Generate First, Then Sample: Enhancing Fake News Detection with LLM-Augmented Reinforced Sampling

Zhao Tong[1,2*], Yimeng Gu[3*], Huidong Liu[4], Qiang Liu[5], Shu Wu[5,6†], Haichao Shi[1], Xiao-Yu Zhang[1†]
[1]Institute of Information Engineering, Chinese Academy of Sciences
[2]School of Cyber Security, University of Chinese Academy of Sciences
[3]Queen Mary University of London, [4]College of Computer Science, Chongqing University
[5]New Laboratory of Pattern Recognition (NLPR),
State Key Laboratory of Multimodal Artificial Intelligence Systems (MAIS),
Institute of Automation, Chinese Academy of Sciences
[6]Key Laboratory of Cyberspace Security, Ministry of Education
tongzhao@iie.ac.cn, yimeng.gu@qmul.ac.uk,
shu.wu@nlpr.ia.ac.cn, zhangxiaoyu@iie.ac.cn

Abstract

The spread of fake news on online platforms has long been a pressing concern. Considering this, extensive efforts have been made to develop fake news detectors. However, a major drawback of these models is their relatively low performance - lagging by more than 20% - in identifying *fake* news compared to *real* news, making them less suitable for practical deployment. This gap is likely due to an imbalance in the dataset and the model's inadequate understanding of data distribution on the targeted platform. In this work, we focus on improving the model's effectiveness in detecting *fake* news. To achieve this, we **first** adopt an LLM to **generate** fake news in three different styles which are later incorporated into the training set, to augment the representation of fake news. **Then**, we apply Reinforcement Learning to dynamically **sample** fake news, allowing the model to learn the optimal real-to-fake news ratio for training an effective fake news detector on the targeted platform. This approach allows our model to perform effectively even with a limited amount of annotated news data and consistently improve detection accuracy across different platforms. Experimental results demonstrate that our approach achieves state-of-the-art performance on two benchmark datasets, improving *fake* news detection performance by 24.02% and 11.06% respectively.

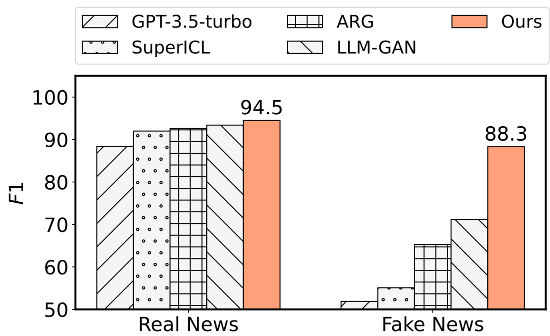


Figure 1: Group-wise F1 comparison of LLM-based models on GossipCop (Shu et al., 2020). Previous models tend to have a large performance gap (by more than 20%) between classifying *fake* news and *real* news. Yet, our model is able to significantly narrow down the gap, achieving 88.3% in F1 on *fake* news and 94.5% on *real* news.

1 Introduction

Fake news generally refers to news that contains false information (Shu et al., 2017). The spread of fake news on online news platforms has long raised significant concerns, particularly when it pertains to key social issues such as public health (Naeem and Bhatti, 2020) and politics (Fisher et al., 2016). Disseminating fake news on these domains often has severe consequences, as it can mislead the public and cause social unrest. Therefore, we argue that it is vital to develop fake news detectors that can effectively identify fake news before it spreads.

To date, numerous efforts have been made to detect fake news. Early approaches focus on building deep learning models to learn better representations of news articles for detection purposes (Wang et al., 2018a). Subsequent works have leveraged the social context of news and external knowledge (Wu et al., 2024). With the advent of Large Language Models (LLMs), recent studies have increasingly exploited the reasoning capabilities of LLMs (Hu et al., 2024; Xu et al., 2023), as well as their enhancement of news content (Park et al., 2024), social context (Nan et al., 2024), and external knowledge (Tian et al., 2024). While these LLM-based

[*]These authors contributed to the work equally.
[†]To whom correspondence should be addressed.

Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 24276–24290
July 27 - August 1, 2025

methods have shown promising results on benchmark datasets, they often struggle to detect *fake* news effectively, especially when compared to their performance in identifying real news (see Fig. 1). However, considering that fake news detection is a task grounded in realistic needs, these models may not fulfill the expectation of real-world online news platforms, where they might fail to identify a large amount of fake news in actual deployment.

In contrast, an ideal fake news detector should fulfill two expectations. *First*, it should achieve promising performances even with a limited amount of news data, especially *fake* news. In reality, fake news is far less prevalent than real news, and is often underrepresented in existing datasets (Shu et al., 2020). This imbalance makes it difficult for previous LLM-based models, which are trained exclusively on the annotated news in the dataset, to effectively identify fake news. We argue that it is crucial to fully utilize available fake news to learn more distinctive features, in order to improve the model's performance in detecting fake news. *Secondly*, an effective fake news detector should have data distribution awareness. In fact, news data on different platforms often follow distinct distribution patterns. In addition, the news label distribution in the training set significantly affects the model's detection performance. However, the optimal fake and real news distribution is often difficult to determine a priori. Therefore, the ability to autonomously learn the optimal ratio of real and fake news will help the model better adapt to the target platform.

To address the aforementioned challenges, we propose GSFND (Generate first and then Sample for Fake News Detection), a novel framework for effective *fake* news detection. GSFND first prompts an LLM to generate fake news in various styles. Then, the original fake news and generated fake news are both included in the training set. This approach helps mitigate the under-representation of fake news in the original dataset. To further inject data distribution awareness into our model, we train a Reinforcement Learning (RL) agent to learn the optimal ratio of real news to fake news (real-to-fake) during model training. Specifically, the RL agent is responsible for sampling fake news according to the selected proportion in each iteration. During training, the agent is encouraged to optimize both the proportion of fake news in the training set and the fake news detector's performance on the validation set. This dynamic adjustment of the real-to-fake ratio, learned from both the semantics of the news and the distribution of news labels, enables the model to gain deeper insights and thus effectively detect fake news.

