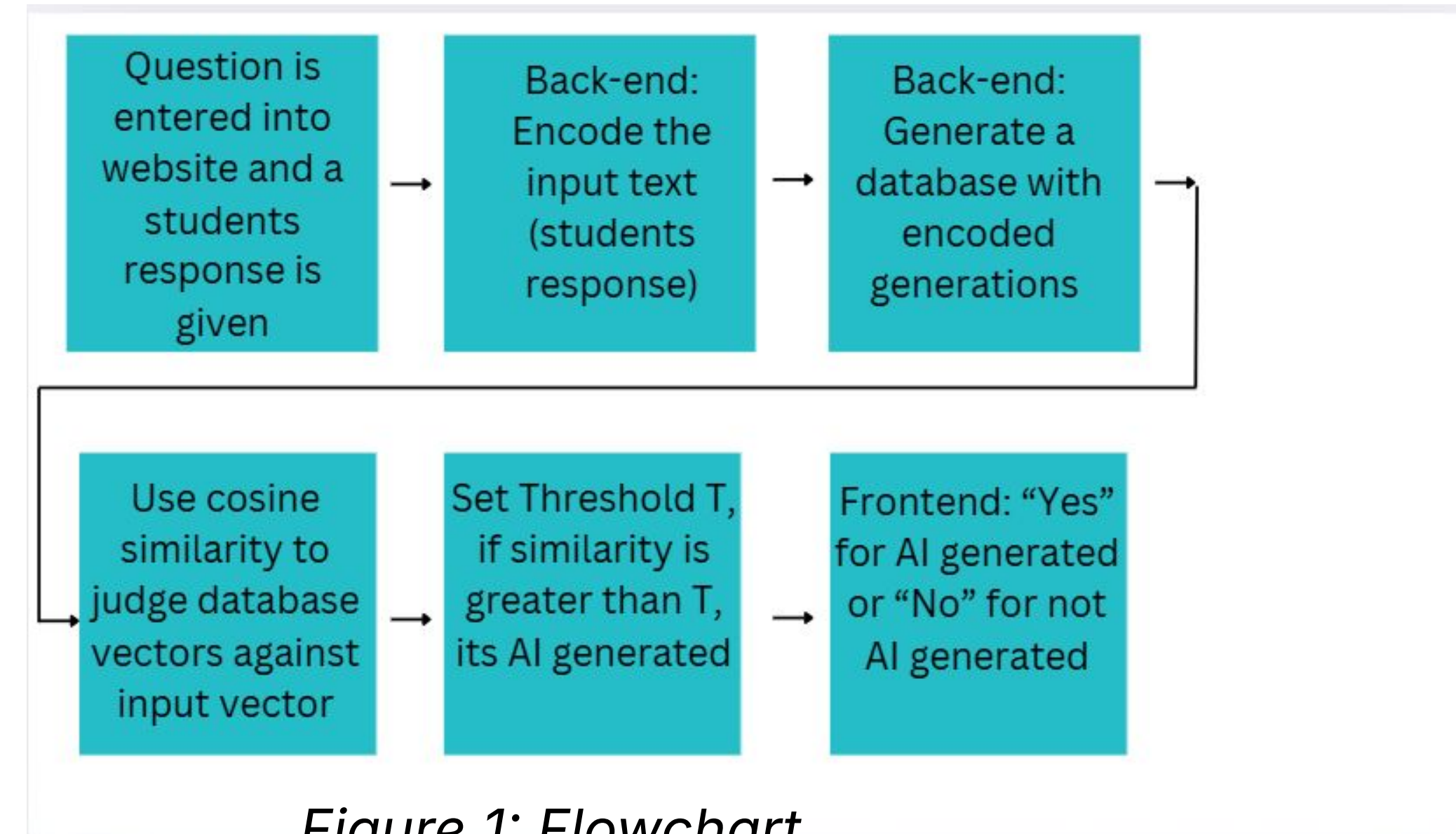

Figure 1: Flowchart

TextVerify is accessible through a website. Which takes 2 inputs: (1) student written text to be tested for AI generation, (2) the prompt. To achieve this, we use a Doc2Vec encoder to encode the student text. Then, we generate responses with AI to the prompt, after that we Doc2Vec encode those responses, and store them in a database. Finally, a threshold (T) is set and the cosine similarity score is calculated. If the cosine similarity score between any vector in the database and the input text vector is higher than T, then the input text is judged as AI Generated. The results are shown to the user.

Figure 3: TextVerify Website Home Page

Results

TextVerify achieved 91.9% accuracy, which is higher than any of the AI detection tools used today. Figure 2 illustrates the accuracy TextVerify has with different types of AI and human written text.

Model/Category	Correctly Classified	Total Tested	Accuracy (%)
GPT-4 Output	10	10	100
Human Written	9	10	90
DeepSeek Output	2	3	66.7
Copilot Output	2	3	66.7
Grok Output	3	3	100
GPT Paraphrased	2	2	100
DeepSeek Paraphrased	2	2	100
Copilot Paraphrased	2	2	100
Grok Paraphrased	2	2	100
Total Accuracy	34	37	91.9

Figure 2: Results

Conclusion and Future Work

I solved the problem of AI Detection using Retrieval Methods with 91.9% accuracy. Some future work includes improvements in AI text detection for other purposes besides school. Also, someone can use retrieval methods along with other methods (watermarking, statistical outlier) to improve classifying shorter text as AI generated or not.

Figure 4: TextVerify Website Report