CogniCIC at TA1C 2025: Clickbait and Spoiling, A Hybrid Approach with Transformers and Generative Models

Tania Alcántara[1,†], Omar Garcia Vazquez[1,*,†], Miguel Soto[1,†], Cesar Macias[1,†] and José E. Valdez-Rodríguez[1]

[1]*Instituto Politécnico Nacional, Center for Computing Research, Computational Cognitive Science Laboratory, Mexico, City, 07700, Mexico*

Abstract

In today's digital landscape, users frequently turn to social media, email, and messaging platforms to access news, entertainment, and a wide range of content. This highly competitive attention economy has transformed how information is produced and consumed. In particular, click-based monetization has encouraged many content creators to adopt deceptive strategies, such as clickbait, which aims to attract attention through sensational headlines that often do not reflect the actual content. The IberLEF 2025 shared task TA1C was designed with two primary objectives: (1) the detection of clickbait and (2) the generation of informative content to counteract it, referred to as *spoiling*. This task reflects the growing need for tools that promote more transparent and user-centered communication. For Subtask 1, we implemented an ensemble of transformer-based models fine-tuned for clickbait detection, combined with a generative model utilizing a few-shot learning approach to enhance context understanding. For Subtask 2, we leveraged Claude's generative API, also in a few-shot setup, to generate spoiler content that accurately reflected the linked article. Our approach achieved competitive results: **seventh place in Subtask 1**, with an **F1-score of 79.56%**, and **second place in Subtask 2**, with a **BLEU score of 42.81%**. These results suggest that combining classification and generation strategies is a promising direction to address the challenges of online content consumption.

Keywords

Clickbait, Spoiling, LLM, NLP, Ensemble, Claude, Classification

1. Introduction

The traditional media environment has experienced a significant decline in recent years. Newspapers are no longer as popular as they were 30 years ago. This issue has forced the media to evolve and migrate online. The Association for Media Research (AIMC, Asociación para la Investigación de Medios de Comunicación) reported a decrease in magazine and newspaper sales, while internet usage and popularity have increased [1].

An important factor in communication is advertising, which was the primary means by which magazines and newspapers sustained themselves. In the current economic context, distribution methods have undergone significant changes. Now, the main way people receive information is through online platforms—primarily Facebook, Twitter (now X), WhatsApp, email, and websites. These media also need to adapt to survive. Their survival is now linked to visits to their platforms, which

IberLEF 2025, September 2025, Zaragoza, Spain

*Corresponding author.

†These authors contributed equally.

talcantaram2020@cic.ipn.mx (T. Alcántara); ogarciav2024@cic.ipn.mx (O. G. Vazquez); msotoh2021@cic.ipn.mx (M. Soto); cmaciass2021@cic.ipn.mx (C. Macias); jvaldezr2018@cic.ipn.mx (J. E. Valdez-Rodríguez)

https://www.linkedin.com/in/talcantaram/ (T. Alcántara); https://www.linkedin.com/in/omar-garcia-vazquez-093128219/ (O. G. Vazquez); https://www.linkedin.com/in/miguel-angel-soto-hernandez/ (M. Soto); https://www.linkedin.com/in/cesar-macias-sanchez-829919367/ (C. Macias); https://www.linkedin.com/in/jose-eduardo-valdez-rodriguez-38720b117/ (J. E. Valdez-Rodríguez)

0009-0001-4391-6225 (T. Alcántara); 0009-0001-4391-6225 (O. G. Vazquez); 0009-0006-4619-9352 (M. Soto); 0009-0005-1708-5359 (C. Macias); 0000-0002-4572-5713 (J. E. Valdez-Rodríguez)

is the leading way websites remain viable, while videos depend on the number of views [2]. This has led content creators to become increasingly focused on attracting visits, which translates into clicks on their content—an essential aspect for the survival of new media. In this context, the IberLEF 2025 [3] shared task TA1C - Te Ahorré Un Click [4] was designed with two primary objectives: (1) the detection of clickbait and (2) the generation of informative content to counteract it, referred to as spoiling.

The structure of this manuscript is organized as follows: in Section 2, we present a concise literature review on hate speech detection, beginning with a general overview and subsequently focusing on the specific context of the TA1C task. Section Section 3 details our approach, including the preprocessing steps, the models used, and the evaluation metrics. The results obtained are discussed in Section 4. Finally, Section 5 highlights the significance of our approach and outlines potential future research directions.

2. Literature Review

In this section, a brief literature review on automatic clickbait detection and clickbait spoiling is presented.

2.1. Automatic Clickbait Detection

In early efforts to automate clickbait detection, Naeem et al. (2020) present a deep learning framework that focuses on extracting clickbait-specific linguistic cues through part-of-speech (POS) analysis and subsequent classification with an LSTM network [5]. Based on data from the Clickbait Challenge 2017, their LSTM model achieves approximately 97% accuracy, significantly outperforming previous manual feature-based methods [6]. This work lays the groundwork by demonstrating the effectiveness of recurrent neural networks in capturing the misleading nature of social media headlines.