In summary, our main contributions are as follows.

- We propose a novel approach GSFND to enhance fake news detection by automatically optimizing the real-to-fake news ratio during model training.
- The model incorporates LLM-generated fake news to address the under-representation of fake news in the original dataset, while also employing reinforced sampling to inject data distribution awareness.
- Our model achieves state-of-the-art (SOTA) performance on two benchmark datasets for fake news detection. Notably, it improves the F1 score on *fake* news by 24.02% and 11.06% on GossipCop (Shu et al., 2020) and Weibo21 (Nan et al., 2021) datasets, respectively.

2 Related Work

2.1 Fake News Detection

Most existing works on fake news detection leverage deep learning models and develop pattern-based approaches. Early attempts (Castillo et al., 2011; Wu et al., 2015) detect fake news through linguistic features. With the advancement of deep learning, subsequent works (Yu et al., 2017; Wang et al., 2018b) have built neural networks to capture the semantic meaning of news and use this to assess its authenticity. Other works also consider the social context of news, including the comment-reply relationship (Wu et al., 2020) and the propagation network (Han et al., 2020; Bian et al., 2020), and incorporate it to represent the news. Later studies (Xu et al., 2022; Hu et al., 2021) further enhance detection by integrating external knowledge related to the news. Augmenting news data is effective but requires considerable effort. In our work, we leverage LLMs to augment news data, which is efficient and capable of providing diverse context.

2.2 LLMs for Fake News Detection

Since the rise of LLMs, numerous studies have explored their role in enhancing fake news detection. Several works take advantage of its reasoning capability, including LLM-generated multi-perspective rationales (Hu et al., 2024), LLM-generated user-news interaction network (Wan et al., 2024), LLMs' self-debate on adversarial rationales (Wang et al., 2024a). Other works use LLMs to augment the news: Nan et al. (2024) uses them to generate user comments, in order to facilitate fake news detection; Ma et al. (2024) obtains effective news embedding and entities from LLMs. Prior researches have shown that LLMs are not yet sufficiently effective in detecting fake news on their own, but they can serve as a valuable support to smaller models.

Although previous works have addressed fake news detection from a wide range of perspectives, they do not focus on improving the detection accuracy on *fake* news. In contrast, our work specifically aims to enhance the *fake* news detection performance via LLM-generated data and reinforcement learning.

2.3 RL for Fake News Detection

Reinforcement Learning (RL) is effective in performing non-continuous actions and is widely applied across various scenarios. For instance, Mosallanezhad et al. (2022) employs RL to learn domain-invariant news features. Gu et al. (2024) explores the use of an RL-based gating mechanism to filter out the noisy modality in multimodal news. Wang et al. (2020) uses RL to automatically select high-quality samples from those with weak labels obtained from the annotator. Guo et al. (2023) benefits from DRL by finding the optimized input news. Yang et al. (2022) generates propagation subgraphs by maximizing the expected accuracy for fake news detection.

In our work, we leverage RL to dynamically determine the optimal real-to-fake ratio of the training set.

3 Preliminary

Fake news detection is the task of determining whether a given news item is real or fake. Formally, let $\mathcal{D} = \{(x_i, y_i)|x_i \in \mathcal{X}, y_i \in \mathcal{Y}\}_{i=1}^{N}$ be a set of news items $\mathcal{X}$ and their authenticity labels $\mathcal{Y}$. N denotes the total number of news items. Given the i-th news item x_i, the goal of fake news detection is to predict its authenticity label $y_i \in \{0, 1\}$, where $y_i = 1$ indicates that the news item x_i is fake, and $y_i = 0$ indicates that it is real.

This task is formulated as a binary classification problem, where the model $f(\cdot)$ learns a mapping $f: \mathcal{X} \rightarrow \mathcal{Y}$ from the news items $\mathcal{X}$ to their authenticity labels $\mathcal{Y}$.

4 Methodology

In this section, we present our proposed approach, GSFND. The overall architecture is illustrated in Fig. 2. GSFND consists of three main components: (*i*) *Multi-perspective fake news generation*, which prompts an LLM to generate fake news in three different styles based on the given fake news; (*ii*) *Reinforced sampling*, which leverages Reinforcement Learning (RL) to learn an optimal ratio of real news to fake news during the training process; and (*iii*) *Fake news detector*, which classifies the news, and its performance on the validation set is used to compute the RL reward.

4.1 Multi-Perspective Fake News Generation

To enrich the fake news in the training set, this module generates fake news from multiple perspectives, effectively addressing the under-representation of fake news in the original dataset. Given the strong text generation capabilities of LLMs, we prompt an LLM to rewrite the original fake news in three different styles, thereby tripling the number of fake news examples in the dataset.

Prompt Design. Inspired by Hu et al. (2024), we design a Chain-of-Thought (CoT) prompt to generate fake news in three specified styles. The structure of the prompt is shown in Fig. 3. In the prompt, we first ask the LLM to list key points of the original news. This is to help the LLM accurately understand the news. Then, the LLM is asked to *rewrite*, *expand* or *disguise* the news while keep the semantics of the news. Here, *rewrite* refers to paraphrasing the news; *expand* refers to adding contents to the news; *disguise* refers to rewriting the news using the tone of official news reports. This requirement ensures that the generated news is still fake news. Examples of generated news can be found in Appendix A.5.

For each piece of fake news, we generate three variations. Hence, we are able to enhance and diversify fake news in the original dataset.

