What Makes You CLIC: Detection of Croatian Clickbait Headlines

Marija Anđelić   Dominik Šipek   Laura Majer   Jan Šnajder
University of Zagreb, Faculty of Electrical Engineering and Computing
TakeLab
{marija.andjelic, dominik.sipek, laura.majer, jan.snajder}@fer.hr

Abstract

Online news outlets operate predominantly on an advertising-based revenue model, compelling journalists to create headlines that are often scandalous, intriguing, and provocative – commonly referred to as *clickbait*. Automatic detection of clickbait headlines is essential for preserving information quality and reader trust in digital media and requires both contextual understanding and world knowledge. For this task, particularly in less-resourced languages, it remains unclear whether fine-tuned methods or in-context learning (ICL) yield better results. In this paper, we compile CLIC, a novel dataset for clickbait detection of Croatian news headlines spanning a 20-year period and encompassing mainstream and fringe outlets. We fine-tune the BERTić model on this task and compare its performance to LLM-based ICL methods with prompts both in Croatian and English. Finally, we analyze the linguistic properties of clickbait. We find that nearly half of the analyzed headlines contain clickbait, and that finetuned models deliver better results than general LLMs.

1 Introduction

Attention-grabbing headlines, a tactic dating back to the printed press, help publishers stand out from the competition (Alves et al., 2016). "Clickbait", a term defined in 2006, describes content deliberately designed to entice clicks (Grammarist, 2023) in the online landscape. Despite driving traffic, clickbait generates predominantly negative audience perception, fostering distrust toward publishers employing such tactics (Blom and Hansen, 2015), its manipulative nature even linking clickbait with fake news (Karadzhov et al., 2017). It exhibits a complex relationship with sentiment (Chakraborty et al., 2017), characterized by hyperbolic positive terminology suggesting strategic emotional manipulation rather than genuine communication. Driven by the negative perception of clickbait, clickbait detection

– the task of automated detection of misleading or sensationalized headlines designed to attract attention – garnered considerable interest within the NLP community. The task progressed from feature-based linguistic approaches (Potthast et al., 2016) to neural architectures (Agrawal, 2016), with transformer-based models demonstrating substantial performance improvements (Zhu et al., 2023). While primarily conducted in English, research exists for less-resourced languages including Italian (Russo et al., 2024), Hungarian (Vincze and Szabó, 2020), Romanian (Gînga and Uban, 2024), and Bulgarian (Karadzhov et al., 2017).

The widespread use of Large Language Models (LLMs), especially using in-context learning (ICL), makes these models a reasonable candidate for clickbait detection. However, LLMs demonstrate notable performance gaps for less-resourced languages across multiple tasks (Rigouts Terryn and de Lhoneux, 2024; Li et al., 2024), where considerably smaller Transformer models pre-trained on a specific language might outperform them (Ljubešić and Lauc, 2021).

In this paper, we address the task of clickbait detection in the Croatian language. We introduce CLIC (**C**lickbait **L**anguage **I**dentification in **C**roatian), a novel human-annotated dataset for the task of clickbait detection. We then train a range of standard ML classifiers and Transformer-based models, and compare their performance with zero- and few-shot LLMs on this task. Additionally, we analyze linguistic features and model failures to provide deeper insight into the clickbait phenomenon.

Our work contributes a valuable new resource for a South Slavic language and deepens understanding of clickbait characteristics in Croatian media. By offering both practical detection methods and novel resources, this work makes way for clickbait neutralization in the Croatian online landscape.

2 Related Work

Clickbait detection datasets are available in different languages, with English being the most represented (Chakraborty et al., 2016; Potthast et al., 2018; Fröbe et al., 2023). Less-resourced languages are also represented. Russo et al. (2024) constructed an Italian corpus of articles from websites known for sensationalist reporting, Vincze and Szabó (2020) created a small corpus downloaded from the Hungarian regional news portals, whereas Gînga and Uban (2024) created an annotated corpus of 10867 articles from the scientific and tech websites published on the Romanian web. To the best of our knowledge, the only existing datasets for Slavic languages are the datasets for Bulgarian (Karadzhov et al., 2017) and Russian (Apresjan and Orlov, 2022).