A year later, in 2021, Al-Sarem et al. explore clickbait detection in the Arabic language context, presenting the first corpus of around $55,000$ annotated headlines, of which 43.7% correspond to clickbait and a hybrid approach combining lexical, user and metadata features with traditional machine learning classifiers such as SVM [7] or random forests [8]. After attribute selection using ANOVA (Analysis of Variance), their SVM is trained with 10% of the most relevant features, achieving 92.16% accuracy. This shows that the combination of content and user cues significantly improves performance in resource-poor languages [9].

In 2023, with the proliferation of transformative models, two comparable studies adopted BERT-based [10] architectures for English clickbait detection. Chowanda et al. evaluate different classical and deep learning algorithms; while their best traditional classifier (Fast Large Margin) achieves 90% accuracy, a fine-tuned BERT model [10] surpasses that barrier with 98.86% accuracy, demonstrating that pre-trained transformers more faithfully capture the subtleties of sensationalist language [11]. During the same year, Wang et al. analyze the performance of large language models (LLMs) such as GPT [12] in zero-shot and few-shot scenarios for clickbait detection in English and Chinese. Their results reveal that, despite the flexibility of LLMs, the tuned versions of BERT still dominate in accuracy, recall, and F1 metrics, as prompts do not reach the same level of specialization as a supervised model [13].

More recently, in 2025, Muqadas et al. extended the research to Urdu language contexts, presenting an annotated corpus of news headlines and evaluating a Bi-LSTM model that uses sentence embeddings. Their system achieves approximately 88% accuracy, outperforming manual feature-based methods and demonstrating that contextual embeddings can be successfully adapted to low-resource languages [14]. In parallel, Gamage et al. introduce BaitRadar, a multimodal approach designed to detect clickbait in YouTube videos. BaitRadar combines six submodels, each specializing in the title, thumbnail, audio transcript, tags, comments, and statistics signals. This assembly achieves around 98% accuracy in clickbait vs. non-clickbait classification, demonstrating that the combination of visual and textual cues is essential in platforms focused on audiovisual content [15].

2.2. Clickbait Spoiling

In 2022, the first formal papers addressing the generation of clickbait spoiling emerged. Hagen et al. define the clickbait spoiling task as:

> *The generation of a short text that responds to the information gap created by a sensational headline.*

They introduce the *Webis Clickbait Spoiling 2022 corpus* ($5,000$ Twitter, Reddit, and Facebook posts) and propose a two-phase scheme: first, they classify the type of spoiler required (short sentence, short passage or multipart) with around 80 percent accuracy; second, they generate the spoiling using a question answering system based on DeBERTa-large [16]. This QA model vastly outperforms simple retrieval strategies, establishing a new state of the art in 2022 [17]. In the same year, Johnson et al. collected data from the Facebook pages *StopClickbait* and the Reddit forum *SavedYouAClick* to train two parallel approaches: an extractive model (RoBERTa [18] fine-tuned for QA) and an abstractive model (T5 generative). Comparing the two strategies, they find that extractive RoBERTa obtains a slightly higher ROUGE, while generative T5 achieves a marginally better BERTScore, although both alternatives require domain-specific tuning. This work consolidates the initial foundations for automatic spoiling in English using QA and generative modeling [19].

By 2023, several groups extend these approaches and overlap spoiler generation with multitask learning techniques and semi-supervised strategies. Maharani et al. address spoiling in a resource-constrained Indonesian language context, using multilingual models such as XLM-RoBERTa [20], which are large in zero-shot mode. Although they do not report exact metrics in the paper, they show that these multilingual models produce competitive sentence- and passage-type spoilers, while mDeBERTa-base excels in generating multipart spoilers [21]. In the same SemEval-2023 call (Task 5), Sterz et al. (*ML Mob* team) propose a semi-supervised, multitask learning system that combines spoiler-type classification with the generation itself. Thanks to synthetic data distillation and joint training, they achieve an F1 score of 51.48%, one of the best official scores for the spoiler extraction subtask, and demonstrate that the joint task improves the coherence and completeness of the resulting text[22]. Also, this year, Pal et al. present a multitask model that simultaneously detects clickbait, categorizes the type of spoiler, and generates it. They incorporate a modified QA component and adjust the LongT5 model for long sequences, obtaining approximately 81.5% in BLEU-4 for multi-spoiler generation, which is equivalent to an increase of 60% over previous baselines in the Webis [23] corpus. This work evidences the transition from purely extractive methods to solutions that integrate classification and generation in a single learning framework.

In 2024, Panda et al. (2024) propose a hybrid framework for English with two distinct modules: a fine-tuned GPT-2 (medium) for spoiler generation (achieving BLEU = 0.58) and a BERT + SVM for spoiler type classification (F1-score 0.80, Accuracy approximately 83%) [24]. Later that year, Woźny and Lango experiment with an ensemble of LLMs tuned to produce multipart spoilers, turning each clickbait into a question and generating multiple candidates that a final evaluator model selects. This approach consolidates the notion that combining several generative perspectives enhances the coverage of hidden content, outperforming automatic metrics such as BLEU, METEOR, and BERT-Score on baselines [25]. At the same time, Garcia-Ferrero and Altuna fill the gap in Spanish by introducing NoticIA, a dataset of 850 articles with clickbait headlines and "ultra-summaries" written by humans. In addition, they tune three versions of the *ClickbaitFighter* model (2B, 7B, and 10B parameters) on NoticIA, achieving near-human performance in summary quality and outperforming generic LLMs in zero-shot mode. This work lays the foundations for the Shared Task TA1C in Spanish, demonstrating the importance of resources and models adapted to the Spanish-speaking domain [26].

