Clickbait Classification Using ChatGPT

Session link: https://chatgpt.com/share/9049b1b6-3bc5-451c-af4d-bfaf21ad2a73

I. Why Errors Happen?

If we give ChatGPT test examples, it will correctly classify all non-clickbait headlines and misclassify several clickbait ones (6 in my case).

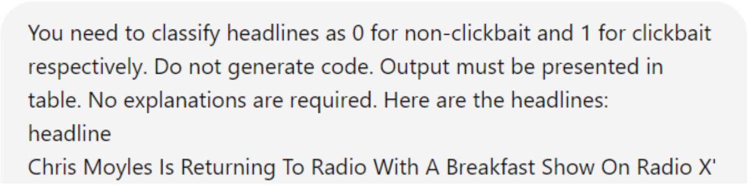
The misclassified headlines were the follows:
- Chris Moyles Is Returning To Radio With A Breakfast Show On Radio X
- How Do You Stay Fit In College
- Tell Us Why You Take Medication For Your Mental Health
- This Web Series Is Helping Trans Teens Find Their Personal Definition Of Beauty
- 41,000 Indians Have Petitioned The Censor Board For More Gender-Sensitive Ratings
- David Bowie's Musical 'Lazarus' Showcases His Enduring Legacy

To discover the reason of this behavior, ChatGPT was asked to define the factors that could make it classify a clickbait message as non-clickbait and to provide the headlines that could be mistakenly classified as non-clickbait.
As can be seen on the picture 1A, the headlines, provided by ChatGPT, coincide with the ones it misclassified. Judging by its' answer, the factors that confuse ChatGPT are Informational Tone and Specificity. Thus, it was decided to reduce their impact on ChatGPT's decision.

II. Approach

For the current task, zero-shot approach was used. The sequence of actions was as follows:

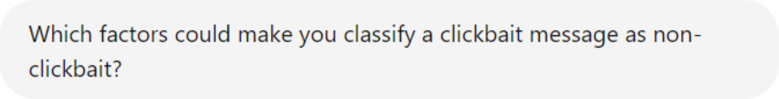
1) <u>Give ChatGPT test examples and obtain the preliminary classification results</u>
   The initial accuracy was 94% (6 examples were wrongly classified as non-clickbait). The head of the prompt is shown on the picture 1.



You need to classify headlines as 0 for non-clickbait and 1 for clickbait respectively. Do not generate code. Output must be presented in table. No explanations are required. Here are the headlines:
headline
Chris Moyles Is Returning To Radio With A Breakfast Show On Radio X'

Picture 1. The head of the first prompt (further headlines are omitted for convenience)

2) Ask ChatGPT to provide factors which could lead to misclassification

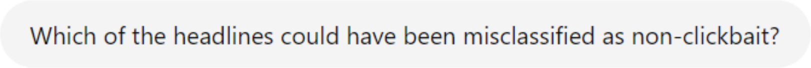


Which factors could make you classify a clickbait message as non-clickbait?

Picture 2. Prompt: Asking ChatGPT for misleading factors

The obtained factors are shown on the picture 2A.

3) Ask ChatGPT to provide headlines which, by its' opinion, might be misclassified



Which of the headlines could have been misclassified as non-clickbait?

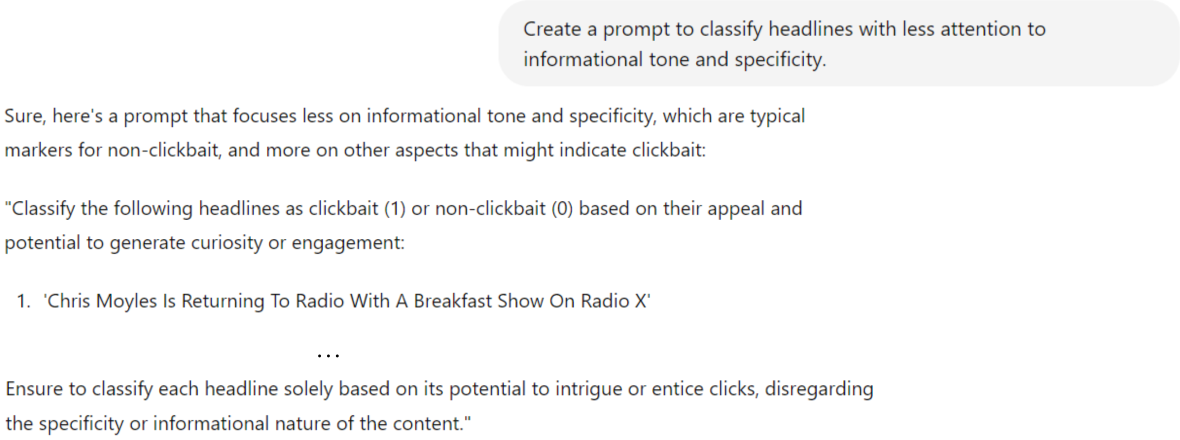
Picture 3. Prompt: Asking ChatGPT which headlines confuse it

