To Click it or not to Click it: An Italian Dataset for Neutralising Clickbait Headlines

Daniel Russo[1,2,*], Oscar Araque[3] and Marco Guerini[2]

[1]University of Trento, Trento, Italy
[2]Fondazione Bruno Kessler, Trento, Italy
[3]Universidad Politécnica de Madrid, Madrid, Spain

Abstract

Clickbait is a common technique aimed at attracting a reader's attention, although it can result in inaccuracies and lead to misinformation. This work explores the role of current Natural Language Processing methods to reduce its negative impact. To do so, a novel Italian dataset is generated, containing manual annotations for classification, spoiling, and neutralisation of clickbait. Besides, several experimental evaluations are performed, assessing the performance of current language models. On the one hand, we evaluate the performance in the task of clickbait detection in a multilingual setting, showing that augmenting the data with English instances largely improves overall performance. On the other hand, the generation tasks of clickbait spoiling and neutralisation are explored. The latter is a novel task, designed to increase the informativeness of a headline, thus removing the information gap. This work opens a new research avenue that has been largely uncharted in the Italian language.

Keywords

clickbait, natural language processing, natural language generation, large language model, language resource

1. Introduction

Accuracy and truthfulness are essential characteristics of journalism. Nevertheless, in an effort to improve revenue, a large number of newspapers and magazines publish *clickbait* articles, a viral journalism strategy that seeks to attract users to click on a link to a page through tactics such as sensationalist stories and catchy headlines that act as bait. The use of these tactics harms the quality of news pieces and thus hinders the ability of citizens to obtain reliable and objective information. The literature distinguishes between two main types of clickbait. (i) *Classical clickbait* [1] embeds within the headlines information gaps, also known as curiosity gaps [2, 3], in order to arouse curiosity in the reader that is forced to access the article's content which is ultimately disappointing. Classical clickbait usually makes use of hyperbolic language, caps lock, demonstrative pronouns and superlative to grasp the user's attention [1, 4, 5]. (ii) *Deceptive clickbait* [5] refers to headlines that resemble traditional media headlines by offering a summary of the article, still leading to content that differs from the reader's expectations. These headlines promise high news value but deliver content with low news value, resulting in reader disappointment.

Although clickbait headlines are considered one of the less harmful forms of fake news, as their main goal is to increase profit by driving traffic to their website [6, 7], they can sometimes pose a danger, especially when they deal with potentially harmful topics such as health and science. To address this problem, Natural Language Processing techniques have been widely employed to detect clickbait headlines, with a particular focus on the English language [8, 9]. Hagen et al. [10] proposed the clickbait spoiling task, i.e., the generation of a short text that satisfies the curiosity induced by a clickbait post.

In light of this, this work addresses the issue of clickbait in the Italian language, studying its characteristics and the possibilities of current technology to reduce its negative impact. In doing so, we have generated a novel Italian dataset that gathers a large collection of clickbait articles, which is made public for the community to use [1]. We named the dataset **ClickBaIT**. This dataset contains manually annotated instances as clickbait/non-clickbait, as well as manually generated spoilers and neutralised headlines. We have also performed a thorough multilingual evaluation, exploiting the availability of English data to complement our dataset in the task of clickbait detection. Finally, this work also explores the use of our annotated dataset and large language models to automatically generate both spoilers and, as a novel task, a neutralised version of clickbait headlines. A graphical illustration of the experimental design is presented in Figure 1.

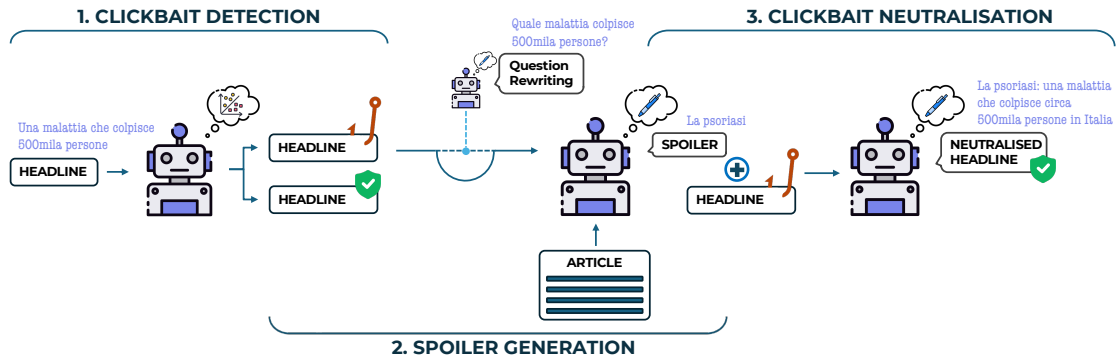[1]The dataset is available in https://github.com/oaraque/ClickBaIT

Figure 1: The experimental design is depicted, encompassing three tasks: *clickbait detection*, *spoiler generation*, and *clickbait neutralisation*. The robot icon represents the language model used for either classification or generation. We utilized *DistilBERT* and *Llama3-8B* for task 1, and *LLaMAntino-3-8B* for tasks 2 and 3. The models were tested for generative tasks using zero-shot, few-shot, and fine-tuning configurations, except for question rewriting, for which we employed a few-shot approach.

2. Related Work

The use of clickbait is common in many news outlets, and thus it has been extensively studied.

There are several works that address clickbait detection: Potthast et al. [8] collected a corpus of clickbait articles, posted by well-known English-speaking newspapers on Twitter, and proposed a set of lexical and semantic features to be used with a Random Forest classifier. Following the general trend in Natural Language Processing (NLP) field, clickbait detection has also been explored using deep learning methods, such as convolutional [11] and recurrent [12] neural networks, as well as more recent Transformer-based approaches [9].

Other works leveraged Natural Language Generations (NLG) strategies to create a piece of text, the *spoiler*, comprising the information needed to fulfil the curiosity gap present in clickbait headlines. This task was proposed by Fröbe et al. [13] with the name of *spoiling generation*. The authors created the *Webis Clickbait Spoiling Corpus 2022*, and cast spoiler generation as a Question Answering task. Eventually, they open the challenge to the community through a SemEval-2023 shared task [13, 14]. The optimal spoiler generator operates with five independent sequence-to-sequence generative models. It selects the best spoiler through a majority vote, determined by comparing edit distances among the outputs [15].