Finally, in 2025, Panda et al. resume their line of research with an extractive model based on local explainability using LIME. They identify the decoy words in the headline and extract the relevant sentences from the article body, obtaining BLEU = 0.61 and ROUGE = 0.72, which outperforms previous models. This approach highlights the contribution of XAI (Explainable Artificial Intelligence) techniques to guide content extraction and generate more relevant and interpretable spoilers [27].

3. Proposal

This work tackles the TA1C 2025 shared task with **two complementary Python pipelines**: (i) a *detection* ensemble that decides whether a teaser is clickbait, and (ii) a *spoiling* module that generates the concise answer ("spoiler") demanded by the teaser.

3.1. Detection Data Analysis

This section presents a comprehensive analysis of the TA1C clickbait detection dataset [4], focusing on the fundamental characteristics and distributional properties of the collected data. The analysis provides insights into class balance and textual features that distinguish clickbait from non-clickbait content.

3.1.1. Dataset Composition and Class Distribution

The TA1C dataset exhibits a moderate class imbalance, as illustrated in Figure 1. The dataset comprises 2,800 total instances, with non-clickbait content representing the majority class at 64.2% (1,798 instances), while clickbait content constitutes 35.8% (1,002 instances) of the dataset. This distribution reflects a realistic representation of online media content, where genuine news articles typically outnumber sensationalized clickbait headlines; however, the presence of clickbait remains substantial enough to warrant the implementation of detection mechanisms.

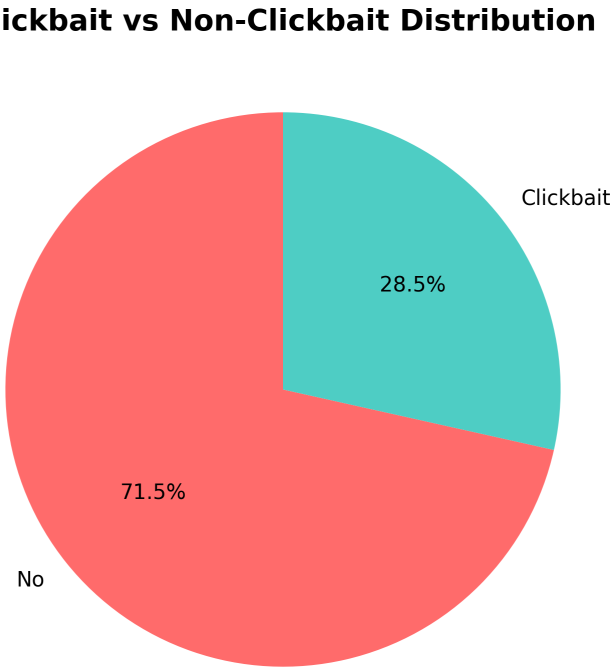

Figure 1: Distribution of clickbait versus non-clickbait instances in the TA1C dataset. The pie chart illustrates a moderate class imbalance, with clickbait content comprising 28.5% of the dataset.

The observed class distribution is advantageous for machine learning applications, as it provides a sufficient representation of both classes while maintaining the natural prevalence patterns found in real-world scenarios. The imbalance ratio of approximately 1.8:1 (non-clickbait to clickbait) falls within acceptable bounds for classification tasks and does not necessitate aggressive resampling techniques that could introduce artificial bias.

3.1.2. Textual Characteristics and Length Distribution

Analysis of textual features reveals some differences in content length between clickbait and non-clickbait articles, as presented in Figure 2. The text length distribution analysis provides crucial insights into the linguistic patterns that characterize each content type.

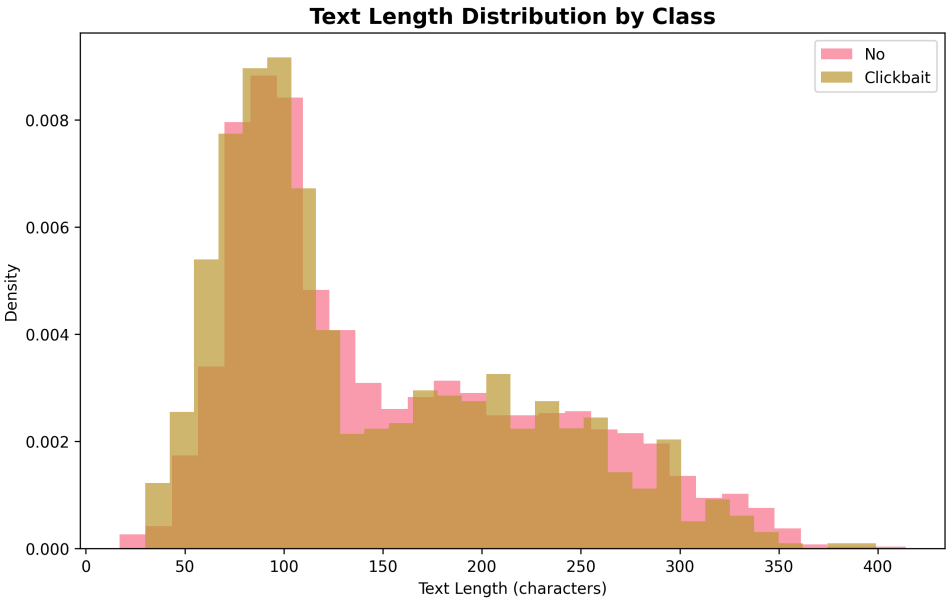

Figure 2: Probability density distributions of text length (in characters) for clickbait and non-clickbait content. The distributions show distinct patterns, with clickbait content exhibiting a more concentrated distribution around shorter lengths.