4.2 Reinforced Sampling

After enhancing the fake news, this module aims to optimize the news label distribution during training, so that the model can further improve *fake* news detection accuracy and better adapt to the characteristics of news from the platform. Specifically, we formulate the task of learning the optimal

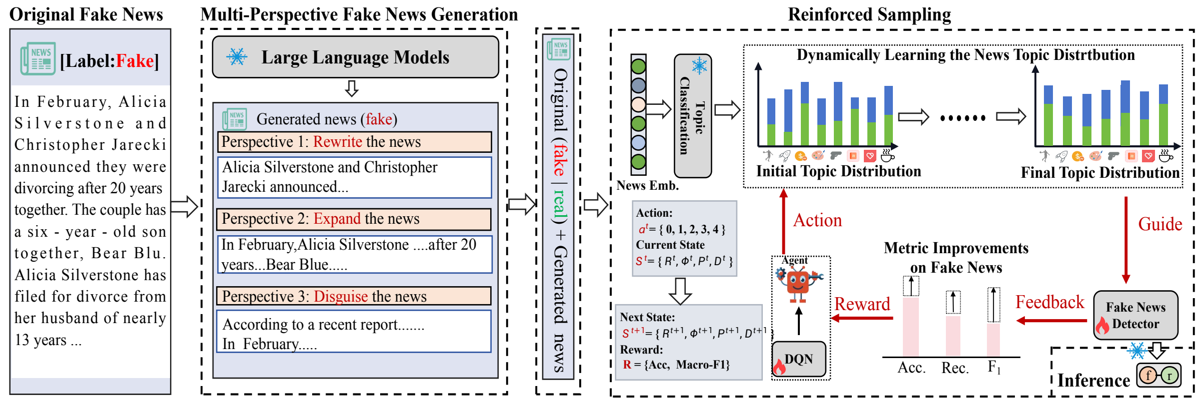


Figure 2: Overall architecture of GSFND. It consists of three components: (*i*) *Multi-perspective fake news generation*, which uses an LLM to generate fake news in 3 different styles based on the given fake news; (*ii*) *Reinforced sampling*, which uses Reinforcement Learning (RL) to learn an optimal real-to-fake news ratio during the training process; (*iii*) *Fake news detector*, which classifies the news and provides its performance for computing the RL award. During inference, its parameters are frozen and the classifier is used directly for detecting fake news.

CoT Prompt for Fake News Generation

Input: Follow the news, please {rewrite | expand | disguise} the news and keep the semantics of the news. The news is <original_news>. Let's think step by step. First, identify the key points of the news. Second, {Rewrite | Expand | Disguise} the news.

Output Format:
Step 1: Identity the key points:
①: <key_point_1>
②: <key_point_2>
Step 2: {Rewrite | Expand | Disguise} the news:
Input News: <reiterated_original_news>
Generated News: <generated_news>

Figure 3: The prompt used for generating news in three specified styles.

sampling ratio a_{opt} as a Markov Decision Process (MDP) (Littman, 2015) and use Deep Q-learning (DQN) (Osband et al., 2016) to dynamically adjust the sampling proportion of synthesized fake news (original fake news + LLM-generated fake news).

The overall learning process is as follows. At each time step t ($1 \leq t \leq T$), the agent observes the current **state** s^t and selects an **action** a^t from the action space $\mathcal{A}$ based on the estimated Q-value $Q(s^t, a^t)$, where the action corresponds to a specific adjustment to the sampling ratio. Here, the Q-value $Q(s^t, a^t)$ represents the expected cumulative reward of taking action a^t in state s^t. After executing the action a^t, the agent receives a **reward** $r(s^t, a^t)$ from the environment, which reflects the effectiveness of the adjusted sampling strategy, and the current state gets updated to a **state** s^{t+1}. Through this iterative interaction with the environment, the Q-value is continuously updated, allowing the agent to optimize the sampling strategy. The ultimate **objective** of DQN is to learn a policy π that maximizes the expected cumulative reward, ensuring that the training data distribution becomes more balanced and significantly enhancing the performance of the downstream fake news detector.

Next, we detailedly describe the **state**, **action**, **reward** and **objective function** defined in our approach.

State. The state s in the MDP consists of four components: $\{R, \Phi, P, D\}$, which capture the key aspects of the data distribution and the classification dynamics:

$$s = \{R, \Phi, P, D\}, \tag{1}$$

$$R = \frac{N_{\text{real}}}{N_{\text{fake}}}, \tag{2}$$

$$\Phi = \left\{ R^{\phi_j} = \frac{N_{\text{real}}^{\phi_j}}{N_{\text{fake}}^{\phi_j}}, \quad j = 1, 2, \ldots, K \right\}, \tag{3}$$

$$P = \{(p_{\text{real}}(x_i), p_{\text{fake}}(x_i)), \quad i = 1, 2, \ldots, N\}, \tag{4}$$

$$D = \{\text{Topic}(x_i), \quad i = 1, 2, \ldots, N\}. \tag{5}$$

(*i*) **Global ratio** R refers to the global (contrast to topic-wise) real-to-fake ratio (in Eq. 2) where N_{real} and N_{fake} denote the number of real and fake news. R helps the agent to understand the overall class imbalance.(Dataset-specific topic distributions are provided in Appendix A.3.) (*ii*) **Topic-wise ratio** Φ represents the topic-wise real-to-fake ratios (in Eq. 3), calculated for each topic ϕ_j, where $N_{\text{real}}^{\phi_j}$ and $N_{\text{fake}}^{\phi_j}$ are the numbers of real and fake news of topic ϕ_j. Φ offers information on topic-specific class imbalance. (*iii*) **Confidence scores** P contains classification confidence scores of a pre-trained model (in Eq. 4). These scores enable the agent to prioritize samples based on the certainty of the model. (*iv*) **Topic label** D (in Eq. 5) represents the topic label (mapped to a numerical value, *e.g.*, 0, 1, 2...) assigned to each news, which adds semantic context into the state. These labels are derived from a pre-trained BERT model, where each unique value corresponds to a specific news topic category. During training, we concatenate the elements of these four components to form a vector representation as state s^t. In this way, it incorporates global imbalance, topic-specific class imbalance, classification confidence, and topic knowledge into the state, allowing the agent to holistically consider the fake news detection performance and make more informed actions.

