# Article information:

Did ChatGPT Write This? This AI Technique Can Help You Identify AI Written Text - MarkTechPost
<https://www.marktechpost.com/2023/02/08/did-chatgpt-write-this-this-ai-technique-can-help-you-identify-ai-written-text/>

# Article summary:

1. ChatGPT is a large language model (LLM) that can understand and generate text replies to given requests.

2. There is a need to identify if an LLM or a human has written something, as malicious use cases could arise from the use of LLMs.

3. A watermarking algorithm has been proposed to identify AI-generated text, which can be integrated into any LLM without requiring it to be trained again and can be detected from a tiny part of the generated text.

# Article rating:

May be slightly imbalanced: The article presents the information in a generally reliable way, but there are minor points of consideration that could be explored further or claims that are not fully backed by appropriate evidence. Some perspectives may also be omitted, and you are encouraged to use the research topics section to explore the topic further.

# Article analysis:

The article provides an overview of the potential risks associated with using large language models such as ChatGPT, and proposes a watermarking algorithm as a solution for identifying AI-generated text. The article is generally well-written and provides clear explanations of the concepts discussed, however there are some areas where more detail could have been provided. For example, while the article mentions that “synthetic data is usually inferior to human-generated content”, it does not provide any evidence or examples to support this claim. Additionally, while the article discusses potential malicious uses of LLMs such as generating fake news or academic writing assignments, it does not explore any counterarguments or possible solutions for these issues beyond the proposed watermarking algorithm. Furthermore, while the article mentions that “money counterfeiting was a huge issue” in the past, it does not provide any information on how this relates to AI-generated content or why this example was chosen over other examples of trustworthiness issues in other industries. Finally, while the article provides links to further resources at the end of the article (e.g., paper and Github), it does not provide any information on how these resources relate to the topics discussed in the article or why they were chosen over other resources available on these topics. In conclusion, while overall this article provides an informative overview of potential risks associated with using large language models such as ChatGPT and proposes a watermarking algorithm as a solution for identifying AI-generated text, there are some areas where more detail could have been provided in order to make it more comprehensive and trustworthy.

# Topics for further research:

* Evidence of synthetic data inferiority
* Counterarguments to malicious uses of LLMs
* Solutions to malicious uses of LLMs
* Money counterfeiting and trustworthiness
* Resources on large language models
* Watermarking algorithms for AI-generated text

# Report location:

<https://www.fullpicture.app/item/b00ad94d8713905f4ca2c0e7191762f9>