Prior work has also identified linguistic patterns linked to clickbait. Chakraborty et al. (2016) found that non-clickbait headlines are generally shorter than clickbait headlines, while Biyani et al. (2016) found that clickbait headlines more often contain uppercase letters, question marks, quotes, exclamations, and other unusual writing patterns, suggesting they are intentionally crafted to appear more attention-grabbing.

Traditional ML methods have proven effective in clickbait detection tasks (Bronakowski et al., 2023; Gînga and Uban, 2024; Chakraborty et al., 2016; Fröbe et al., 2023), demonstrating how clickbait relies on clear linguistic features. Fine-tuned Transformer models, however, achieve superior performance, reaching F1 scores of up to 0.89 (Gînga and Uban, 2024; Fröbe et al., 2023; Indurthi et al., 2020).

LLMs have demonstrated remarkable capabilities across NLP tasks, including clickbait detection as shown by Zhu et al. (2023), who achieved state-of-the-art results using zero-shot and few-shot approaches. However, to our knowledge, none of those techniques have been applied to Croatian or other South Slavic languages.

3 Dataset

Since no suitable resource existed, we created a dataset of Croatian news headlines using TakeLab Retriever (Dukić et al., 2024), an AI-driven search engine and database for Croatian news outlets. This tool covers both mainstream and fringe outlets, ensuring a diverse representation. We used the Take-Lab retriever to sample data from 2000 to 2024,

and unlike previous studies that specifically targeted clickbait-heavy sources, our approach sampled broadly across the entire Croatian web – covering 32 outlets. A total of 5000 headlines were collected by extracting 200 headlines published on 25 randomly selected dates.

Eight volunteers carried out the annotation across multiple rounds, with each headline reviewed by five annotators to avoid ties. To better reflect the real-world scenario, where news consumers are often influenced by clickbait titles alone, annotators evaluated headlines without access to the full article. This approach differs from previous studies (Chakraborty et al., 2016; Gînga and Uban, 2024; Vincze and Szabó, 2020), which provided the annotators with full article content with the goal of clickbait neutralization. We employed a binary classification scheme with two labels (*clickbait* and *not clickbait*), along with an additional *invalid* label to allow annotators to filter out artifacts captured during web scraping, such as navigational items, advertising content and metadata. At the end of the annotation process, due to time constraints and volunteer availability, only 3000 randomly selected headlines were selected from a total of 30 news portals out of the original 5000. The resulting annotated dataset is made publicly available in its entirety.[1]

Out of the 3,000 total annotated headlines (examples shown in Table 4), 77 were labeled as *invalid*, and 16 were duplicates. This resulted in a final dataset of 2,907 annotated headlines, with 1,536 labeled *clickbait* (52.84%) and 1,371 (47.16%) as *not clickbait*, making the corpora relatively balanced and in accordance with previous work Gînga and Uban (2024); Chakraborty et al. (2016). The annotators achieved an inter-annotator agreement of 0.53 using the Fleiss-kappa score, categorized as moderate agreement. This confirms the subjective nature of clickbait detection, both in its linguistic ambiguity and in how readers perceive manipulative techniques.

Observing the clickbait distribution over the years, the clickbait to non-clickbait headline ratio is consistent at around 50% (Fig. 2), indicating no significant increase in clickbait on our sample.

---

[1]Dataset available at: `https://takelab.fer.hr/data/clic`

Model	Accuracy	F1
Logistic regression	0.63	0.67
Simple neural network	0.59	0.61
SVM	0.63	0.66
BERT-base-multilingual	0.72	0.71
BERTić	0.78	0.78