Action. The action space $\mathcal{A} = \{0, 1, 2, 3\}$ corresponds to four distinct proportions of fake news: sampling *50%* (action 0), *100%* (action 1), *150%* (action 2) and *200%* (action 3). That is to say, each action $a^t \in \mathcal{A}$ results in one fake news proportion ρ_{fake}. This can be expressed in Eq. 6:

$$\rho_{\text{fake}}(a^t) = 0.5 \times (a^t + 1), \tag{6}$$

where a^t represents the selected action at time step t. The optimal action taken by the agent at time step t is formulated by $a^t = \underset{a}{argmax}\, Q(s^t, a)$, where the Q-value $Q(s^t, a)$ is the accumulated rewards of taking the action a.

In this way, we obtain 4 distinct fake news proportions: $[50\%, 100\%, 150\%, 200\%]$, within the range of $[0.50, 2.00]$. This range is carefully designed based on two key considerations: *first*, a proportion below 0.50 would under-represent fake news in the training process, potentially reducing the model's ability to detect fake news; *second*, a proportion above 2.00 would over-represent fake news, introducing reverse data imbalance and model bias. Additionally, this range aligns with the diversity and coverage achieved by our multi-perspective fake news generation strategy (Sec. 4.1), which employs three distinct prompts. This ensures the generated dataset is both comprehensive and representative, providing a solid foundation for robust model training.

This dynamic adjustment mechanism allows the agent to explore and exploit different sampling strategies, so that it can eventually learn the optimal sampling proportions for fake news.

Reward. To guide the agent in learning an optimal sampling strategy, the reward function must reflect the classifier's performance on real and fake news. Thus, we combine two key metrics: accuracy (Acc) and the macro-average F1-score (Macro-F_1). Accuracy evaluates overall prediction performance, while Macro-F_1 emphasizes balanced performance across classes, particularly in imbalanced datasets. This combination ensures the reward function captures both *overall* and *class-specific* performance, preventing the model from favoring the majority class (*e.g.*, real news) at the expense of the minority class (*e.g.*, fake news).

Therefore, the reward function is formulated as:

$$r(s^t, a^t) = \alpha \cdot \text{Acc} + \beta \cdot \text{Macro-}F_1, \tag{7}$$

where s^t is the state at time step t, and a^t is the action executed by the agent. The weights α and β are hyperparameters that balance the contributions of accuracy and Macro-F_1. To ensure that the reward reflects the model's generalization performance, we compute both metrics on the validation set.

Note that the weights α and β can be adjusted according to the application scenarios. On imbalanced datasets, a larger β emphasizes Macro-F1, improving the minority class performance. Conversely, on balanced datasets, a larger α guides the model to focus on optimizing overall accuracy. This design enhances the flexibility of the model in real-world applications.

Objective Function. We choose to use deep Q-Learning (Osband et al., 2016) to learn an optimal policy for the agent. In implementation, we adopt a multilayer perceptron (DQN in Fig. 2) to estimate the $Q(\cdot)$ function. The deep Q-learning network takes the state s^t as input and outputs the Q-value of all possible actions. We update the network by minimizing the following loss function L_q:

$$\begin{aligned} L_q &= \mathbb{E}_{s^t, a^t}\Big[\big(Y^t - Q(s^t, a^t)\big)^2\Big], \\ \text{where } Y^t &= \mathbb{E}_{s^{t+1}}\Big[r^t + \lambda \max_{a^{t+1}} Q(s^{t+1}, a^{t+1} | s^t, a^t)\Big]. \end{aligned} \tag{8}$$

In the above equation, $Y^t - Q\left(s^t, a^t\right)$ is the temporal difference error, where Y^t is the target value of $Q\left(s^t, a^t\right)$. r^t is the reward of taking the action a^t, computed by Eq.7, and $\lambda \underset{a^{t+1}}{max}\, Q\left(s^{t+1}, a^{t+1}|s^t, a^t\right)$ is the future reward estimated by the current deep Q-learning network with s^{t+1}.

In order to train the deep Q-learning network, we collect some offline samples in advance, *i.e.*$\{(s^t, a^t, r^t, s^{t+1})|1 \leq t \leq T\}$, as input to feed the network. For the LLM generated data sampling, we adjust the sampling ratio by actions $a^t = \{0, 1, 2, 3\}$ and train the current network for one epoch with new sampling. After one epoch of training, we compute the reward r^t and obtain the next state s^{t+1}. We keep doing this until all the states have been traversed. The details are presented in Algorithm 1.

Algorithm 1 Training of the Reinforced Sampling.

Require:
Synthesized fake news: the original fake news and the generated fake news.

Ensure:
The samples, denoted as $\{(s^t, a^t, r^t, s^{t+1})\}$, are used for training the deep Q-learning network (DQN).