The headlines are presented on the picture 1A.

4) Find common factors for the provided headlines

By carefully reading ChatGPT's explanations, I found that common factors were Informational Tone and Specificity. In this case it was obvious, otherwise one could ask ChatGPT to define them.

5) Ask ChatGPT to construct an enhanced prompt w.r.t these factors



Create a prompt to classify headlines with less attention to informational tone and specificity.

Sure, here's a prompt that focuses less on informational tone and specificity, which are typical markers for non-clickbait, and more on other aspects that might indicate clickbait:

"Classify the following headlines as clickbait (1) or non-clickbait (0) based on their appeal and potential to generate curiosity or engagement:

1. 'Chris Moyles Is Returning To Radio With A Breakfast Show On Radio X'

…

Ensure to classify each headline solely based on its potential to intrigue or entice clicks, disregarding the specificity or informational nature of the content."

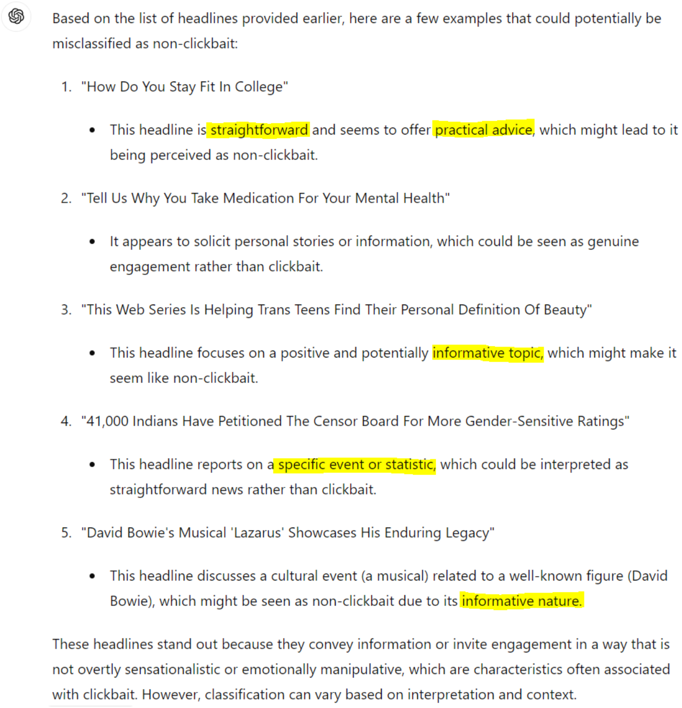
Picture 4. Asking ChatGPT to construct an enhanced prompt. Headlines are omitted for convenience

ChatGPT was asked to pay **less** attention to the found factors. It may appear counter intuitive: should not we focus **more** on these factors instead? However, it could be seen as applying feature engineering, like in any other machine learning method. If including the feature worsens the outcome, we often decide to omit it.

After prompting, the accuracy raised to 100%. Thus, no further steps were taken.