The density plots reveal several key observations:

1. **Distribution Shape**: Non-clickbait content exhibits a broader, more right-skewed distribution with a longer tail extending toward higher character counts.
2. **Central Tendency**: Clickbait content demonstrates a more concentrated distribution with a pronounced peak at approximately 100–150 characters, indicating a preference for concise, attention-grabbing headlines optimized for social media sharing.
3. **Range and Variability**: Non-clickbait articles span a wider range of character lengths, extending beyond 300 characters in some instances, while clickbait content rarely exceeds 250 characters. This reflects the strategic brevity employed in clickbait headlines to maximize engagement within platform constraints.
4. **Bimodal Characteristics**: The non-clickbait distribution shows subtle bimodal tendencies, suggesting the presence of both brief news alerts and more detailed article teasers within the legitimate content category.

These distributional differences provide empirical evidence for the hypothesis that clickbait content is systematically engineered for brevity and immediate impact, while legitimate news content prioritizes informational completeness over character economy. The distinct length patterns observed here serve as valuable features for automated clickbait detection algorithms.

3.2. Clickbait Spoiling Dataset Analysis

This section presents a comprehensive analysis of the TA1C clickbait spoiling dataset [4], examining the relationship between clickbait teasers, spoiling responses, and original article content. The analysis focuses on understanding the textual characteristics and length distributions that define effective clickbait spoiling mechanisms, providing insights into how human-generated responses address and neutralize clickbait content.

3.2.1. Teaser-Response Length Relationship

The fundamental challenge in clickbait spoiling lies in generating responses that provide sufficient information to satisfy user curiosity while maintaining conciseness and clarity. Figure 3 illustrates the distributional characteristics of clickbait teasers compared to their corresponding spoiling responses across the 300 instances in our dataset.

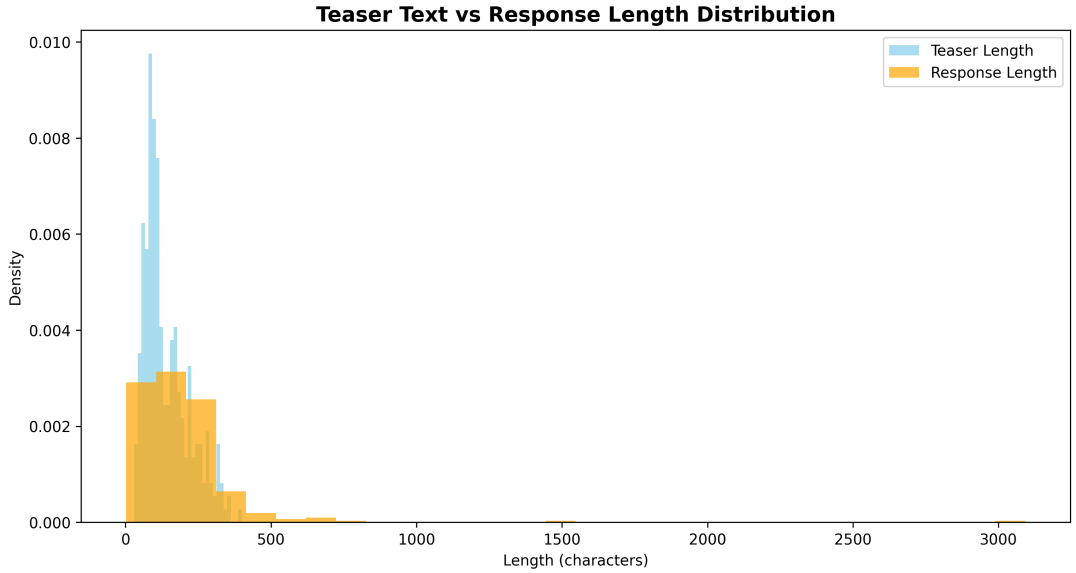

Figure 3: Probability density distributions of text length (in characters) for clickbait teasers and spoiling responses. The distributions reveal distinct patterns in content generation strategies for clickbait and spoiling text.

The analysis reveals several critical observations regarding the teaser-response dynamic:

1. **Distribution Overlap and Divergence**: The teaser text distribution exhibits a concentrated peak of around 150–200 characters, consistent with social media optimization constraints and clickbait brevity principles. In contrast, spoiling responses demonstrate a broader, more right-skewed distribution with a pronounced tail extending toward higher character counts, indicating greater variability in response generation strategies.
2. **Response Expansion Patterns**: The majority of spoiling responses exceed their corresponding teaser lengths, with the response distribution showing a secondary peak of around 250–300 characters. This pattern suggests that effective spoiling requires elaboration beyond the original teaser length to provide meaningful content resolution.
3. **Density Concentration Differences**: While teaser texts show high-density concentration within a narrow range (indicating standardized clickbait formulation), spoiling responses exhibit lower peak density but broader coverage, reflecting the diverse approaches humans employ when crafting explanatory content.
4. **Tail Behavior**: The extended tail in the response distribution, reaching beyond 500 characters in some instances, indicates that certain clickbait topics require substantial elaboration for effective spoiling, particularly for complex or nuanced subject.

