

“ Fortunately, being an AI sleuth is relatively easy, thanks to some emerging smart SaaS (software as a service) tools

INGIMAGE®



THE AI CONTENT CONUNDRUM

Merilee Kern outlines how sentiment degradation by AI is impacting the age of misinformation



With industries across the board becoming increasingly adept at using AI driven large language models (LLMs) such as OpenAI's ChatGPT, Anthropic's Claude 2 and Meta AI's Llama 2, it begs one obvious question: how do we ensure that all the content out there isn't generated by AI?

Fortunately, being an AI sleuth is relatively easy, thanks to some emerging smart SaaS (software as a service) tools. Such solutions will become crucial for those looking to

identify the use of artificial intelligence in content creation, dealing with challenges related to false positives in AI content detection, discerning if AI material has yet been detected and even if AI 'distilled' content has been reconfigured to be more neutral in sentiment.

A study conducted on original content being watered down or repositioned to a more impartial format (that doesn't accurately reflect the originator's voice or purpose) has revealed how pervasive this is and indicated what the relevant implications are.

Merilee Kern (MBA) is an internationally regarded brand strategist and analyst. She is also the creator and host of the Savvy Ventures business TV show that airs on FOX Business TV and Bloomberg TV (she can be contacted at www.TheLuxeList.com and www.SavvyVentures.tv).

phrase another text can offer speed and ease in content production but it comes with caveats.

For example, there might be a good reason for coverage of a news event to have highly negative or positive sentiment. Dampening those qualities will prevent readers from perceiving how potentially troublesome or heartening an event might be.

Outside of news content, publishers may desire to convey a particular kind of sentiment to evoke feelings in readers and a neutral scoring story will struggle to do so.

On the other hand, there could be uses for producing texts with more neutral sentiment that read more like ‘just the facts.’ Publishers may want to consider the tone and purpose of a piece, and know that LLMs could modify texts in ways that affect those goals.

Q: What is sentiment analysis, which was the benchmark employed for your AI paraphrased content study?

A: Sentiment analysis is the process of analysing and categorising texts as positive, neutral or negative, and to what degree.

It’s often used to assess opinions and feelings expressed in reviews or open-ended questions in surveys. Many of the stories in this study had their sentiment made more neutral after generative AI rewrote them.

On the sentiment analysis scale, 1 is highly negative, 5 is highly positive and 3 is neutral. LLMs tended to move a story’s sentiment closer to 3, whether the original writing was more negative or positive. In the aggregate, the rewritten articles had their sentiment flattened.

Q: How was the study methodology undertaken – and what were the key findings and data points?

A: We analysed 100 articles for their sentiment, or how positive or negative they were, and then had them rewritten by three large language models – viz. OpenAI’s ChatGPT, Anthropic’s Claude 2 and Meta AI’s Llama 2.

The new texts’ sentiment scores were then analysed for any changes. The 100 articles utilised in the study, each from popular websites, were rated by Sapien.IO’s sentiment analysis for how positive, neutral or negative each was.

We had three different LLMs – ChatGPT, Claude 2 and Llama 2 – each paraphrase the articles and then analyse the sentiment of the new texts. These ratings were compared to the original articles’ sentiment rating. The score, along with each rewritten article’s word count, was analysed for any relationships.

Key findings include the substantiation that LLM rewrites moved the sentiment scores closer to the middle or neutral part of the scale and the resulting sentiment scores differed by LLM. Llama 2 had the most positive orientation scores with Claude 2 having the most negative.

It’s also important to note that rewritten articles were made shorter than the original, which could be part of the reason that sentiment scores changed.

Overall, the analysis showed that no more than half a point in difference between the original article’s average sentiment analysis score of 2.54 (slightly more negative than neutral) and the LLMs’ rewrite averages of 2.72 (Claude 2), 2.95 (ChatGPT) and 3.08 (Llama 2).

However, those differences became pronounced when considering articles that originally held sentiment scores of 1 or 5. In those cases, the rewrites differed by more than a point and up to 1.5 points on average, and pulled towards a neutral 3. If the original scored 1, the rewrites averaged 2.35. When the original was 5, the rewrites averaged 3.56.

Q: You mentioned that LLM rewrites often resulted in fewer words and this could have impacted the results. Can you elaborate on this?

A: Yes, a possible explanation for the neutralisation in sentiment could be that all three LLMs reduced the number of words when they rewrote articles. Claude 2 reduced words by a notable 43.5 percent compared to 13.5 percent by ChatGPT and 15.6 percent by Llama 2.

While shortening an article can be desirable for some purposes, the reduction might eliminate details or potent phrases that indicate how negative or positive the sentiment of the story is. Losing those details or descriptive words could explain part of the movement toward a rating of 3 (neutral) for stories with either the most positive or negative sentiment.

This study was small but the data displayed suggests a slightly positive correlation between sentiment scores and word counts – with longer texts receiving higher scores. The trend was highlighted by comparing the three LLMs to each other.

Across all levels of sentiment in the original articles, Claude 2 consistently had both the lowest sentiment scores and word counts, while Llama 2 had the highest sentiment scores and word counts.

Originality.AI’s findings note that some of the most popular LLMs used to rewrite or paraphrase another text are making content more neutral in sentiment, and altering the nature and objective of the original (written) work.

Founder and CEO of Originality.AI Jonathan Gillham provides some insights into the study along with implications of its disconcerting findings.

Q: Your study found that popular LLMs such as OpenAI’s ChatGPT, Anthropic’s Claude 2 and Meta AI’s Llama 2 are officially making content more neutral in sentiment. Why does this matter?

A: Employing LLMs to rewrite or para-