III. Appendix

Pictures

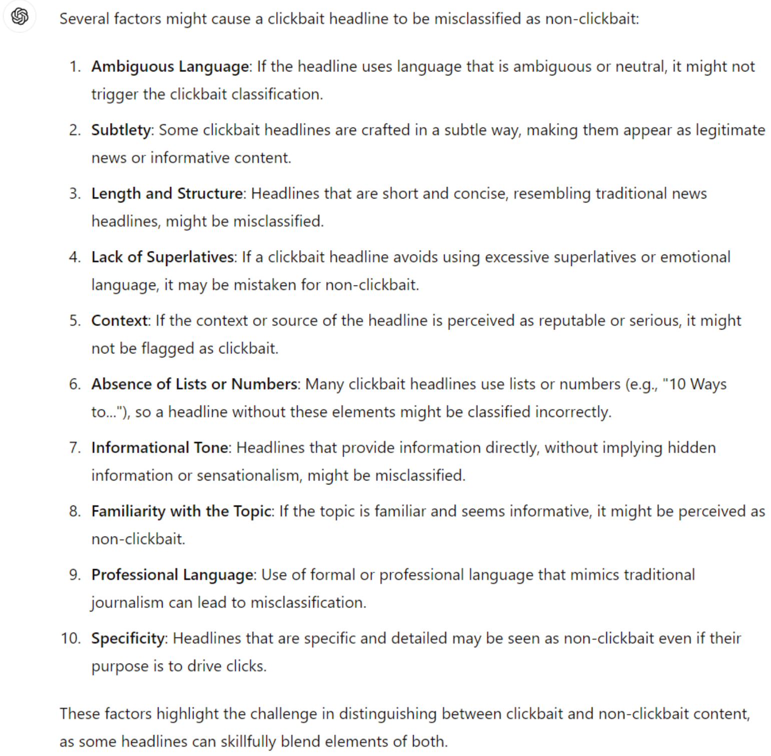


Based on the list of headlines provided earlier, here are a few examples that could potentially be misclassified as non-clickbait:

1. "How Do You Stay Fit In College"
    - This headline is straightforward and seems to offer practical advice, which might lead to it being perceived as non-clickbait.
2. "Tell Us Why You Take Medication For Your Mental Health"
    - It appears to solicit personal stories or information, which could be seen as genuine engagement rather than clickbait.
3. "This Web Series Is Helping Trans Teens Find Their Personal Definition Of Beauty"
    - This headline focuses on a positive and potentially informative topic, which might make it seem like non-clickbait.
4. "41,000 Indians Have Petitioned The Censor Board For More Gender-Sensitive Ratings"
    - This headline reports on a specific event or statistic, which could be interpreted as straightforward news rather than clickbait.
5. "David Bowie's Musical 'Lazarus' Showcases His Enduring Legacy"
    - This headline discusses a cultural event (a musical) related to a well-known figure (David Bowie), which might be seen as non-clickbait due to its informative nature.

These headlines stand out because they convey information or invite engagement in a way that is not overtly sensationalistic or emotionally manipulative, which are characteristics often associated with clickbait. However, classification can vary based on interpretation and context.

Picture 1A. ChatGPT's answer to the following question: "Which of the headlines could have been misclassified as non-clickbait?" The key words, which point to the common misclassification factor, are highlighted in yellow.



Several factors might cause a clickbait headline to be misclassified as non-clickbait:

1. **Ambiguous Language**: If the headline uses language that is ambiguous or neutral, it might not trigger the clickbait classification.
2. **Subtlety**: Some clickbait headlines are crafted in a subtle way, making them appear as legitimate news or informative content.
3. **Length and Structure**: Headlines that are short and concise, resembling traditional news headlines, might be misclassified.
4. **Lack of Superlatives**: If a clickbait headline avoids using excessive superlatives or emotional language, it may be mistaken for non-clickbait.
5. **Context**: If the context or source of the headline is perceived as reputable or serious, it might not be flagged as clickbait.
6. **Absence of Lists or Numbers**: Many clickbait headlines use lists or numbers (e.g., "10 Ways to..."), so a headline without these elements might be classified incorrectly.
7. **Informational Tone**: Headlines that provide information directly, without implying hidden information or sensationalism, might be misclassified.
8. **Familiarity with the Topic**: If the topic is familiar and seems informative, it might be perceived as non-clickbait.
9. **Professional Language**: Use of formal or professional language that mimics traditional journalism can lead to misclassification.
10. **Specificity**: Headlines that are specific and detailed may be seen as non-clickbait even if their purpose is to drive clicks.

These factors highlight the challenge in distinguishing between clickbait and non-clickbait content, as some headlines can skillfully blend elements of both.

Picture 2A. ChatGPT's answer to the following question: "Which factors could make you classify a clickbait message as non-clickbait?"