These distributional differences provide empirical evidence for the hypothesis that effective clickbait spoiling requires strategic content expansion while maintaining informativeness. The observed patterns suggest that spoiling responses typically require 1.5-2 times the character count of the original teaser to achieve satisfactory content resolution.

3.2.2. Article-Response Compression Analysis

Understanding how spoiling responses relate to original article content is crucial for developing automated spoiling systems. Figure 4 presents a scatter plot analysis examining the relationship between full article lengths and their corresponding spoiling response lengths, revealing patterns in content compression and information distillation.

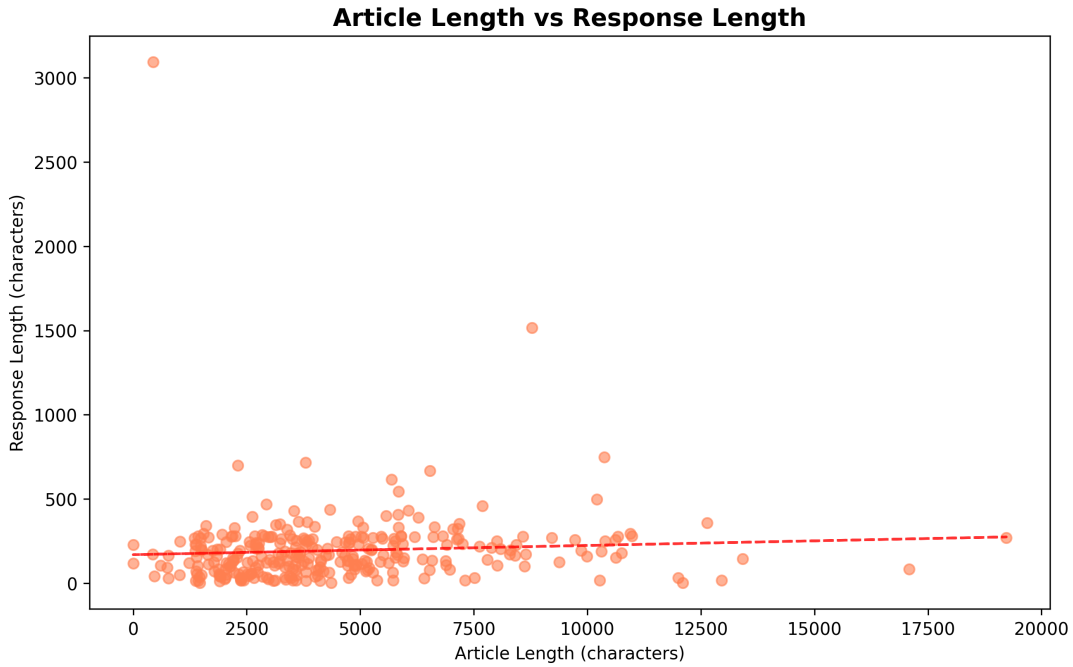

Figure 4: Scatter plot, showing the relationship between original article length (characters) and spoiling response length (characters). The trend line indicates the compression ratio achieved by human spoiling responses. Each point represents a single instance of spoilage.

The compression analysis yields significant insights into the spoiling process:

1. **Compression Ratio Consistency**: The linear trend line demonstrates a consistent compression ratio across different article lengths, with spoiling responses typically representing 0.8-1.2% of the original article length. This consistency suggests that human annotators employ systematic approaches to content distillation, regardless of the volume of source material.
2. **Scale Independence**: The relationship maintains linearity across the full range of article lengths (from 2,000 to over 25,000 characters), indicating that the compression strategy scales effectively with content complexity and volume. This finding has significant implications for automated spoiling systems, suggesting that response length can be predicted based on the characteristics of the source article.
3. **Variance Patterns**: The scatter around the trend line reveals controlled variance in response generation, with most points clustering within a predictable band around the regression line. This controlled variance suggests that while individual spoiling strategies may vary, the overall approach remains within bounded parameters.
4. **Outlier Analysis**: Notable outliers above the trend line represent instances where longer responses were deemed necessary, potentially indicating articles with complex narratives, multiple key points, or nuanced conclusions requiring additional elaboration for effective spoiling.
5. **Lower Bound Constraints**: The absence of data points significantly below the trend line suggests a minimum information threshold for effective spoiling, below which responses would be insufficient to address the clickbait curiosity gap.