Table 1: Performance comparison of various models

	Model	Naive	D	E	F	DE	DF	FE	DFE
English	Mistral 7B	0.66	0.63	0.69	0.43	0.71	0.46	0.73	0.72
	Phi 3.5	0.70	0.60	0.56	0.52	0.59	0.45	0.62	0.57
	gemma 7b	0.70	0.70	0.67	0.70	0.69	0.70	0.71	0.69
	gpt-4.1	0.60	0.43	0.68	0.36	0.46	0.30	0.50	0.41
Croatian	Mistral 7B	0.42	0.64	0.25	0.61	0.53	0.58	0.60	0.55
	Phi 3.5	0.44	0.49	0.44	0.66	0.48	0.63	0.59	0.59
	gemma 7b	0.70	0.50	0.62	0.67	0.59	0.57	0.62	0.67
	gpt-4.1	0.63	0.54	0.59	0.34	0.49	0.41	0.50	0.45

Table 2: F1 scores for various configurations (best in green); D = definition, F = features, E = explanation

4 Clickbait Detection

Baselines. As baselines, we use several standard ML models, all utilizing TF-IDF vectorization of headlines as input features: logistic regression, SVM with a linear kernel (both optimized via GridSearchCV), and a simple neural network with two hidden layers using ReLU activation and sigmoid output layer for binary classification. For all experiments, we use an 80-10-10 split for training, validation, and testing, respectively. We also report the majority class baseline (0.55), representing the accuracy achievable by predicting the most frequent class in our test set.

Fine-tuned Transformers. We finetune BERT (Devlin et al., 2019) and BERTić (Ljubešić and Lauc, 2021), an Electra-based Transformer model pre-trained on south-Slavic languages, including Croatian.

LLMs. For our experiments, we use ICL with both zero-shot and few-shot prompting for various LLMs, including the closed-source GPT 4.1, and smaller open-source models Gemma7B, Phi-3.5-mini, and Mistral7B-Instruct.

For prompting the models, we use various configurations based on the amount of instructions and demonstrations added to the naive prompt (containing just the instruction to classify clickbait). To do so, we draft prompt components: clickbait definition (D), features of clickbait articles (F), and few-shot examples (E) containing both clear and ambiguous cases. We construct combinations of the D, F, and E components to be able to isolate which component attributes mostly to model performance. Also, we translate the prompts to Croatian to compare whether the model performance is higher using English prompts or using the same language as the examples are in. The full prompts are available in Appendix 5.

Results. Table 1 shows the classification performance of all implemented models. All tested ML methods outperform both the majority and random baselines, confirming that clickbait detection fundamentally relies on linguistic patterns. While our baseline models rely solely on TF-IDF vectorization for feature extraction, without the sophisticated linguistic feature engineering employed in prior work (e.g., Potthast et al. (2016) and Chakraborty et al. (2017) used structural, lexical, and syntactic features), we still achieve reasonable performance. This suggests that even simple lexical features can capture many of the distinguishing patterns of clickbait in Croatian.

The fine-tuned Transformer models achieved F1 scores of 71.4% (BERT) and 77.6% (BERTić) respectively, with BERTić expectedly coming out on top. These results coincide with the findings of previous authors.

The results for LLMs with English prompts are shown in Table 2. All tested LLMs outperform both the majority and random baselines, and perform comparably to the ML baseline. However, compared to the results of a fine-tuned BERTić model, all of the LLMs underperform. Gemma7B is the most consistent out of the group with regards to prompt design and incorporation of clickbait features. In all models, the inclusion of clickbait features has been shown to have the most drastic effect on the model's eventual performance, and combining features with few-shot examples produces the highest performance. The importance of examples can be seen when examining the poor performance of a combination of only definition and features (DF), which underscores the critical role of few-shot learning approaches in this task. All of the models show a significant discrepancy between results for prompts in Croatian and English. An analysis of precision versus recall reveals a clear difference between languages. When prompted in English, models tend to be more liberal, favoring recall, whereas prompts in other languages lead models to prioritize precision.

118

Clickbait headlines		Non-clickbait headlines	
Part of speech	**Ratio**	**Part of speech**	**Ratio**
Interjection	∞	Proper noun	1.54×
Determiner	2.32×	Symbol	1.44×
Particle	2.14×	Number	1.31×
Pronoun	1.90×	Other	1.29×
Auxiliary verb	1.53×	Adposition	1.19×

Table 3: Relative frequency of part-of-speech tags in clickbait vs. non-clickbait headlines. Values represent the frequency with which each tag appears in its respective category.