Regarding the languages studied, the majority of works are based on English. Other works were performed in Chinese [16], Turkish [17, 18] and Spanish [19, 20]. To the best of our knowledge, this is the first work that fully addresses the study of clickbait detection and spoiling in the Italian language. Moreover, we propose a novel task, i.e., *clickbait neutralisation*, which aims at filling the curiosity gap by rewriting the headline levering the information of the spoiler.

3. Dataset

3.1. Dataset Creation

Data were collected from fourteen news websites[2], notorious for acting as news aggregators, engaging in plagiarism, lacking fact-checking, and using sensational headlines to draw in readers. In all the websites, articles are labelled according to specific categories; we decided to focus on four macro-categories: *health*, *science*, *economy*, and *environment*. These categories have been selected to cover some of the most frequent - and potentially hazardous - domains where clickbait is usually found. Since the categories varied a lot from website to website, we manually mapped each category into one of the four macro categories under analysis. Two annotators, knowledgeable in the area, were then provided with the headlines and the related articles and were asked to label whether a headline was clickbait. For aiding in this task, we have used as reference the clickbait measure as computed by Arthur et al. [21]. Eventually, given the clickbait dataset, the two annotators were required to extract the gold spoilers from the article's text and to produce the neutralised forms for each headline. To this end, we employed an author reviewer strategy [22]: an LLM (ChatGPT `gpt-3.5-turbo-0125`[3]) was used to generate both the spoilers and the neutralised forms (author component)[4], and the native Italian speaking annotators were asked to manually post-edited the generations (reviewer component).[5] This procedure was proven to be more effective and less time-consuming than writing the data

---

[2]Essere Informati, TGNewsItalia, Voxnews, DirettaNews, Informati, Italia, Jeda News, News Cronaca, TG5Stelle, TG24-ore, ByoBlu, Mag24, WorldNotix, lo sapevi che, Fortementein
[3]https://chat.openai.com
[4]In Appendix A.3 we provide the prompt employed
[5]Details in Appendix A.2

Category	Headline	Article	Clickbait	Spoiler	Neutralised title
Health	Frutto o fiore? gustosissima e attraente, una celebrità sulle nostre tavole, sveliamo chi è	Tutti la conosciamo, immancabile sulle nostre tavole, celebre in tutto il mondo ma misteriosa la sua natura, frutto da gustare o fiore...	True	La fragola	Fragola: gustosissima e attraente, una celebrità sulle nostre tavole
Science	Scoperto un metallo che si auto-ripara. Scienziati sbalorditi	Il recente esperimento ha rivelato un fenomeno straordinario...	True	Il platino	Il metallo che si auto-ripara: il platino
Health	Una malattia che colpisce 500mila persone	Parliamo di una malattia sistemica cronica mediata dal sistema immunitario che interessa...	True	La psoriasi colpisce circa 500 mila persone	La psoriasi: una malattia che colpisce circa 500mila persone in Italia
Environment	Zanzare, ecco come eliminarle senza insetticidi	Con l'arrivo del caldo, anche le zanzare si fanno largo nelle nostre case o nei nostri giardini...	True	Per eliminare una volta per tutte le zanzare dalla vostra casa, dovreste acquistare un pipistrello	Zanzare, ecco come eliminarle senza insetticidi: basta acquistare un pipistrello

Table 1
An excerpt of the presented dataset showing the most relevant fields. Article bodies are shortened for space reasons. Translated text can be found in Table 9 (Appendix B).

from scratch [23]. To assess the amount of post-editing required, we employed *Human-targeted Translation Edit Rate* [HTER; 24]. HTER quantifies the minimum edit distance, which is the least number of editing operations needed, between a machine-generated text and its post-edited counterpart. HTER values exceeding 0.4 indicate low-quality outputs; under such circumstances, rewriting the text from scratch or extensive post-editing would necessitate comparable effort [25].

The obtained HTER results for the spoiler generation (*0.4*) are higher than those computed upon the neutralisation (*0.3*), in par or slightly lower than the 0.4 threshold. The high HTER values, especially for the spoiler annotation, can be attributed to the model's tendency to generate spoilers comprising more details than those necessary to fill the curiosity gap. While in some cases a simple deletion was sufficient, in others the annotator had to rewrite the spoiler almost completely. Regarding the annotation of the neutralisation texts, the higher results are a consequence of the spoiler generation, as the model was required to generate them simultaneously.

With this, we have generated the *golden* set of the dataset, in which all the instances were manually annotated. Further details regarding the dataset creation can be found in Appendix A. To expand this set, we have used a clickbait classifier (see Sect. 4.1) to automatically detect clickbait headlines. This new set of data, automatically annotated, constitutes the *silver* set of our dataset. Several examples of dataset entries are provided in Table 1.

3.2. Dataset Analysis

The complete *ClickBaIT* dataset consists of 4,144 entries. Each entry includes the following fields: (i) **source website**, that specifies the source of the article; (ii) **publication date**, which is captured from the original source; (iii) **headline text**; (iv) **article text**; (v) **original URL**; (vi) **macro category** inferred from the original category extracted from the source; (vii) **image URL** associated with the article as specified in the source; (viii) **clickbait annotation**; (ix) the **associated spoiler**; and (x) the **neutralised version** of the title.

Table 2 shows the main statistics of the final version of the dataset. The golden set is manually annotated and thus contains high-quality information. Additionally, the silver set has been annotated automatically as described and therefore contains a larger number of instances.

To gain a deeper understanding of the content of the dataset we have used Variationist [26], a tool that allows to inspect useful statistics and patterns in textual data. Upon inspection of the data, we have detected several patterns frequently used for generating the curiosity gap.

Of course, one of the most common strategies used in

Set	Clickbait (%)	Non-clickbait (%)	Total
Golden	698 (53%)	629 (47%)	1,327
Silver	1,563 (56%)	1,224 (44%)	2,787
Total	2,261	1,853	4,114

Table 2
Size of the presented dataset, considering both golden and silver sets.

clickbait headlines is the formulation of a question that is later answered in the article, even though sometimes it is not. In the instance "*Quanto è green il gas?*" (*How green is gas?*) the article explains that gas is not considered green. Another frequent strategy we have detected is the introduction to the content of the article, which invites the reader to click it: *Beve un cucchiaio di aceto di mele nell'acqua tutti i giorni, ecco cosa succede* (*Drinks a tablespoon of apple cider vinegar in water every day, this is what happens*).