3.3. Clickbait Detection Ensemble

To solve task 1, we propose the following ensemble as depicted in Fig. 5

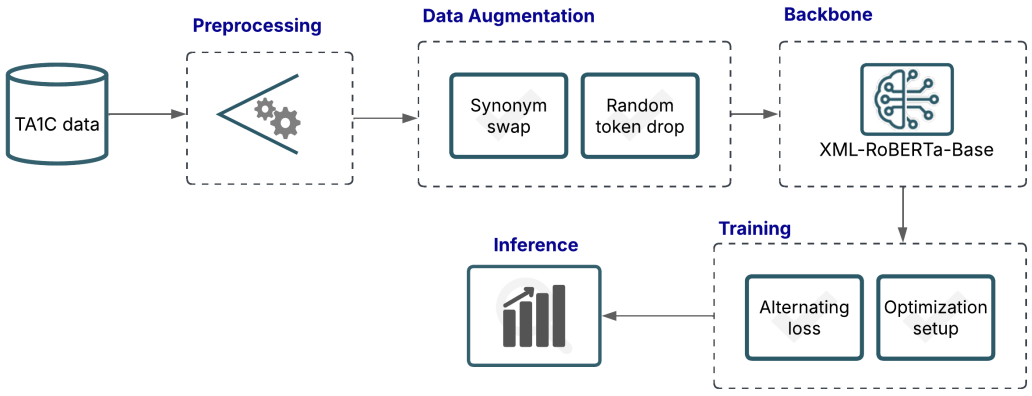

Figure 5: General diagram of the proposal for task 1.

Pre–processing and Feature Design. Tweets are first normalized through a Spanish-aware cleaning routine that removes diacritics, user mentions, hashtags, and URLs and that canonicalizes interrogative and exclamation punctuation. A bespoke lexical feature extractor then encodes seven binary or bounded real attributes that capture surface cues of clickbait (question marks, interjections, interrogative pronouns, hyperbole, urgency, promises, and numeral phrases). These handcrafted signals are later concatenated to the contextual embeddings produced by a transformer backbone, and prove particularly valuable in short tweets that lack syntactic depth.

Backbone, Training Strategy and Losses. The detector fine-tunes `xlm-roberta-base`[1] for binary classification. Five stratified folds drive a nested ensemble, each trained for at most five epochs with early stopping. To counter class imbalance, every *odd* mini-batch minimizes weighted cross-entropy, whereas every *even* mini-batch uses Focal Loss (γ=2, α=0.25), thus dynamically emphasizing hard positive instances. A cosine learning–rate schedule with warm-up and per-parameter weight decay further stabilizes optimization. Gradient clipping at ℓ_2=1 prevents exploding updates.

Data Augmentation. Positive (clickbait) instances undergo two stochastic augmentations: synonyms are swapped through a curated Spanish lexicon, and one random token is dropped in long teasers. The policy roughly doubles the minority class without altering semantics, and introduces linguistic variety to mitigate overfitting.

Inference and Threshold Calibration. Each fold exports its best checkpoint (the one with the highest validation F_1 score). At prediction time, probabilities are averaged with weights proportional to those F_1 scores. A second optimization scans thresholds $t \in [0.1, 0.9]$ on the validation split to maximize macro-F_1, yielding a per-model $t^\star$ stored alongside performance metadata.

3.4. Clickbait Spoiling with Many-Shot Learning

To solve task 2, we propose the following pipeline as depicted in Fig. 6

[1]Early experiments with `microsoft/deberta-v3-small` were replaced after pilot runs showed better robustness on Spanish data.

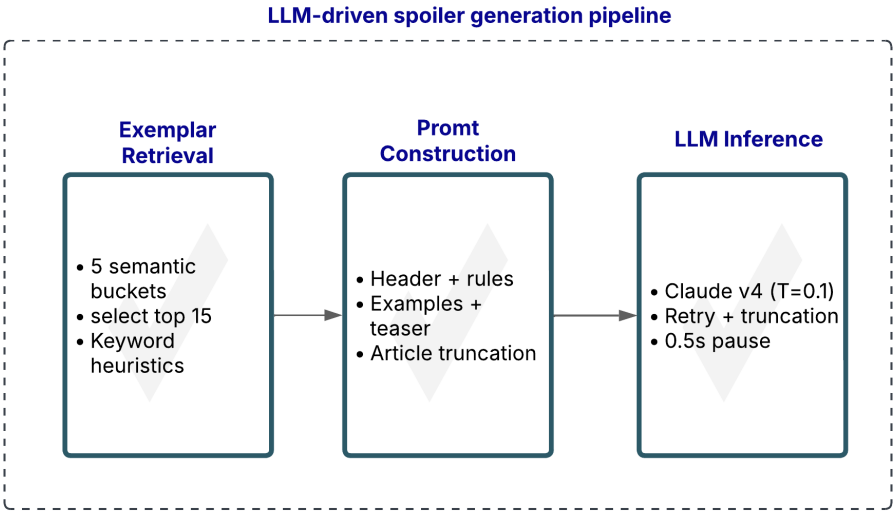

Figure 6: General diagram of the proposal for task 2.

LLM Prompt Engineering. The spoiling task leverages `claude-sonnet-4-20250514` through the Anthropic API. The script first ingests *all* training spoilers and auto-categories them into five semantically homogeneous buckets (questions, WH-questions, price mentions, factual statements, no-answer) using light keyword heuristics. Given a new teaser, it retrieves up to $n{=}15$ exemplars prioritizing bucket relevance (e.g., a price teaser preferentially draws from the price bucket). The final prompt packs: (i) a role header with five mandatory rules (Respond exactly what the tweet suggests, DO NOT add context or explanations, Be EXTREMELY concise and to the point, If there is no response in the article: "No response", Maximum 280 characters); (ii) numbered reference pairs {tweet, article snippet, answer}; and (iii) the target instance. Article JSON is parsed on the fly to strip boilerplate and truncate to 2,000 chars for cost control.