1: **for** a **in** all actions **do**
2: Compute and obtain the current state $s^t = \{R^t, \Phi^t, P^t, D^t\}$.
3: Take the action a to adjust the sampling proportion for fake news.
4: Sample from the synthesized fake news.
5: Train the DQN network by Eq.8 for one epoch.
6: Compute the sampling ratio of real news to fake news according to Eq. 6, and obtain the updated state $s^{t+1} = \{R^{t+1}, \Phi^{t+1}, P^{t+1}, D^{t+1}\}$.
7: Compute the reward r^t according to Eq. 7.
8: **end for**
9: **if** s^{t+1} is an unseen state **then**
10: **Go** to step 2.
11: **end if**

4.3 Detection-Sampling Synergy

The detector is responsible for classifying news and providing performance metrics to compute the RL reward during training. Given an input news text x_i, the detector encodes it into a feature vector $\mathbf{h}_i$ using a BERT model as PLM. This vector is then passed through a fully connected layer followed by a softmax activation to output the probabilities for the two classes (*real* or *fake*), formulated as:

$$\mathbf{h}_i = \text{PLM}(x_i), \quad \hat{y}_i = \text{softmax}(\mathbf{W} \cdot \mathbf{h}_i + \mathbf{b}), \tag{9}$$

where $\mathbf{W}$ and $\mathbf{b}$ are trainable parameters.

The training of the Reinforced Sampling framework (Sec. 4.2) and the detector are mutually beneficial. Reinforced Sampling dynamically adjusts the proportions of real and fake news in the training data, providing a balanced and diverse dataset that helps the detector learn stronger classification boundaries and generalize better. Meanwhile, the detector's performance (measured by Accuracy and Macro-F1) serves as a feedback signal to guide the RL agent in optimizing its sampling policy.

In summary, RL-sampled training data improves the robustness and accuracy of the fake news detector. In turn, it provides more effective feedback to enhance the Reinforced Sampling framework (Sec. 4.2).

5 Experimental Design

5.1 Datasets

We evaluate our model on two widely used datasets for fake news detection: the Chinese dataset Weibo21 (Nan et al., 2021) and the English dataset GossipCop (Shu et al., 2020). Following prior work (Hu et al., 2024), we split the above datasets into training, validation, and test set (Tab. 2). We set the real-to-fake ratio to a fixed value (*e.g.*, 1:1) in the validation and test set to ensure a fair and consistent evaluation of the detector's performance.

5.2 Implementation Details

Model Architecture. In DQN, we build a three-layer fully connected neural networks with one further hidden layer of 12 units to estimate the $Q(\cdot)$ function. Each fully connected layer is followed by a ReLU activation function. In the Fake News Detector, we use the pre-trained BERT as the feature extractor (PLM).

Training. The experiments are conducted on 4 NVIDIA GeForce RTX 4090 GPUs (24GB). The model is optimized with Adam (Kingma, 2014). The batch size is 16. The learning rate is 10^{-4}. The model is trained for 10 epochs.

LLM Choice. ChatGLM-6B (GLM et al., 2024) demonstrates strong comprehension and generation capabilities in both Chinese and English, as well as in texts that seamlessly blend both languages. We select ChatGLM-6B as the large language

Table 1: Performance comparison between GSFND and baselines on Weibo21 and GossipCop datasets. The best performance is highlighted in **bold**. The second best performance is <u>underlined</u>. Δ *Improve* represents our method's relative improvements over the second best performance in percentage.

Method		Weibo21				GossipCop			
		mac F1	Acc.	F1-real	F1-fake	mac F1	Acc.	F1-real	F1-fake
Deep-learning -based	BERT	0.753	0.754	0.769	0.737	0.765	0.862	0.916	0.615
	EANN	0.754	0.756	0.773	0.736	0.763	0.864	0.918	0.608
	Publisher-Emo	0.761	0.763	0.784	0.738	0.766	0.868	0.920	0.611
	ENDEF	0.765	0.766	0.779	0.751	0.768	0.865	0.918	0.618
LLM-based	GPT-3.5-turbo	0.725	0.734	0.774	0.676	0.702	0.813	0.884	0.519
	SuperICL	0.757	0.759	0.779	0.734	0.736	0.864	0.920	0.551
	ARG	0.784	0.786	0.804	0.764	0.790	0.878	0.926	0.653
	LLM-GAN	<u>0.804</u>	<u>0.806</u>	<u>0.812</u>	<u>0.796</u>	<u>0.823</u>	<u>0.896</u>	<u>0.934</u>	<u>0.712</u>
	GSFND	**0.880**	**0.899**	**0.875**	**0.884**	**0.914**	**0.925**	**0.945**	**0.883**
	Δ *Improve*	9.45%	11.54%	7.76%	11.06%	10.94%	3.24%	1.18%	24.02%

Table 2: Statistics of the fake news detection datasets.

#	Weibo21			GossipCop		
	Train	Val	Test	Train	Val	Test
Real	2,331	1,172	1,137	2,878	1,030	1,024
Fake	2,873	779	814	1,006	244	234
Total	5,204	1,951	1,951	3,884	1,274	1,258

model (LLM) for generating fake news (Sec. 4.1). Specifically, we utilize the ChatGLM-6B-scale version, as it supports inference and computation on consumer-grade GPUs (Cai et al., 2024). This makes our method more accessible and deployable on lightweight hardware for future applications.

5.3 Baselines

Most fake news detection methods rely on deep learning, with recent approaches incorporating LLMs. Therefore, we select 9 representative methods as baselines from both categories, including (*i*) **Deep-learning-based (DL-based) method:** BERT (Devlin et al., 2018), EANN (Wang et al., 2018a), Publisher-Emo (Zhang et al., 2021) and ENDEF (Zhu et al., 2022); (*ii*) **LLM-based method:** SuperICL (Xu et al., 2023), ARG (Hu et al., 2024), and LLM-GAN (Wang et al., 2024b). Additionally, we prompt GPT-3.5-turbo (OpenAI, 2022) via CoT (Wei et al., 2022) with few-shot demonstrations for fake news detection. Details of the baselines and evaluation metrics are provided in Appendices A.1 and A.2, respectively.