Another usual pattern is the reference to enumerations, frequently using round and manageable numbers such as 10, 8, and 5. This can be done for introducing numbered content, as in "*Le 10 fantasie femminili più segrete*" (*The 10 most secret female fantasies*), or even to generate a reaction in the reader: "*Hai solo 10 secondi per salvarti. Ecco cosa devi fare:*" (*You only have 10 seconds to save yourself. Here's what you have to do:*). Other means can be used to make headlines noticeable, such as introducing text in all caps, using striking vocabulary or even punctuation marks, as in "*[ALLARME] Truffa AUTO USATE, fate attenzione!*" (*[ALERT] USED CAR scam, beware!*).

See Table 8 (Appendix A.2) for a collection of patterns that have been considered during the manual annotation of the dataset. Besides, Appendix B includes a graphical summary of the dataset, while its interactive version can be accessed online.[6] Details are provided in Appendix C.

4. Experimental Design

The experimental design comprises three steps: clickbait detection, spoiler generation and clickbait neutralisation.

4.1. Clickbait Detection

This is the first and most basic task aimed at addressing the clickbait phenomenon. To explore the effect of using additional data in the training process, we use the Webis-Clickbait-17 [27], an English dataset containing clickbait that is also annotated in a binary fashion.

Following the insights by Araque et al. [28], we use the training on English data to improve the classification of Italian data. The main idea is to harness the availability of large amounts of English data, generating a compound dataset with a lower amount of Italian instances. To do so, a multilingual mixture dataset is created so that 35% of the final dataset comprises Italian instances, while the rest are in English.

We model the detection challenge as a binary classification task: clickbait/non-clickbait. To study the complexity of the task, we explore two different models for classification: (i) a DistilBERT [29] (`distil-base-`

`multilingual-cased`[7]) model trained in a multilingual setting, and (ii) the Llama3-8B language model (`meta-llama/Meta-Llama-3-8B`[8]). The composed dataset has been split into train and test splits, which have been used to fine-tune and evaluate these models, respectively.

To assess the effect of using a mixture of both English and Italian instances in the dataset, we evaluate the performance of the two models in a monolingual setting (e.g., fine-tuning in Italian and predicting in the same language) as well as the multilingual variant (e.g., fine-tuning in English and Italian text, and predicting on Italian instances).

4.2. Spoiler Generation

The spoiler generation task consists in generating a short message that fulfils the curiosity gap present in a given clickbait title, by extracting the information from the linked article. To this end, we tested `LLaMAntino-3-ANITA-8B-Inst-DPO-ITA` (LLaMAntino-3-8B hereafter) [30] on our clickbait dataset. The model was tested both in in-context learning (zero- and few-shot) and fine-tuning settings.

Building on prior research that frames spoiler generation as a Question Answering task [31], we prompt the model to rewrite clickbait headlines as questions and extract the corresponding answers, i.e., the spoilers, from the linked articles.

4.3. Clickbait Neutralisation

The best-performing configuration was employed for the neutralisation of the clickbait headlines. To this end, we instructed the LLM to perform a style transfer task, from a clickbait headline style to a more journalistic one, while integrating the spoiler information into the original headline.

5. Results and Discussion

5.1. Evaluation Metrics

Firstly, for the evaluation of the clickbait detection task we use the macro-averaged precision, recall and f-score. This allows us to assess the performance even in an unbalanced scenario. For the generation tasks, we assessed lexical similarity through ROUGE score [32] and semantic similarity. For the latter, text embeddings, computed using `sentence-bert-base-italian-xxl-uncased`[9], were compared using cosine similarity.

---

[6] https://oaraque.github.io/ClickBaIT/clickbait.html

[7] https://huggingface.co/distilbert-base-multilingual-cased
[8] https://huggingface.co/meta-llama/Meta-Llama-3-8B
[9] https://huggingface.co/nickprock/sentence-bert-base-italian-xxl-uncased

	zero-shot			few-shot			fine-tuning		
	R1	RL	SemSim	R1	RL	SemSim	R1	RL	SemSim
headlines	0.189	0.157	0.567	0.250	0.221	0.667	0.260	0.234	0.659
questions	0.271	0.249	0.645	0.286	0.258	0.630	0.250	0.224	0.646

Table 3
`LLaMAntino-3-8B` results for the spoiler generation task. We report ROUGE 1 and L (R1, RL) and semantic similarity (SemSim).

5.2. Clickbait detection

Table 4 shows the results of the evaluation in the task of clickbait classification. As expected, introducing data instances in English improves the performance in Italian. In the case of classification in Italian, we see a staggering improvement for the Llama3 model of 8.43 points. This further supports previous results [28]. We argue that augmenting the training set with instances in a diverse language is an effective strategy that can be generalised to other tasks.

We also see that the best model for the classification of clickbait is the one obtained with Llama3, trained with both English and Italian data. Hence, we use this model to predict on the silver set of our dataset.

Test	Train	Model	Prec.	Rec.	M-F1
EN	EN	DistilBERT	67.15	70.34	66.94
		Llama3	68.42	66.46	67.18
	EN+IT	DistilBERT	70.28	70.14	70.12
		Llama3	**71.20**	**71.15**	**71.15**
IT	IT	DistilBERT	68,85	70.47	68.65
		Llama3	66.96	67.19	67.07
	EN+IT	DistilBERT	72.87	74.85	71.77
		Llama3	**76.32**	**75.51**	**75.50**

Table 4
Results for Clickbait detection. The 'Test' and 'Train' columns indicate the languages of the test and train sets, respectively.

5.3. Spoiler Generation Results

Results for the spoiler generation task are reported in Table 3. We evaluated the capabilities of `LLaMAntino-3-8B` in both in-context learning scenarios (zero- and few-shot) and through fine-tuning. As inputs, we used clickbait headlines and questions generated by ChatGPT, instructing the model to execute a Question Answering task for the latter. When using headlines as input, few-shot and fine-tuning approaches outperform zero-shot methods. Few-shot approaches demonstrate higher performance in terms of semantic similarity, while fine-tuning exhibits stronger lexical adherence to the source document, as reflected in ROUGE scores. This can be attributed to the

few examples provided in the few-shot approach, which make the model aware of the task while allowing more creative outputs (resulting in lower ROUGE scores). Conversely, the fine-tuned model learned from the training data to adhere more closely to the source article, which comes at the expense of producing semantically richer responses (evidenced by lower SemSim scores).

