Document Title: **Document — 11/28/2024**

We are <u>highly confident</u> this text is ai generated
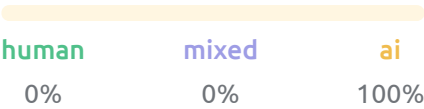
Classification	**Probability Breakdown**	**Plagiarism Scan**

ai

100% probability AI generated

● highly confident

human 0%   mixed 0%   ai 100%

?

Plagiarism scan not activated

AI Scanned Document

☐ AI Content

Document — 11/28/2024

Weaknesses:

It appears that your CONDITIONAL ALIGNMENT primarily facilitates attention-based interaction between the GPT-generated prompts and the corresponding content. However, I am unclear about the specific significance of taking the dot product of these two attention maps. From my perspective, your alignment module seems to merely apply attention mechanisms followed by a dot product, which does not appear to introduce any substantive algorithmic novelty. Could you elaborate further on the theoretical or empirical contributions this approach provides beyond the existing methods?

Your multimodal fusion module appears to simply concatenate features from different modalities and feed them into a transformer encoder. This approach is quite common and widely adopted in existing literature.

In a word, it seems that the paper primarily applies the teacher-student model paradigm to the domain of multi-modal sentiment analysis (MSA), incorporating GPT-generated content as prompts. While the motivation is sound, the implementation appears somewhat simplistic, lacking sufficient innovation to substantiate a significant contribution.

The selection of baselines in your comparison is quite limited, and notably, none of the baselines are from 2024. Given that this field remains highly active and rapidly evolving, I strongly recommend including more recent baselines from 2024 to provide a more comprehensive and current evaluation of your proposed approach.

The analysis presented in the "EFFECT OF EACH COMPONENT" section appears rather superficial and lacks depth, raising the concern that it may have been generated by AI without sufficient refinement or critical examination.

Since the goal is to train a student model with reduced complexity, it would be highly informative to include a comparison of parameter counts with other baselines. Such a comparison would help substantiate claims regarding the efficiency and compactness of the student model relative to existing approaches.

Questions:

It appears that your CONDITIONAL ALIGNMENT primarily facilitates attention-based interaction between the GPT-generated prompts and the corresponding content. However, I am unclear about the specific significance of taking the dot product of these two attention maps. From my perspective, your alignment module seems to

merely apply attention mechanisms followed by a dot product, which does not appear to introduce any substantive algorithmic novelty. Could you elaborate further on the theoretical or empirical contributions this approach provides beyond the existing methods?

Your multimodal fusion module appears to simply concatenate features from different modalities and feed them into a transformer encoder. This approach is quite common and widely adopted in existing literature.

In a word, it seems that the paper primarily applies the teacher-student model paradigm to the domain of multimodal sentiment analysis (MSA), incorporating GPT-generated content as prompts. While the motivation is sound, the implementation appears somewhat simplistic, lacking sufficient innovation to substantiate a significant contribution.

The selection of baselines in your comparison is quite limited, and notably, none of the baselines are from 2024. Given that this field remains highly active and rapidly evolving, I strongly recommend including more recent baselines from 2024 to provide a more comprehensive and current evaluation of your proposed approach.

The analysis presented in the "EFFECT OF EACH COMPONENT" section appears rather superficial and lacks depth, raising the concern that it may have been generated by AI without sufficient refinement or critical examination.

Since the goal is to train a student model with reduced complexity, it would be highly informative to include a comparison of parameter counts with other baselines. Such a comparison would help substantiate claims regarding the efficiency and compactness of the student model relative to existing approaches.

FAQ

What is GPTZero?

GPTZero is the leading AI detector for checking whether a document was written by a large language model such as ChatGPT. GPTZero detects AI on sentence, paragraph, and document level. Our model was trained on a large, diverse corpus of human-written and AI-generated text, with a focus on English prose. To date, GPTZero has served over 2.5 million users around the world, and works with over 100 organizations in education, hiring, publishing, legal, and more.

When should I use GPTZero?

Our users have seen the use of AI-generated text proliferate into education, certification, hiring and recruitment, social writing platforms, disinformation, and beyond. We've created GPTZero as a tool to highlight the possible use of AI in writing text. In particular, we focus on classifying AI use in prose. Overall, our classifier is intended to be used to flag situations in which a conversation can be started (for example, between educators and students) to drive further inquiry and spread awareness of the risks of using AI in written work.

Does GPTZero only detect ChatGPT outputs?

No, GPTZero works robustly across a range of AI language models, including but not limited to ChatGPT, GPT-4, GPT-3, GPT-2, LLaMA, and AI services based on those models.

What are the limitations of the classifier?

The nature of AI-generated content is changing constantly. As such, these results should not be used to punish students. We recommend educators to use our behind-the-scene Writing Reports as part of a holistic assessment of student work. There always exist edge cases with both instances where AI is classified as human, and human is classified as AI. Instead, we recommend educators take approaches that give students the opportunity to demonstrate their understanding in a controlled environment and craft assignments that cannot be solved with AI. Our classifier is not trained to identify AI-generated text after it has been heavily modified after generation (although we estimate this is a minority of the uses for AI-generation at the moment). Currently, our classifier can sometimes flag other machine-generated or highly procedural text as AI-generated, and as such, should be used on more descriptive portions of text.

I'm an educator who has found AI-generated text by my students. What do I do?

Firstly, at GPTZero, we don't believe that any AI detector is perfect. There always exist edge cases with both instances where AI is classified as human, and human is classified as AI. Nonetheless, we recommend that educators can do the following when they get a positive detection:

 Ask students to demonstrate their understanding in a controlled environment, whether that is through an in-person assessment, or through an editor that can track their edit history (for instance, using our Writing Reports through Google Docs). Check out our list of several recommendations on types of assignments that are difficult to solve with AI.
 Ask the student if they can produce artifacts of their writing process, whether it is drafts, revision histories, or brainstorming notes. For example, if the editor they used to write the text has an edit history (such as Google Docs), and it was typed out with several edits over a reasonable period of time, it is likely the student work is authentic. You can use GPTZero's Writing Reports to replay the student's writing process, and view signals that indicate the authenticity of the work.
 See if there is a history of AI-generated text in the student's work. We recommend looking for a long-term pattern of AI use, as opposed to a single instance, in order to determine whether the student is using AI.