5 Analysis

Feature Analysis. We conducted computational and linguistic analyses to identify the linguistic patterns that distinguish clickbait from legitimate headlines and understand our models' detection mechanisms. Initial attention masking and saliency analysis failed to yield coherent patterns shown in Table 6, with random words receiving high importance scores. Therefore, we employed Part of Speech (PoS) tagging, revealing clear statistical differences between clickbait and non-clickbait content. Results in Table 3 show pronouns, auxiliary words, and determiners are significantly overrepresented in clickbait articles ($p<0.05$, chi-squared tests), aligning with findings by Gînga and Uban (2024). Interjections appear exclusively in clickbait. Chi-squared tests confirm statistical significance for most differences, with determiners, particles, and pronouns showing the strongest clickbait association. Conversely, non-clickbait articles contain significantly more proper nouns and numbers.

Dataset Cartography. The dataset cartography method, proposed by Swayamdipta et al. (2020), enables the identification of hard-to-learn instances during training, as well as ambiguous and potentially mislabeled instances. Figure 1 shows the dataset cartography for the fine-tuned BERTić model. By analyzing confidence, variability, and correctness across training epochs, we gain insight into training dynamics, and cluster the train-set instances into regions – *easy-to-learn*, *ambiguous*, *misleading*, and *hard-to-learn*.

In cases where variability is high and confidence is around 0.5, we identify the space of *ambiguous* instances. Upon inspection, those instances are characterized by balanced clickbait elements, contextual dependency, and mixed use of language devices. The *misleading* region represents instances with low variability and medium confidence. These instances mostly lack classic clickbait markers and
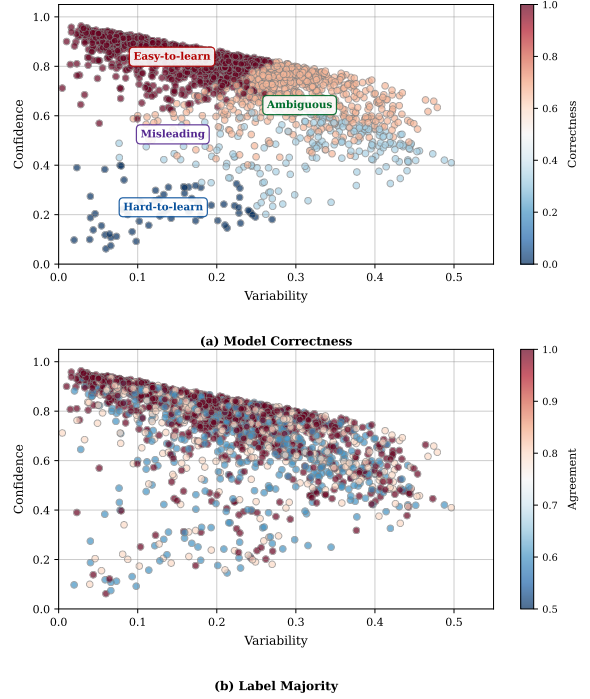

(a) Model Correctness

(b) Label Majority

Figure 1: Dataset cartography visualization showing the classification of headlines by learnability. (a) Model correctness indicates how well the model performs on various examples. (b) Label majority displays the level of agreement among annotations.

are mostly short, direct statements.

Comparison with Annotators. Since clickbait is a subjective classification task, where instances with lower agreement could indicate complex or ambiguous instances, we compare BERTić and LLM performance with annotator agreement levels. For BERTić, we again utilize dataset cartography, but instead of the *correctness* dimension, we display the majority proportion for the given instance. Figure 1(b) shows no distinct regions. This suggests a decoupling between human label variation and model difficulty, meaning instances that humans find ambiguous are not necessarily the same ones that challenge the model.

Error analysis. For a qualitative error analysis of LLMs, we construct an intersection of misclassified instances across prompt variants, then analyze the resulting subset. Examples are shown in Table 4. For non-clickbait instances consistently labeled as clickbait, we find that they mostly consist of sensationalist and dramatic expressions used in factual headlines – exclamation marks, quotation marks, quotes, all caps, numbers – which are features mostly appearing in clickbait instances.