Interestingly, casting spoiler generation as a question-answering task yields higher results in the zero-shot setting compared to using headlines as input. However, the results for few-shot and fine-tuning scenarios tend to be on par. This can be explained by the fact that headlines may contain multiple gaps that the human-annotated dataset accounted for, but the non-supervised "question generation" module could not fully capture. Generally, this approach leads to sufficiently good results; however, we believe that more attention should be given to the quality of the questions, either through more efficient prompts or with human-generated/curated data.

5.4. Clickbait Neutralisation Results

In Table 5, we report the results for clickbait neutralisation. For this task, we prompted `LLaMAntino-3-8B` with a few-shot approach, employing the spoilers generated with the three configurations of the previous experiments (headlines as input). Using spoilers generated with the fine-tuned models leads to higher results both for lexical and semantic metrics. Interestingly, scores tend to increase when the training complexity of the input data increases. In Table 6 we report examples of headlines along with their generated spoilers (through the fine-tuned model) and their neutralisation.

input data	R1	RL	SemSim
zero-shot	0.250	0.212	0.675
few-shot	0.265	0.223	0.706
fine-tuning	**0.286**	**0.247**	**0.715**

Table 5
Neutralisation generation results. Automatically generated spoilers from the previous experiments were used as input for the few-shot generation of the data. We report ROUGE 1 and L (R1 and RL) and the semantic similarity scores.

Headline	Spoiler	Neutralisation
"Juventus in Serie B": perché c'è panico tra i tifosi, la scoperta delle ultime ore	15 punti di penalizzazione	Juventus in grave difficoltà: 15 punti di penalizzazione e il rischio di cadere in Serie B
Lutto tremendo nello sport italiano, morto giovanissimo dopo un malore	"Samuel Dilas era un giocatore di pallacanestro che militava nel Virtus Lumezzane a Brescia, in Serie B" e "aveva 24 anni" e "era alto 206 centimetri" e "nato a Novellara (Reggio Emilia)" e "aveva un padre di nome Torsen, una madre di nome Chiara e una sorella minore di nome Maia" e "era in convalescenza dopo una polmonite" e "era arrivato alla Virtus Lumezzane nella scorsa stagione".	Tragico decesso del pallacanestrista Samuel Dilas, 24 anni, ex convalescente da polmonite e giocatore della Virtus Lumezzane
Un papà si rifiuta di mangiare accanto a un bambino Down di 5 anni, il cameriere decide di fare questo	Il cameriere ha sbattuto fuori il maleducato padre che voleva essere spostato a causa della presenza di un bambino con sindrome di Down.	Un cameriere espelle un cliente maleducato che chiede di essere spostato per non sedersi accanto a un bambino con sindrome di Down.
E' doloroso e si forma tra le dita dei piedi, ecco come rimuoverlo	"L'occhio di pernice è causato principalmente dalla pressione della scarpa che favorisce la formazione di un'ispessimento di pelle che provoca dolore, in quanto è soggetto all'attrito tra le dita. Per rimuovere l'occhio di pernice è fondamentale ammorbidire prima la zona interessata per poi provare a rimuovere l'ispessimento utilizzando rimedi naturali senza dolore e in modo semplice."	Come rimuovere l'occhio di pernice, un problema di pressione e attrito causato dalle scarpe
La chiamano "LA BOMBA" la miscela che in sole 24-48 ore elimina influenza, raffreddore e tosse	Lo zenzero è un rimedio naturale per il trattamento di tosse, raffreddore e influenza. La miscela limone, zenzero e miele è ideale per alleviare i sintomi delle comuni malattie. Basta prendere 2 o 3 cucchiai della miscela naturale, riempire una tazza con acqua calda e lasciare in infusione per 3 o 4 minuti.	Miscela naturale di limone, zenzero e miele allevia i sintomi di tosse, raffreddore e influenza in pochi giorni.

Table 6
Examples of clickbait headlines, along with the automatically generated spoiler and neutralised version.

6. Conclusion

This work presents *ClickBaIT*, a novel Italian dataset for clickbait modelling, as well as a diverse set of experiments to assess the effectiveness of current models for clickbait detection, spoiling and neutralisation. The dataset includes news articles that have been manually annotated to indicate the presence of clickbait, spoilers associated with clickbait headlines, and their respective neutral headlines.

The experiments explore the effectiveness of current NLP methods for the modelling of clickbait headlines in Italian through *ClickBaIT*. The evaluation for clickbait detection shows how training data can be augmented in a multilingual setting, which leads to classification improvements that are in line with previous research [28]. The generation experiments, for both spoiling and neutralisation, evidence that the evaluated model does benefit from in-domain knowledge extracted from the proposed dataset. As seen, these informed generations are more accurate and align better with the golden text.

Considering the effect of clickbait, we argue that while there are initially harmless articles, lack of accuracy can have a detrimental effect on readers. This is clear when considering certain sensitive domains such as health. Thus, we hope that this work facilitates future research on the topic for example, by addressing the link between clickbait and misinformation, considering both in a unified framework.

Acknowledgments

References

[1] K. Scott, You won't believe what's in this paper! clickbait, relevance and the curiosity gap, Journal of Pragmatics 175 (2021) 53–66. URL: https://www.sciencedirect.com/science/article/pii/S0378216621000229. doi:https://doi.org/10.1016/j.pragma.2020.12.023.

[2] J. N. Blom, K. R. Hansen, Click bait: Forward-reference as lure in online news headlines, Journal of Pragmatics 76 (2015) 87–100. URL: https://www.sciencedirect.com/science/article/pii/S0378216614002410. doi:https://doi.org/10.1016/j.pragma.2014.11.010.

[3] G. Loewenstein, The psychology of curiosity: A review and reinterpretation, Psychological Bulletin 116 (1994) 75–98. doi:10.1037/0033-2909.116.1.75.