6 Result and Analysis

6.1 Overall Performance

From Tab. 1, we can see that our proposed method GSFND achieves the highest accuracy (Acc.) and outperforms most baselines in mac F1 and class-wise F1-scores. Notably, it improves the mac F1 by 9.45% on Weibo21 and 10.94% on GossipCop.(The effectiveness validation of our method across diverse LLM-based models and datasets is provided in Appendix A.4).

Detection on *fake* news. More importantly, our method significantly enhances *fake* news detection performance, improving F1-fake by 24.02% on GossipCop and 11.06% on Weibo21. While the baselines achieve promising results in detecting real news, their performance in fake news detection lags behind by as much as 30% compared to real news identification. At the same time, our method demonstrates robust performance in identifying real news as well. For example, it improves the F1-real by 7.76% on Weibo21. The experimental result shows that our method outperforms the baselines in detecting *fake* news and exhibits a more balanced ability to identify both real and fake news.

DL-based *vs.*LLM-based. In general, we observe that DL-based methods underperform LLMs-based methods on both datasets, except for directly prompting GPT-3.5-turbo which achieves the lowest results. This suggests that LLMs is helpful in providing pre-predicted labels (SuperICL), rationales (ARG), explanations (LLM-GAN) and augmented news (Ours) for fake news detection. However, LLM's few-shot prediction has limited performances (GPT-3.5-turbo).

6.2 Ablation Study

To demonstrate the effectiveness of different components in GSFND, we ablate LLM-generated fake news (G), Reinforced Sampling (RS) and Topic

Classification (TC).

Tab. 3 presents the result of ablating each component from GSFND. GSFND-X denotes removing X from GSFND. We observe that removing LLM-generated fake news (GSFND-G) results in the most significant performance drop across both datasets, with a greater drop on Weibo21. This suggests that incorporating a wider variety of fake news styles into the dataset is crucial for enhancing the model's ability to detect fake news, as it helps the model capture more distinctive features. Additionally, removing the reinforced sampling (GSFND-RS) causes the second most significant performance drop across both datasets, highlighting the importance of data distribution awareness for effective fake news detection. Finally, removing the topic classification (GSFND-TC) also causes a notable performance drop. We assume that considering topic semantics helps the RL agent in learning the optimal ratio across various topics, thereby benefiting overall fake news detection.

In short, these results confirm the importance of all components in our proposed GSFND for effective fake news detection.

Table 3: Evaluation results of ablating G, RS and TC from GSFND on both datasets.

Dataset	Method	mac F1	Acc.	F1-real	F1-fake
Weibo21	GSFND-G	0.634	0.681	0.653	0.615
	GSFND-RS	0.780	0.745	0.773	0.786
	GSFND-TC	0.789	0.769	0.783	0.794
	GSFND	**0.880**	**0.899**	**0.875**	**0.884**
GossipCop	GSFND-G	0.725	0.672	0.717	0.732
	GSFND-RS	0.816	0.763	0.801	0.831
	GSFND-TC	0.831	0.793	0.825	0.836
	GSFND	**0.914**	**0.925**	**0.945**	**0.883**

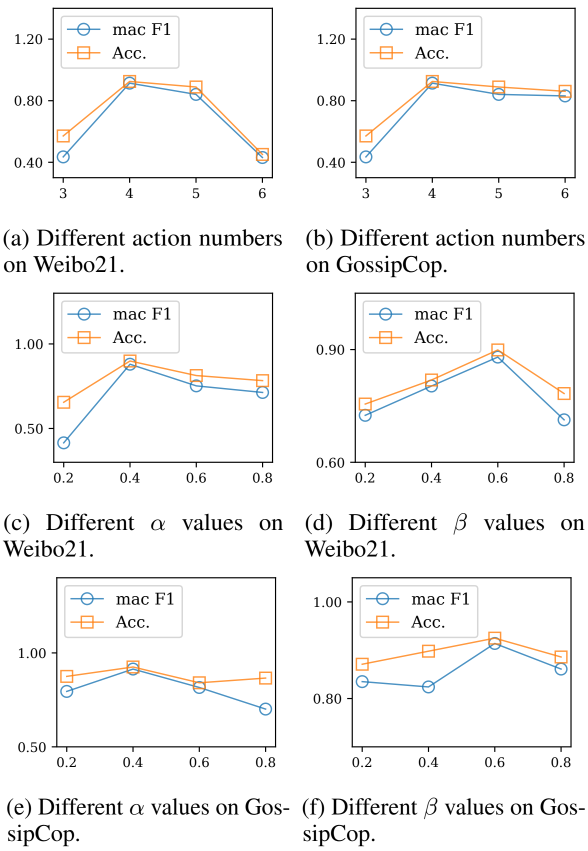
6.3 Sensitivity Analysis

The learned optimal real-to-fake ratio for GossipCop and Weibo21 are 1:0.5 and 1:1.5, respectively. We conjecture that this value is related to the noise level of the targeted news platforms. As a social media platform, Weibo features more informal and colloquial content, which may lead the model to determine a higher proportion of fake news necessary for training a more effective detector.

In addition, this section gives an in-depth analysis on how the hyperparameters, $|\mathcal{A}|$, α and β, affect the performance of GSFND.