Title	True	Pred
Senzacija: Niko Kovač novi trener Bayerna	✗	✓
Pametna kuna bira najbolju poslovnu ideju! Prijavite se!	✗	✓
Irak: U 24 sata ubijena petorica američkih vojnika	✗	✓
Lalovac ne pada daleko od Linića	✓	✗
Urednik mu nije trebao. Znao je sve o pisanju	✓	✗
Sin na listi HDZ-a, a otac u izbornom povjerenstvu	✓	✗

Table 4: Examples of misclassified headlines. ✓= click-bait, ✗= not clickbait

For instances that annotators labeled clickbait, but consistently classified as non-clickbait, we find no sensational or shocking expressions, the tone is neutral and serious, but the full context is missing. Concealing crucial information is a clear characteristic of clickbait articles, leading to these errors.

6 Conclusion

In this paper, we addressed the task of clickbait detection for Croatian with a new dataset. We evaluated LLM-based ICL methods and fine-tuned BERTić against traditional ML approaches. Our experiments show that fine-tuned BERTić achieves the best overall performance, while LLMs improve when clickbait features are explicitly included in prompts.

7 Limitations and Risks

Limitations. Our current approach focuses solely on headline annotation for clickbait detection. This method, while effective for initial classification, does not allow for an assessment of the semantic relationship between headlines and their corresponding article content. Consequently, headlines that are sensationalized but ultimately accurate might be misclassified. Additionally, our findings are based on Croatian-language data. Therefore, the generalizability of these results to other languages or cultural contexts, which may have distinct clickbait conventions, remains to be explored.

Risks. The linguistic patterns identified in this work could enable more sophisticated clickbait generation that evades detection systems, potentially exacerbating rather than mitigating the problem. Another possible risk is that, in the event of deploying our models, there is a risk of misclassifying legitimate news as clickbait, potentially suppressing real journalism. If any biases are present, they may be amplified in the automatic content filtering process.

References

Amol Agrawal. 2016. Clickbait detection using deep learning. In *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*, pages 268–272.

Liliana Alves, Nuno Antunes, Olga Agrici, Carlos Sousa, and Célia Ramos. 2016. Click Bait: You Won't Believe What Happens Next! *Journal of Social, Technological and Environmental Science*, 5:196–213.

Valentina Apresjan and Alexander Orlov. 2022. Pragmatic mechanisms of manipulation in Russian online media: How clickbait works (or does not). *Journal of Pragmatics*, 195:91–108.

Prakhar Biyani, Kostas Tsioutsiouliklis, and John Blackmer. 2016. "8 Amazing Secrets for Getting More Clicks": Detecting Clickbaits in News Streams Using Article Informality. *Proceedings of the AAAI Conference on Artificial Intelligence*, 30.

Jonas Nygaard Blom and Kenneth Reinecke Hansen. 2015. Click bait: Forward-reference as lure in online news headlines. *Journal of Pragmatics*, 76:87–100.

Mark Bronakowski, Mahmood Al-khassaweneh, and Ali Al Bataineh. 2023. Automatic Detection of Clickbait Headlines Using Semantic Analysis and Machine Learning Techniques. *Applied Sciences*, 13(4).

Abhijnan Chakraborty, Bhargavi Paranjape, Sourya Kakarla, and Niloy Ganguly. 2016. Stop Clickbait: Detecting and Preventing Clickbaits in Online News Media. *CoRR*, abs/1610.09786.

Abhijnan Chakraborty, Rajdeep Sarkar, Ayushi Mrigen, and Niloy Ganguly. 2017. Tabloids in the Era of Social Media? Understanding the Production and Consumption of Clickbaits in Twitter.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

David Dukić, Marin Petričević, Sven Ćurković, and Jan Šnajder. 2024. Takelab retriever: AI-Driven Search Engine for Articles from Croatian News Outlets.

Maik Fröbe, Benno Stein, Tim Gollub, Matthias Hagen, and Martin Potthast. 2023. SemEval-2023 Task 5: Clickbait Spoiling. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 2275–2286, Toronto, Canada. Association for Computational Linguistics.