[4] K. Scott, R. Jackson, When everything stands out, nothing does, Relevance theory, figuration, and continuity in pragmatics 8 (2020) 167–192.

[5] K. Scott, "deceptive" clickbait headlines: Relevance, intentions, and lies, Journal of Pragmatics 218 (2023) 71–82. URL: https://www.sciencedirect.com/science/article/pii/S0378216623002643. doi:https://doi.org/10.1016/j.pragma.2023.10.004.

[6] S. Zannettou, M. Sirivianos, J. Blackburn, N. Kourtellis, The web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans, J. Data and Information Quality 11 (2019). URL: https://doi.org/10.1145/3309699. doi:10.1145/3309699.

[7] E. Aïmeur, S. Amri, G. Brassard, Fake news, disinformation and misinformation in social media: a review, Social Network Analysis and Mining 13 (2023) 30.

[8] M. Potthast, S. Köpsel, B. Stein, M. Hagen, Clickbait detection, in: Advances in Information Retrieval: 38th European Conference on IR Research, ECIR 2016, Padua, Italy, March 20–23, 2016. Proceedings 38, Springer, 2016, pp. 810–817.

[9] P. Rajapaksha, R. Farahbakhsh, N. Crespi, Bert, xlnet or roberta: The best transfer learning model to detect clickbaits, IEEE Access 9 (2021) 154704–154716. doi:10.1109/ACCESS.2021.3128742.

[10] M. Hagen, M. Fröbe, A. Jurk, M. Potthast, Clickbait spoiling via question answering and passage retrieval, in: S. Muresan, P. Nakov, A. Villavicencio (Eds.), Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 7025–7036. URL: https://aclanthology.org/2022.acl-long.484. doi:10.18653/v1/2022.acl-long.484.

[11] A. Agrawal, Clickbait detection using deep learning, in: 2016 2nd International Conference on Next Generation Computing Technologies (NGCT), 2016, pp. 268–272. doi:10.1109/NGCT.2016.7877426.

[12] S. Kaur, P. Kumar, P. Kumaraguru, Detecting clickbaits using two-phase hybrid cnn-lstm biterm model, Expert Systems with Applications 151 (2020) 113350. URL: https://www.sciencedirect.com/science/article/pii/S0957417420301755. doi:https://doi.org/10.1016/j.eswa.2020.113350.

[13] M. Fröbe, B. Stein, T. Gollub, M. Hagen, M. Potthast, SemEval-2023 task 5: Clickbait spoiling, in: A. K. Ojha, A. S. Doğruöz, G. Da San Martino, H. Tayyar Madabushi, R. Kumar, E. Sartori (Eds.), Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023), Association for Computational Linguistics, Toronto, Canada, 2023, pp. 2275–2286. URL: https://aclanthology.org/2023.semeval-1.312. doi:10.18653/v1/2023.semeval-1.312.

[14] A. K. Ojha, A. S. Doğruöz, G. Da San Martino, H. Tayyar Madabushi, R. Kumar, E. Sartori (Eds.), Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023), Association for Computational Linguistics, Toronto, Canada, 2023. URL: https://aclanthology.org/2023.semeval-1.0.

[15] H. Kurita, I. Ito, H. Funayama, S. Sasaki, S. Moriya, Y. Mengyu, K. Kokuta, R. Hatakeyama, S. Sone, K. Inui, TohokuNLP at SemEval-2023 task 5: Clickbait spoiling via simple Seq2Seq generation and ensembling, in: A. K. Ojha, A. S. Doğruöz, G. Da San Martino, H. Tayyar Madabushi, R. Kumar, E. Sartori (Eds.), Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023), Association for Computational Linguistics, Toronto, Canada, 2023, pp. 1756–1762. URL: https://aclanthology.org/2023.semeval-1.243. doi:10.18653/v1/2023.semeval-1.243.

[16] T. Liu, K. Yu, L. Wang, X. Zhang, H. Zhou, X. Wu, Clickbait detection on wechat: A deep model integrating semantic and syntactic information, Knowledge-Based Systems 245 (2022) 108605. URL: https://www.sciencedirect.com/science/article/pii/S0950705122002714. doi:https://doi.org/10.1016/j.knosys.2022.108605.

[17] Şura Genç, E. Surer, Clickbaittr: Dataset for clickbait detection from turkish news sites and social media with a comparative analysis via machine learning algorithms, Journal of Information Science 49 (2023) 480–499. doi:10.1177/01655515211007746.

[18] A. Geçkil, A. A. Müngen, E. Gündogan, M. Kaya, A clickbait detection method on news sites, in: 2018 IEEE/ACM International Conference on Ad-

vances in Social Networks Analysis and Mining (ASONAM), 2018, pp. 932–937. doi:10.1109/ASONAM.2018.8508452.

[19] C. Oliva, I. Palacio-Marín, L. F. Lago-Fernández, D. Arroyo, Rumor and clickbait detection by combining information divergence measures and deep learning techniques, in: Proceedings of the 17th International Conference on Availability, Reliability and Security, ARES '22, Association for Computing Machinery, New York, NY, USA, 2022. URL: https://doi.org/10.1145/3538969.3543791. doi:10.1145/3538969.3543791.

[20] I. García-Ferrero, B. Altuna, Noticia: A clickbait article summarization dataset in spanish, arXiv preprint arXiv:2404.07611 (2024).

[21] T. E. C. L. Arthur, A. T. Cignarella, S. Frenda, M. Lai, M. A. Stranisci, A. Urbinati, et al., Debunker assistant: a support for detecting online misinformation, in: Proceedings of the Ninth Italian Conference on Computational Linguistics (CLiC-it 2023), volume 3596, Federico Boschetti, Gianluca E. Lebani, Bernardo Magnini, Nicole Novielli, 2023, pp. 1–5.

[22] S. S. Tekiroğlu, Y.-L. Chung, M. Guerini, Generating counter narratives against online hate speech: Data and strategies, in: D. Jurafsky, J. Chai, N. Schluter, J. Tetreault (Eds.), Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 2020, pp. 1177–1190. URL: https://aclanthology.org/2020.acl-main.110. doi:10.18653/v1/2020.acl-main.110.