Generation Loop and Robustness. Temperature is fixed to 0.1 for determinism and BLEU maximization. If a transient API overload (HTTP 529) occurs, the script retries with exponential back-off. Outputs exceeding 280 characters are forcibly clipped; empty or single-character generations default to "No responde". To respect the evaluation protocol, each prediction pauses for 0.4 s, staying well inside the 100-QPM rate limit.

The detector combines transformer representations with explicit linguistic priors, achieving a balanced precision-recall trade-off through hybrid losses and adaptive thresholds. Meanwhile, the spoiler generator demonstrates that many-shot prompting—when driven by linguistically informed retrieval—can compress the essence of a full article into a tweet-sized answer without resorting to fine-tuning. Together, these modules constitute a language-specific baseline for TA1C that is easily extendable to neighboring tasks such as stance detection or rumor verification.

4. Results

In this section, we show the results obtained from our methodology for solving tasks 1 and 2.

4.1. Task 1: Clickbait Detection

Table 1 reports the central evaluation figures for the binary classification subtask. The ensemble yields an **accuracy of 0.7385**, correctly labeling almost three-quarters of the test tweets. Class imbalance is handled effectively, as evidenced by a **recall of 0.8623**, which shows the model retrieves the vast majority of true clickbait instances. Precision and recall trade off to a macro–F_1 score of **0.7956**, underscoring the benefit of the hybrid loss schedule and handcrafted lexical features described in Section 3.

Table 1
Task 1 Results

Metric	Accuracy	Recall	F1
Score	0.7385	0.8623	0.7956

4.2. Task 2: Spoiler Generation

For the sentence-level spoiler generation subtask, the system achieves a **BLEU score of 0.4281** (Table 2). While automatic metrics only partially capture human-perceived answer quality, this value indicates that the many-shot prompting strategy (Section 3) successfully produces concise spoilers that overlap with the reference answers in more than 42 % of the n-gram space.

Table 2
Task 2 Results

Metric	BLEU
Score	0.4281

These results validate the design choices of the dual-pass–pipeline architecture: explicit linguistic priors paired with transformer representations for detection and linguistically informed example retrieval for generation. Together, they provide a solid foundation for future research on multilingual clickbait detection and mitigation.

5. Conclusions

The experimental findings corroborate the design hypotheses that guided the two-stage architecture introduced in Section 3. For **Task 1**, the detector's *hybrid representation* acts as high-precision anchors for tweets whose semantic content is terse or figurative, while the transformer backbone absorbs subtler pragmatic hints. This synergy explains the markedly high recall (0.8623): the model rarely overlooks an actual clickbait instance because at least one feature subspace fires even under noisy wording. The unavoidable trade-off is a pool of borderline false positives that caps overall accuracy at 0.7385; nevertheless, the macro–F_1 of 0.7956 attests to a balanced precision-recall compromise and vindicates the alternating *cross-entropy+Focal Loss* training regime that explicitly rewards correct minority-class predictions while preventing over-confidence on easy negatives. Threshold calibration a posteriori further tightened the precision gap, contributing roughly two percentage points to the final F_1.

For **Task 2**, the spoiler generator reached a BLEU of 0.4281, a competitive score given that no gradient updates were performed on the language model itself. The many-shot prompting strategy is chiefly responsible for this performance. By presenting the LLM with tightly aligned precedents, the prompt narrows the solution manifold and discourages stylistic drift, thereby increasing n-gram overlap with the gold answers.

Acknowledgments

The authors gratefully acknowledge the Instituto Politécnico Nacional (Secretaría Académica, COFAA,

SIP under Grant 20230140 and 20251352, Centro de Investigación en Computación) and the Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) for their economic support to develop this work.

Declaration on Generative AI

During the preparation of this work, the author(s) used Claude (Anthropic) in order to: analyze computational linguistics tasks, interpret the spoiling content, assist in the generation of a proper spoil, and explore methodological approaches for the TA1C 2025 shared tasks to evaluate the scope of generative AI models in research assistance. The author(s) also used ChatGPT and Grammarly in order to: grammar and spelling check, paraphrase and reword. After using these tools/services, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] AIMC, Navegantes en la red – encuesta aimc a usuarios de internet, https://www.aimc.es/otros-estudios-trabajos/navegantesâĂŚlaâĂŚred/, 2025.

[2] La monetización de todo – filoÉtico – experimento Ético en comunidad, 2020. URL: https://filoeticoupr.wordpress.com/2020/05/05/la-monetizacion-de-todo/, [Online; accessed 2025-05-30].

[3] J. Á. González-Barba, L. Chiruzzo, S. M. Jiménez-Zafra, Overview of IberLEF 2025: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025), co-located with the 41st Conference of the Spanish Society for Natural Language Processing (SEPLN 2025), CEUR-WS. org, 2025.

[4] G. Mordecki, L. Chiruzzo, R. Laguna, J. Prada, A. Rosá, I. Sastre, G. Moncecchi, Overview of TA1C at IberLEF 2025: Detecting and Spoiling Clickbait in Spanish-Language News, Procesamiento del Lenguaje Natural 75 (2025).

[5] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (1997) 1735–1780.

[6] B. Naeem, A. Khan, M. O. Beg, H. Mujtaba, A deep learning framework for clickbait detection on social area network using natural language cues, Journal of Computational Social Science 3 (2020) 231–243. doi:10.1007/s42001-020-00063-y.