Raluca-Andreea Gînga and Ana Sabina Uban. 2024. SciTechBaitRO: ClickBait Detection for Romanian Science and Technology News. In *Proceedings of the Third Workshop on NLP for Positive Impact*, pages 188–201, Miami, Florida, USA. Association for Computational Linguistics.

Grammarist. 2023. Clickbait. Accessed: 2025-04-29.

Vijayasaradhi Indurthi, Bakhtiyar Syed, Manish Gupta, and Vasudeva Varma. 2020. Predicting clickbait strength in online social media. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4835–4846, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Georgi Karadzhov, Pepa Gencheva, Preslav Nakov, and Ivan Koychev. 2017. We built a fake news / click bait filter: What happened next will blow your mind! In *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017*, pages 334–343, Varna, Bulgaria. INCOMA Ltd.

Zihao Li, Yucheng Shi, Zirui Liu, Fan Yang, Ali Payani, Ninghao Liu, and Mengnan Du. 2024. Language Ranker: A Metric for Quantifying LLM Performance Across High and Low-Resource Languages.

Nikola Ljubešić and Davor Lauc. 2021. BERTić - the transformer language model for Bosnian, Croatian, Montenegrin and Serbian. In *Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing*, pages 37–42, Kiyv, Ukraine. Association for Computational Linguistics.

Martin Potthast, Tim Gollub, Kristof Komlossy, Sebastian Schuster, Matti Wiegmann, Erika Patricia Garces Fernandez, Matthias Hagen, and Benno Stein. 2018. Crowdsourcing a large corpus of clickbait on Twitter. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1498–1507, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Martin Potthast, Sebastian Köpsel, Benno Stein, and Matthias Hagen. 2016. Clickbait Detection. volume 9626, pages 810–817.

Ayla Rigouts Terryn and Miryam de Lhoneux. 2024. Exploratory Study on the Impact of English Bias of Generative Large Language Models in Dutch and French. In *Proceedings of the Fourth Workshop on Human Evaluation of NLP Systems (HumEval) @ LREC-COLING 2024*, pages 12–27, Torino, Italia. ELRA and ICCL.

Daniel Russo, Oscar Araque, and Marco Guerini. 2024. To Click It or Not to Click It: An Italian Dataset for Neutralising Clickbait Headlines. In *Proceedings of the 10th Italian Conference on Computational Linguistics (CLiC-it 2024)*, pages 829–841, Pisa, Italy. CEUR Workshop Proceedings.

Swabha Swayamdipta, Roy Schwartz, Nicholas Lourie, Yizhong Wang, Hannaneh Hajishirzi, Noah A. Smith, and Yejin Choi. 2020. Dataset cartography: Mapping and diagnosing datasets with training dynamics. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 9275–9293, Online. Association for Computational Linguistics.

Veronika Vincze and Martina Katalin Szabó. 2020. Automatic detection of Hungarian clickbait and entertaining fake news. In *Proceedings of the 3rd International Workshop on Rumours and Deception in Social Media (RDSM)*, pages 58–69, Barcelona, Spain (Online). Association for Computational Linguistics.

Yi Zhu, H. Wang, Ye Wang, Yun Li, Yunhao Yuan, and Jipeng Qiang. 2023. Clickbait Detection via Large Language Models. *ArXiv*, abs/2306.09597.