[23] D. Russo, S. Kaszefski-Yaschuk, J. Staiano, M. Guerini, Countering misinformation via emotional response generation, in: H. Bouamor, J. Pino, K. Bali (Eds.), Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Singapore, 2023, pp. 11476–11492. URL: https://aclanthology.org/2023.emnlp-main.703. doi:10.18653/v1/2023.emnlp-main.703.

[24] M. Snover, B. Dorr, R. Schwartz, L. Micciulla, J. Makhoul, A study of translation edit rate with targeted human annotation, in: Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers, Association for Machine Translation in the Americas, Cambridge, Massachusetts, USA, 2006, pp. 223–231. URL: https://aclanthology.org/2006.amta-papers.25.

[25] M. Turchi, M. Negri, M. Federico, Coping with the subjectivity of human judgements in MT quality estimation, in: Proceedings of the Eighth Workshop on Statistical Machine Translation, Association for Computational Linguistics, Sofia, Bulgaria, 2013, pp. 240–251. URL: https://aclanthology.org/W13-2231.

[26] A. Ramponi, C. Casula, S. Menini, Variationist: Exploring multifaceted variation and bias in written language data, arXiv preprint arxiv:2406.17647 (2024). URL: https://arxiv.org/abs/2406.17647.

[27] M. Potthast, T. Gollub, K. Komlossy, S. Schuster, M. Wiegmann, E. Garces Fernandez, M. Hagen, B. Stein, Crowdsourcing a Large Corpus of Clickbait on Twitter, in: E. Bender, L. Derczynski, P. Isabelle (Eds.), 27th International Conference on Computational Linguistics (COLING 2018), Association for Computational Linguistics, 2018, pp. 1498–1507. URL: https://aclanthology.org/C18-1127/.

[28] O. Araque, M. F. L. Corniel, K. Kalimeri, Towards a multilingual system for vaccine hesitancy using a data mixture approach., in: Proceedings of the 9th Italian Conference on Computational Linguistics, 2023.

[29] V. Sanh, L. Debut, J. Chaumond, T. Wolf, Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter, arXiv preprint arXiv:1910.01108 (2019).

[30] M. Polignano, P. Basile, G. Semeraro, Advanced natural-based interaction for the italian language: Llamantino-3-anita, 2024. arXiv:2405.07101.

[31] M. Woźny, M. Lango, Generating clickbait spoilers with an ensemble of large language models, arXiv preprint arXiv:2405.16284 (2024).

[32] C.-Y. Lin, Rouge: A package for automatic evaluation of summaries, in: Text summarization branches out, 2004, pp. 74–81.

scienza	insetti, animali, AI, scienza, smartphone, Spazio, tecnologia, TECNOLOGIE, SCIENZA, ufo, biochimica, eclissi, bomba atomica, terra piatta, idroelettrico, temperatura, coltivazione, robot, fisica quantistica, macchie solari, ricerca, vulcano, titanio, universo, fotovoltaico, intelligenza, iPhone, hacker, microonde, motori di ricerca, onde elettromagnetiche, tecnologia, sole, scienza, radioterapia, pesticidi, armi chimiche, comete, case farmaceutiche, psichiatria, smartphone, formiche, elettrodomestici, solare, macrobiologi, mondo, lampadine a basso consumo, tecnologia, scienze-e-tech, scienza, scienza, innovazione, scienza, tecnologia-2, animali intelligenti, funzione cognitiva, microchip, cani, samsung, wi fi, tecnologia-e-tv, SCIENZE, TECNOLOGIA, bioetica, biologia, fisica, covid, coronavirus
salute	Salute, CORONAVIRUS, VAIOLO SCIMMIE, TUBERCOLOSI, SALUTE, SCABBIA, AIDS, salute, hiv, cocaina, antidepressivi, veleni, infezioni, carne, tabacco, infibulazione, fluoro, alcool, alimentari, aids, antibatterico, dieta, insetticida, cibo, benessere, farmaci, digitopressione, caffè, sigarette, ministero della salute, autismo, limoni, cure naturali, paracetamolo, cancro, antiossidante, droga, olio, medicina alternativa, fragole, vegetariano, eroina, dislessia, veleno, zenzero, virus, psicologia, biologico, magnesio, frutta, psicofarmaci, pollo al cloro, fiori di bach, medico, sonno, birra, vitamina e, ulivi, proteine, stress, banana, pensieri negativi, tumori, benzodiazepine, latte, miele, cuore, epilessia, longevità, marijuana, diabete, sale, ibernazione, vecchiaia, fegato, vegan, prevenzione, dentifricio, cervello, sistema immunitario, sodio, suicidio, rimedi naturali, maltempo, canapa, pillola, mal di gola, depressione, psiche, alimentazione, ebola, aspartame, dentifricio senza fluoro, tiroide, mangiare, cure proibite, Alzheimer, smog, gas, malattie, calamità, mammografia, verdura, aloe, masticazione, farmaco, igiene, batteri, medicina, vitamina c, epatite c, forfora, energia, vaccini, ormoni, flora batterica, sorbitolo, antibiotici, piedi, obesità, arsenico, cortisolo, chemioterapia, contraccezione, Neurotrasmettitori, semi, melograno, celiachia, Coca cola, salute-benessere, salute, salute-e-benessere, bellezza, dimagrante, benessere, salute-benessere, rimedi-naturali, pianeta-mamma, grano antico, acqua ossigenata, alimetnazione, ansia, dentisti, curcuma, casa-e-cucina, hobby-e-sport, SPORT, crescita-consapevolezza, la-salute-che-viene, sport, stile-di-vita, consigli, lifestyle, pomodori
ambiente	Cambiamenti climatici, energia, energia elettrica, Natura, AMBIENTE, ECOLOGIA, global warming, geoingegneria, alberi, pianeta terra, natura, inquinamento, mare, terra, manipolazione climatica, clima, rinnovabili, Dissesto idrogeologico, ecologia, ambiente, green, ambiente-attuale, ecologia, salute-benessere, natura, ambiente, METEO, tempesta solare, astronomia, acido
economia	affari-online, economia, ECONOMIA, consumi-risparmi, microchip r-fid, bollo auto, tasso d'interesse, finanza, bollette, banche, profitto, spese, economia-finanza, economia, economia, economia-dellanima, fisco-e-tasse, economia, economia, economia, economia-e-finanza

Table 7
Split of the categories into the four macro-categories.