Note that $|\mathcal{A}|$ represents the number of values that a can take, ranging from 0 to $|\mathcal{A}|-1$. While experimenting, we compute $\rho_{\text{fake}}(a^t) = \rho_{\min} + (\rho_{\max} - \rho_{\min}) \cdot \left(\frac{a^t}{a_{\max}}\right)^p \cdot \delta$ for $|\mathcal{A}| \neq 4$, where $\delta = 1 - \frac{|a^t-4|}{4}$ and $a_{\max} = |\mathcal{A}| - 1$. $\frac{a^t}{a_{\max}}$ normalizes the agent's action a^t to $[0, 1]$, ensuring flexible scaling. p adjusts the sensitivity of the sampling ratio to a^t, allowing fine control over changes. When a^t is near 4, λ is close to 1, which reduces the penalty. Using this equation is to keep the range of sampling proportions unchanged, still from $\rho_{\min} = 0.5$ to $\rho_{\max}$=2. Experiments on different sampling space sizes $|\mathcal{A}|$ (Figs. 4a and 4b) indicate that the optimal sampling space consists of four actions.

During training, we introduce two hyperparameters, α and β, in the reward function (Eq. 7). Experiments (Figs. 4c, 4e for α, and Figs. 4d, 4f for β) validate the importance of both *overall* and *class-specific* performance in guiding the model to determine the optimal ratio. Concretely, the model places slightly greater emphasis on balancing class-specific performance than on overall performance, with a weight distribution of 0.6 *vs.*0.4.



(a) Different action numbers on Weibo21.
(b) Different action numbers on GossipCop.
(c) Different α values on Weibo21.
(d) Different β values on Weibo21.
(e) Different α values on GossipCop.
(f) Different β values on GossipCop.

Figure 4: Effect of different $|\mathcal{A}|$, α and β. The y-axis represents mac F1 and Acc. The x-axis is the different values of $|\mathcal{A}|$, α and β.

7 Conclusion

This paper introduces GSFND, an effective *fake* news detector that demonstrates a more balanced ability to identify both real and fake news. GSFND first prompts a large language model (LLM) to augment the given fake news in various styles and then employs Reinforcement Learning to dynamically learn an optimal real-to-fake news ratio during the training process. Experimental results validate the effectiveness and superiority of our approach. These advantages enable GSFND to excel in *fake* news identification across multiple news platforms, thereby enhancing its real-world applicability.

Limitations

Despite the effectiveness of our proposed GSFND, there are several limitations. First, in this work, we do not discuss experimental results with alternative LLMs, such as variants of Llama and GPT. This is because ChatGLM has achieved satisfactory performance, and this module primarily demonstrates the concept of leveraging LLMs for fake news generation. Second, we constrain the RL-agent's actions (sampling proportions) to a limited set of discrete values. In future work, we aim to expand the action space to allow selection from a continuous range of values.

Acknowledgements

This work is supported by Open Foundation of Key Laboratory of Cyberspace Security, Ministry of Education of China (No.KLCS20240205), and the National Natural Science Foundation of China (NSFC) (Grant 62376265).

References

Tian Bian, Xi Xiao, Tingyang Xu, Peilin Zhao, Wenbing Huang, Yu Rong, and Junzhou Huang. 2020. Rumor detection on social media with bi-directional graph convolutional networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 549–556.

Christina Boididou, Katerina Andreadou, Symeon Papadopoulos, Duc Tien Dang Nguyen, Giulia Boato, Michael Riegler, Yiannis Kompatsiaris, et al. 2015. Verifying multimedia use at mediaeval 2015. In *MediaEval 2015*, volume 1436. CEUR-WS.

Fenglong Cai, Dong Yuan, Zhe Yang, and Lizhen Cui. 2024. Edge-llm: A collaborative framework for large language model serving in edge computing. In *2024 IEEE International Conference on Web Services (ICWS)*, pages 799–809. IEEE.

Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 675–684.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Marc Fisher, John Woodrow Cox, and Peter Hermann. 2016. Pizzagate: From rumor, to hashtag, to gunfire in dc. *Washington Post*, 6:8410–8415.

Team GLM, Aohan Zeng, Bin Xu, Bowen Wang, Chenhui Zhang, Da Yin, Dan Zhang, Diego Rojas, Guanyu Feng, Hanlin Zhao, et al. 2024. Chatglm: A family of large language models from glm-130b to glm-4 all tools. *arXiv preprint arXiv:2406.12793*.

Yimeng Gu, Ignacio Castro, and Gareth Tyson. 2024. Detecting multimodal fake news with gated variational autoencoder. In *Proceedings of the 16th ACM Web Science Conference*, pages 129–138.

Zhen Guo, Qi Zhang, Qisheng Zhang, Lance M Kaplan, Audun Jøsang, Feng Chen, Dong H Jeong, and Jin-Hee Cho. 2023. Detecting intents of fake news using uncertainty-aware deep reinforcement learning. In *2023 IEEE International Conference on Web Services (ICWS)*, pages 722–724. IEEE.

Yi Han, Shanika Karunasekera, and Christopher Leckie. 2020. Graph neural networks with continual learning for fake news detection from social media. *arXiv preprint arXiv:2007.03316*.

Beizhe Hu, Qiang Sheng, Juan Cao, Yuhui Shi, Yang Li, Danding Wang, and Peng Qi. 2024. Bad actor, good advisor: Exploring the role of large language models in fake news detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 22105–22113.

Linmei Hu, Tianchi Yang, Luhao Zhang, Wanjun Zhong, Duyu Tang, Chuan Shi, Nan Duan, and Ming Zhou. 2021. Compare to the knowledge: Graph neural fake news detection with external knowledge. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 754–763.

Diederik P Kingma. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Michael L Littman. 2015. Reinforcement learning improves behaviour from evaluative feedback. *Nature*, 521(7553):445–451.

Xiaoxiao Ma, Yuchen Zhang, Kaize Ding, Jian Yang, Jia Wu, and Hao Fan. 2024. On fake news detection with llm enhanced semantics mining. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 508–521.

Ahmadreza Mosallanezhad, Mansooreh Karami, Kai Shu, Michelle V Mancenido, and Huan Liu. 2022. Domain adaptive fake news detection via reinforcement learning. In *Proceedings of the ACM Web Conference 2022*, pages 3632–3640.