Level	Prompt
English	Clickbait headlines involve subtle and manipulative techniques to attract attention and pique readers' curiosity so they click on the article. It is important to distinguish clickbait from irrelevance, where uninteresting or gossipy headlines are not necessarily clickbait. Clickbait headlines often contain the following features: 1) Sensationalism – Does it use emotional words like "shocking," "unbelievable," or "must-see"? 2) Missing Information – Does it leave out key details, forcing the reader to click? 3) Manipulative Language – Does it promise "one simple trick" or "secrets they don't want you to know"? Is the following headline clickbait? Answer with Yes or No. Headline: "Što se događa sa slavnom pjevačicom: Novi imidž razočarao fanove" Answer: Yes Headline: "Policajac tužio svog načelnika za klevetu" Answer: Yes Headline: "Vaterpolisti dubrovačkog Juga osvojili Hrvatski kup" Answer: No Headline: "Sudarila se jahta s trajektom kod Biograda: 'Grunulo je, djeca su plakala, prestravili smo se...'" Answer: No headline Answer:
Croatian	Clickbait naslovi uključuju suptilne i manipulativne tehnike kojima se privlači pozornost i potiče znatiželja čitatelja kako bi kliknuli na članak. Važno je razlikovati clickbait s irelevantnošću pri čemu nezanimljivi ili trač naslovi ne moraju nužno biti clickbait. Clickbait naslovi često sadržavaju sljedeće značajke: 1) Senzacionalizam – Koristi li emocionalne riječi poput "šokantno", "nevjerojatno" ili "morate vidjeti"? 2) Izostavljanje informacija – Izostavlja li ključne pojedinosti, tjera li čitatelja da klikne? 3) Manipulativni jezik – Obećava li "jedan jednostavan trik" ili "tajne koje ne žele da znate"? Je li navedeni naslov clickbait naslov? Odgovori s Da ili Ne. Headline: "Što se događa sa slavnom pjevačicom: Novi imidž razočarao fanove" Odgovor: Da Headline: "Policajac tužio svog načelnika za klevetu" Odgovor: da Headline: "Vaterpolisti dubrovačkog Juga osvojili Hrvatski kup" Odgovor: Ne Headline: "Sudarila se jahta s trajektom kod Biograda: 'Grunulo je, djeca su plakala, prestravili smo se...'" Odgovor: No headline Odgovor:

Table 5: System prompts in English and Croatian used for inference in EFD (D = definition, F = features, E = explanation) configuration.
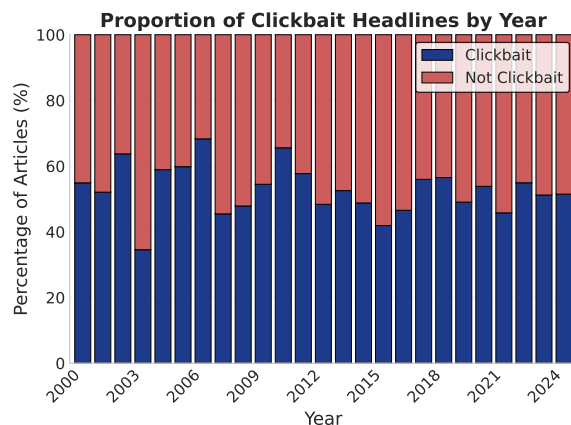

Figure 2: Percentage of news headlines classified as clickbait versus non-clickbait for each year in the dataset.

Table 6: Top 20 influential words based on attention mapping (mean IG scores)

Croatian Word	Mean IG	English Translation
cvatu	395.6445	bloom/flourish
Glumica	14.8237	actress
loš	12.4242	bad
zajedničke	11.3327	joint/common
ekonomija	10.7566	economy
Ruši	10.4093	demolishes/topples
njega	10.3574	him/his
zločinačka	8.7656	criminal
nezainteresirani	7.0193	uninterested
FUUUUUUJ	6.6199	expression of disgust
Pokrenut	6.5441	launched/initiated
petka	6.3973	Friday
SAČIĆ	6.0077	surname
Paltrow	5.8773	surname (Paltrow)
vječna	5.8190	eternal
oko	5.5182	eye/around
zadnje	5.4024	last/final
autogol	5.2721	own goal
stampeda	5.2050	stampede
Satelit	4.8463	satellite

News Outlet	Number of Articles
vecernji.hr	505
glas-slavonije.hr	394
index.hr	346
24sata.hr	330
net.hr	278
dnevnik.hr	209
jutarnji.hr	182
tportal.hr	155
slobodnadalmatija.hr	103
narod.hr	76
hrt.hr	58
direktno.hr	55
hr.n1info.com	38
dnevno.hr	35
novilist.hr	30
rtl.hr	25
lupiga.com	24
h-alter.org	20
telegram.hr	16
geopolitika.news	6

Table 7: Number of articles by news outlet (minimum 5 articles per outlet).