A. Dataset Creation Details

A.1. Category Assessment

In Table 7 we report how the heterogeneous categories scraped directly from the misleading websites were divided into the four macro-category of *scienza* (science), *salute* (well-being), *ambiente* (environment), *economia* (economy).

A.2. Annotation Guidelines

Three components of our datasets were subject to human intervention to: (i) determine if the headline was clickbait, (ii) identify the related article's spoiler, that is, the information required to satisfy the curiosity gap within the headline, and (iii) revise the headline to include the spoiler information, thereby neutralizing it. During all three annotation stages, we employed a machine-human collaboration to expedite the work of annotators. The an-

notators received both a score indicating how much the headline was clickbait and automatic ChatGPT gpt-3.5-turbo-0125 generated suggestions for the spoilers and the neutralized versions of the headlines. Below, we have outlined the annotation guidelines that the annotators were to follow.

Clickbait labelling In order to select the clickbait headlines present in the scraped data, the annotators were provided with specific guidelines. Table 8 provides the main key points taken into consideration in order to label the data.

Spoiler post-editing For the post-editing of the spoiler the annotator was required to spot in the headline the information gap and to check if the generated spoiler was providing that information checking the related article. If the model failed to find the proper spoiler, the annotator had to rewrite it sticking as much as possible to

Characteristic	Original example (IT)	Translated example (EN)
Lack of essential information, i.e., the subject the article is talking about	*"Ora riposa in pace". Calcio in lutto, morto uno dei grandi protagonisti dell'Italia*	*"Now rest in peace". Football in mourning, one of Italy's great protagonists dead*
Sensationalist tone	*Fan ubriaca le salta addosso sul palco. La sua reazione è incredibile e sconvolge tutti i presenti*	*Drunk fan jumps on her on stage. Her reaction is incredible and shocks everyone present*
Questions raised but answered in the article body	*Tratti della nostra colonna: quali sono? Come evitare lesioni?*	*Traits of our column: what are they? How to avoid injuries?*
Enumeration of elements	*10 cibi per sbarazzarsi del gonfiore di stomaco e pancia*	*10 foods to get rid of bloated stomach and tummy*
Use of capitalization	*INFARTO: sopravvivere quando si è soli. Hai solo 10 secondi per salvarti. Ecco cosa devi fare:*	*HEART ATTACK: surviving when alone. You only have 10 seconds to save yourself. Here's what you have to do:*
Introduction of the content without actually giving the information	*Zanzare, ecco come eliminarle senza insetti-cidi*	*Mosquitoes, this is how to eliminate them without insecticide*
Use of quotations that do not give information	*Omicron, Ilaria Capua: "Ecco perché i vacci-nati si infettano di più rispetto a prima"*	*Omicron, Ilaria Capua: "This is why the vacci-nated get more infected than before"*

Table 8
Key points used for the annotation of the dataset. Please note that some instances can exemplify more than one point.

the document's text. If the spoiler was correct but added extra info, the annotator had to keep those extra information only if those were essential for having a complete headline. If the spoiler was correct, then the annotator could leave it as it was.

Neutralised Clickbait Post-Editing   The annotator was required to check if the neutralised forms comprises both the headline and the spoiler information. If the spoiler was very long (e.g., long listing), then the annotator had to summarise the spoiler as much as possible aiming to embed in the final novel headline enough information to reduce or remove the information gap. If the model failed at addressing the spoiler information in the neutralised version of the headline, then the annotator had to manually add it. Moreover, the annotator was required to remove sensationalist tones as much as possible, if this tone was still creating useless curiosity in the reader.

A.3. Author Component Instruction

Hereafter, we provide the instruction employed to automatically generate spoilers and the neutralised versions of the clickbait headlines through ChaGPT `gpt-3.5-turbo-0125`.

> I have a clickbait headline and its corresponding article, both written in Italian.

The clickbait headline typically omits key information to create a curiosity gap for the reader. Your task is to extract this missing information, known as a "spoiler," from the article's text. The spoiler can be a single keyword, a short text passage, or a list of keywords. Once you have identified the spoiler, rewrite the clickbait headline by incorporating this information to eliminate the curiosity gap. The output must be in JSON format and written in Italian. The JSON should include two entries: one called "spoiler" that contains the extracted spoiler(s), and another called "new_headline" that has the revised headline.

Example Input:

Clickbait headline: "Questo attore ha fatto qualcosa di incredibile sul set di un famoso film!" Article: "Durante le riprese del film 'Il Gladiatore', l'attore Russell Crowe ha deciso di fare un gesto di grande generosità donando una parte significativa del suo stipendio al fondo per i membri della troupe."

Example Output:

{"spoiler": "Russell Crowe ha donato una parte significativa del suo stipen-
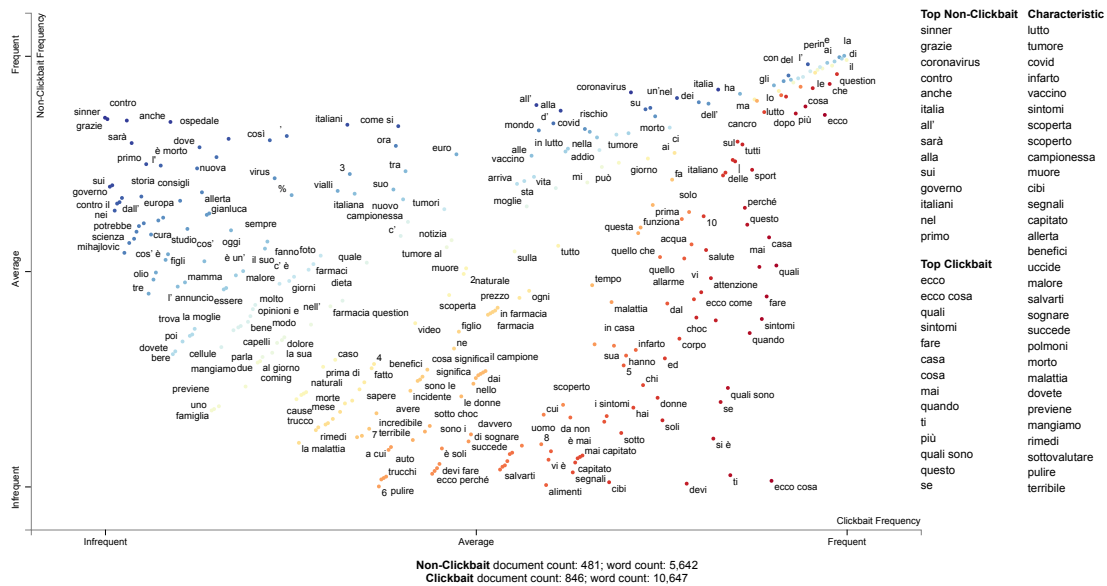
Figure 2: Frequency of words for both clickbait and non-clickbait categories. On the right, most frequent words for each class, and both (Characteristic). An interactive version of the graph can be accessed at the following link https://oaraque.github.io/clickIT/clickbait.html

