Appendix for Generating clickbait spoilers with an ensemble of large language models

Mateusz Woźny[1] and **Mateusz Lango**[1,2]

[1] Poznan University of Technology, Faculty of Computing and Telecommunications, Poznan, Poland
[2] Charles University, Faculty of Mathematics and Physics, Prague, Czech Republic
mateusz.wozny@student.put.edu.pl, mlango@cs.put.edu.pl

A Basic dataset characteristics

All experiments were performed on Webis Clickbait 22 corpora (Hagen et al., 2022), which basic characteristics are presented in Table 1. During data preparation, problematic examples for which the beginning of the answer could not be found in the article were removed from the training, validation, and test dataset. Such data preparation is needed to train question-answering models. As a result, the test dataset consisted of 988 examples, instead of 1,000.

B Customized prompts for each spoiler type used in the Vicuna model

- phrase spoilers: "Below is a question paired with a context for which generate an answer. \n Write an answer as short as possible (max 5 words). Use only words from context.\n Question:\n q\n Context: a \n Answer:"

- passage spoilers: "Below is a question paired with a context for which generate an answer. \n Write an answer which will be from one to three sentences. Use only words from context. \n Question:\n q\n Context: a \n Answer:"

- multi-part spoilers: "Below is a question paired with a context for which generate an answer. \n Write an answer which is multi-part which means it contains multiple phrases or sentences from the given text. Use only words from context.\n Question:\n q\n Context: a \n Answer:"

C Examples of generated spoilers

We provide some additional examples of generated spoilers by different approaches and the final spoiler selected by the ensemble in Table 2.

References

Matthias Hagen, Maik Fröbe, Artur Jurk, and Martin Potthast. 2022. Webis clickbait spoiling corpus 2022.

	Train	Validation	Test	Total number
Phrase	1367	335	423	2125
Passage	1274	322	403	1999
Multi-part	559	143	174	876
Total	3200	800	1000	5000

Table 1: Number of examples in train/val/test Webis Clickbait 22 corpora used for experiments

Clickbait	Reference	DeBERTa	LLaMA	Vicuna
Agency might plant a garden on the moon.	NASA	*NASA*	arabidopsis plant basil and turnip	lunar sunlight
10 habits of incredibly happy people	1 they slow down to appreciate life s little pleasure 2 they exercise 3 they spend money on other people 4 they surround themselves with the right people 5 they stay positive	they have a growth mindset	*1 they slow down to appreciate life s little pleasure 2 they exercise 3 they spend money on other people 4 they surround themselves with the right people 5 they stay positive*	*1 they slow down to appreciate life s little pleasure 2 they exercise 3 they spend money on other people 4 they surround themselves with the right people 5 they stay positive*
Kanye West Reveals His Favorite Song... Of All Time.	All along the Watchtower the Jimi Hendrix cover p2cm	all along the watchtower	*all along the watchtower the jimi hendrix cover*	ye ha compared himself to jimi hendrix in the past in a 2012 interview with w magazine he said visiting my mind is like visiting the herme s factory shit is real
You'll never believe what else your gut bacteria can do.	having some control over your gut's bacterial population might actually prevent cancer and save your life	healthy gut bug can stop cancer	can prevent cancer	*having some control over your gut's bacterial population might actually prevent cancer and save your life*
Outcast star 'Patrick Fugit' has a better Dungeons	bandy swiftpet	*bandy swiftpet*	*bandy swiftpet*	patrick fugit
This Is How Much Data The Internet Gets Through In One Minute.	18 264 840 megabyte	*18 264 840 megabyte*	3 6 billion	3 6 billion
Netflix Just Did Something It's Only Ever Done TWICE Before.	end a show	end a show	*bloodline*	*bloodline*

Table 2: Several examples of spoilers generated by different methods. The spoilers in italics were selected by pointwise ranker.