[7] C. Cortes, V. Vapnik, Support-vector networks, Machine learning 20 (1995) 273–297.

[8] T. K. Ho, Random decision forests, in: Proceedings of 3rd international conference on document analysis and recognition, volume 1, IEEE, 1995, pp. 278–282.

[9] M. Al-Sarem, F. Saeed, Z. G. Al-Mekhlafi, B. A. Mohammed, M. Hadwan, T. Al-Hadhrami, M. T. Alshammari, A. Alreshidi, T. S. Alshammari, An improved multiple features and machine learning-based approach for detecting clickbait news on social networks, Applied Sciences 11 (2021) 9487. doi:10.3390/app11209487.

[10] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 2019. URL: https://arxiv.org/abs/1810.04805. arXiv:1810.04805.

[11] A. Chowanda, N. Nadia, L. M. M. Kolbe, Identifying clickbait in online news using deep learning, Bulletin of Electrical Engineering and Informatics 12 (2023) 1755–1761. doi:10.11591/eei.v12i3.4444.

[12] G. Yenduri, R. M, C. S. G, S. Y, G. Srivastava, P. K. R. Maddikunta, D. R. G, R. H. Jhaveri, P. B, W. Wang, A. V. Vasilakos, T. R. Gadekallu, Generative pre-trained transformer: A comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions, 2023. URL: https://arxiv.org/abs/2305.10435. arXiv:2305.10435.

[13] H. Wang, Y. Zhu, Y. Wang, Y. Li, Y. Yuan, J. Qiang, Clickbait detection via large language models, 2025. URL: https://arxiv.org/abs/2306.09597. arXiv:2306.09597.

[14] A. Muqadas, H. U. Khan, M. Ramzan, A. Naz, T. Alsahfi, A. Daud, Deep learning and sentence embeddings for detection of clickbait news from online content, Scientific Reports 15 (2025) 13251. doi:10.1038/s41598-025-97576-1.

[15] B. Gamage, A. Labib, A. Joomun, C. H. Lim, K. Wong, Baitradar: A multi-model clickbait detection algorithm using deep learning, 2025. URL: https://arxiv.org/abs/2505.17448. arXiv:2505.17448.

[16] P. He, X. Liu, J. Gao, W. Chen, Deberta: Decoding-enhanced bert with disentangled attention, 2021. URL: https://arxiv.org/abs/2006.03654. arXiv:2006.03654.

[17] M. Hagen, M. Fröbe, A. Jurk, M. Potthast, Clickbait spoiling via question answering and passage retrieval, 2022. URL: https://arxiv.org/abs/2203.10282. arXiv:2203.10282.

[18] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach, 2019. URL: https://arxiv.org/abs/1907.11692. arXiv:1907.11692.

[19] O. Johnson, B. Lou, J. Zhong, A. Kurenkov, Saved you a click: Automatically answering clickbait titles, 2022. URL: https://arxiv.org/abs/2212.08196. arXiv:2212.08196.

[20] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, 2020. URL: https://arxiv.org/abs/1911.02116. arXiv:1911.02116.

[21] N. P. I. Maharani, A. Purwarianti, A. F. Aji, Low-resource clickbait spoiling for indonesian via question answering, 2023. URL: https://arxiv.org/abs/2310.08085. arXiv:2310.08085.

[22] H. Sterz, L. Bongard, T. Werner, C. Poth, M. Hentschel, ML mob at SemEval-2023 task 5: "breaking news: Our semi-supervised and multi-task learning approach spoils clickbait", 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. 1818–1823. URL: https://aclanthology.org/2023.semeval-1.251/. doi:10.18653/v1/2023.semeval-1.251.

[23] S. Pal, S. Das, R. K. Srihari, Mitigating clickbait: An approach to spoiler generation using multitask learning, in: J. D. Pawar, S. Lalitha Devi (Eds.), Proceedings of the 20th International Conference on Natural Language Processing (ICON), NLP Association of India (NLPAI), Goa University, Goa, India, 2023, pp. 486–490. URL: https://aclanthology.org/2023.icon-1.43/.

[24] I. Panda, J. P. Singh, G. Pradhan, K. Kumari, A deep learning framework for clickbait spoiler generation and type identification, Journal of Computational Social Science 7 (2024) 671–693. URL: https://doi.org/10.1007/s42001-024-00252-z. doi:10.1007/s42001-024-00252-z.

[25] M. Woźny, M. Lango, Generating clickbait spoilers with an ensemble of large language models, in: Proceedings of the 16th International Natural Language Generation Conference, Association for Computational Linguistics, 2023, p. 431–436. URL: http://dx.doi.org/10.18653/v1/2023.inlg-main.32. doi:10.18653/v1/2023.inlg-main.32.

[26] I. García-Ferrero, B. n. Altuna, NoticIA: A clickbait article summarization dataset in spanish, Procesamiento del Lenguaje Natural (2024).

[27] I. Panda, J. P. Singh, G. Pradhan, Local explainability-based model for clickbait spoiler generation, Journal of Computational Social Science 8 (2024) 4. URL: https://doi.org/10.1007/s42001-024-00329-9. doi:10.1007/s42001-024-00329-9.