dio al fondo per i membri della troupe",
"new_headline": "Russell Crowe ha fatto qualcosa di incredibile sul set di 'Il Gladiatore': ha donato una parte significativa del suo stipendio al fondo per i membri della troupe"}

Please ensure the output is formatted in JSON as specified and that all content is in Italian.

Now do it for the following headline.

Clickbait headline: "{headline}"

Article:"{article}"

B. Additional Dataset Details

B.1. Dataset Visualisation

Figure 2 shows a frequency-based visualization of the dataset. It considers the frequency of appearance of relevant uni and bi-grams for both the clickbait and non-clickbait categories. The figure shows common strategies that are frequent in clickbait content, such as the use of "ecco cosa" (*this is what*) or "quali sono" (*what are*) that can be seen in the lower right part.

B.2. Dataset Excerpt Translation

Table 9 includes the English translations for the Italian examples presented in Table 1.

C. Experimental Design Details

C.1. Question Generation

Questions were generated with ChatGPT `gpt-3.5-turbo-0125` using the following prompt:

You will be provided with a clickbait headline written in Italian. Your task is to generate a question that addresses any missing or vague information in the headline. Here are some examples:

Headline: Si chiama la benedizione di Dio: rimuove l'alta pressione, il diabete e il grasso nel sangue Question: Che cosa viene chiamato 'benedizione di Dio'?

Headline: "Emorragia cerebrale". Italia in apprensione per il suo campione: ricoverato in condizioni gravissime

Question: Chi è il campione?

Please generate the question in Italian, ensuring it seeks to clarify the ambiguous or incomplete details present in the headline.

Category	Headline	Article	Clickbait	Spoiler	Neutralised title
Health	Fruit or flower? Tasty and attractive, a celebrity on our tables, we reveal who she is	We all know it, inevitable on our tables, world-famous, but mysterious is its nature, fruit to enjoy or flower to decorate?	True	The strawberry	Strawberry: tasty and attractive, a celebrity on our tables
Science	Self-repairing metal discovered. Scientists astounded	The recent experiment revealed an extraordinary phenomenon...	True	Platinum	The metal that repairs itself: platinum
Health	A disease that affects 500,000 people	We are talking about a chronic immune-mediated systemic disease that affects about 1.8 million patients...	True	Psoriasis affects about 500,000 people	Psoriasis: a disease that affects about 500,000 people in Italy
Environment	Mosquitoes, here's how to get rid of them without insecticides	With the arrival of hot weather, mosquitoes also make their way into our homes or gardens...	True	To eliminate mosquitoes from your home once and for all, you should buy a bat	Mosquitoes, here's how to get rid of them without insecticides: just buy a bat

Table 9
Translated from the original Italian. An excerpt of the presented dataset showing the most relevant fields. Article bodies are shortened for space reasons.

C.2. Spoiler Generation

For the zero-shot spoiler generation task we employed the following prompt:

> Ti verranno forniti un titolo clickbait e il suo articolo corrispondente. Il titolo clickbait di solito omette, o non esplicita, informazioni chiave per creare curiosità nel lettore. Estrai dall'articolo le informazioni mancanti o vaghe nel titolo che servono per colmare questa curiosità. La risposta può essere un messaggio estremamente coinciso oppure un elenco. Formatta la risposta nel seguente modo. "Risposta: <output>"
>
> Titolo: {headline}
>
> Articolo: {article}

The same instruction was employed with the fine-tuned model. For few-shot generation of the spoiler, we enriched the instruction with two examples.

When casting spoiler generation as a Question Answering task, the following instruction was employed:

> Ti verrà fornita una domanda e un documento. Trova nel documento le informazioni per rispondere alla domanda. La risposta può essere un messaggio conciso oppure un elenco. Formatta la risposta nel seguente modo. "Risposta: <output>"

C.3. Fine-Tuning Details

The `LLaMAntino-3-8B` [30] model underwent training on a single Ampere A40 GPU with 48GB of memory, employing the QLoRA strategy with a low-rank approximation of 64, a low-rank adaptation of 16, and a dropout rate of 0.1. It was set to evaluate every 50 steps, with a batch size of 4, across 3 epochs, using a learning rate of 10^{-4}.

In the clickbait detection experiments, the `DistilBERT` and `Llama3-8b` models have been fine-tuned on the same GPU. The DistilBERT model has been trained on 10 epochs with a learning rate of $2 \cdot 10^{-4}$. For the Llama3 model, we have used QLoRa with the same characteristics as described above, trained on two epochs, with a learning rate of $2 \cdot 10^{-4}$.

C.4. Neutralised Clickbait Generation

The following system prompt (enriched with three examples) has been utilised with LLaMAntino-3-8B:

> Ti verrano forniti due testi: un titolo clickbait e un testo, chiamato spoiler, che contiene le informazioni mancanti nel titolo. Il tuo compito è di riscrivere il titolo clickbait integrando le informazioni dello spoiler. Il nuovo titolo deve essere informativo, privo di toni sensazionalistici, e breve. Se Lo spoiler contine tante informazioni, puoi riassumerle in concetti più generali.
>
> Titolo: {headline}
>
> Spoiler: {spoiler}

D. Ethical Statement

No specific ethical conflicts have been reported during the development of this work. The dataset was compiled from publicly available sources. It is important to acknowledge that the examples in this document are not indicative of the authors' opinions or beliefs. Additionally, the ideas or assertions contained within these texts may be misleading or harmful; therefore, the dataset should be utilized strictly for research purposes.