Salman Bin Naeem and Rubina Bhatti. 2020. The covid-19 'infodemic': a new front for information professionals. *Health Information & Libraries Journal*, 37(3):233–239.

Qiong Nan, Juan Cao, Yongchun Zhu, Yanyan Wang, and Jintao Li. 2021. Mdfend: Multi-domain fake news detection. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 3343–3347.

Qiong Nan, Qiang Sheng, Juan Cao, Beizhe Hu, Danding Wang, and Jintao Li. 2024. Let silence speak: Enhancing fake news detection with generated comments from large language models. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pages 1732–1742.

OpenAI. 2022. Chatgpt: Optimizing language models for dialogue. Accessed: 2023-08-13.

Ian Osband, Charles Blundell, Alexander Pritzel, and Benjamin Van Roy. 2016. Deep exploration via bootstrapped dqn. *Advances in neural information processing systems*, 29.

Sungwon Park, Sungwon Han, and Meeyoung Cha. 2024. Adversarial style augmentation via large language model for robust fake news detection. *arXiv preprint arXiv:2406.11260*.

Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2020. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big data*, 8(3):171–188.

Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, 19(1):22–36.

Jacob-Junqi Tian, Hao Yu, Yury Orlovskiy, Tyler Vergho, Mauricio Rivera, Mayank Goel, Zachary Yang, Jean-Francois Godbout, Reihaneh Rabbany, and Kellin Pelrine. 2024. Web retrieval agents for evidence-based misinformation detection. *arXiv preprint arXiv:2409.00009*.

Herun Wan, Shangbin Feng, Zhaoxuan Tan, Heng Wang, Yulia Tsvetkov, and Minnan Luo. 2024. Dell: Generating reactions and explanations for llm-based misinformation detection. *arXiv preprint arXiv:2402.10426*.

Bo Wang, Jing Ma, Hongzhan Lin, Zhiwei Yang, Ruichao Yang, Yuan Tian, and Yi Chang. 2024a. Explainable fake news detection with large language model via defense among competing wisdom. In *Proceedings of the ACM on Web Conference 2024*, pages 2452–2463.

Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao. 2018a. Eann: Event adversarial neural networks for multi-modal fake news detection. In *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*, pages 849–857.

Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao. 2018b. Eann: Event adversarial neural networks for multi-modal fake news detection. In *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*, pages 849–857.

Yaqing Wang, Weifeng Yang, Fenglong Ma, Jin Xu, Bin Zhong, Qiang Deng, and Jing Gao. 2020. Weak supervision for fake news detection via reinforcement learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 516–523.

Yifeng Wang, Zhouhong Gu, Siwei Zhang, Suhang Zheng, Tao Wang, Tianyu Li, Hongwei Feng, and Yanghua Xiao. 2024b. Llm-gan: Construct generative adversarial network through large language models for explainable fake news detection. *arXiv preprint arXiv:2409.01787*.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

Ke Wu, Song Yang, and Kenny Q Zhu. 2015. False rumors detection on sina weibo by propagation structures. In *2015 IEEE 31st international conference on data engineering*, pages 651–662. IEEE.

Yike Wu, Yang Xiao, Mengting Hu, Mengying Liu, Pengcheng Wang, and Mingming Liu. 2024. Towards robust evidence-aware fake news detection via improving semantic perception. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 16607–16618.

Zhiyuan Wu, Dechang Pi, Junfu Chen, Meng Xie, and Jianjun Cao. 2020. Rumor detection based on propagation graph neural network with attention mechanism. *Expert systems with applications*, 158:113595.

Canwen Xu, Yichong Xu, Shuohang Wang, Yang Liu, Chenguang Zhu, and Julian McAuley. 2023. Small models are valuable plug-ins for large language models. *arXiv preprint arXiv:2305.08848*.

Weizhi Xu, Junfei Wu, Qiang Liu, Shu Wu, and Liang Wang. 2022. Evidence-aware fake news detection with graph neural networks. In *Proceedings of the ACM web conference 2022*, pages 2501–2510.

Ruichao Yang, Xiting Wang, Yiqiao Jin, Chaozhuo Li, Jianxun Lian, and Xing Xie. 2022. Reinforcement subgraph reasoning for fake news detection. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2253–2262.

Feng Yu, Qiang Liu, Shu Wu, Liang Wang, Tieniu Tan, et al. 2017. A convolutional approach for misinformation identification. In *IJCAI*, pages 3901–3907.

Xueyao Zhang, Juan Cao, Xirong Li, Qiang Sheng, Lei Zhong, and Kai Shu. 2021. Mining dual emotion for fake news detection. In *Proceedings of the web conference 2021*, pages 3465–3476.

Yongchun Zhu, Qiang Sheng, Juan Cao, Shuokai Li, Danding Wang, and Fuzhen Zhuang. 2022. Generalizing to the future: Mitigating entity bias in fake news detection. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2120–2125.

A Appendix

A.1 Baseline Details

In this section, we introduce the details of the baselines.

- **BERT** (Devlin et al., 2018) is a widely adopted pre-trained deep learning model for natural language understanding, which we fine-tune specifically for the task of fake news detection.
- **EANN** (Wang et al., 2018a) leverages auxiliary signals and is designed to minimize the influence of event-related features. For the auxiliary task, the publication month is used as the label.
- **Publisher-Emo** (Zhang et al., 2021) integrates emotional analysis of news comments with textual features to enhance the performance of fake news detection.
- **ENDEF** (Zhu et al., 2022) employs causal learning to mitigate entity bias, thereby improving generalization on distribution-shifted fake news datasets.
- **SuperICL** (Xu et al., 2023) enhances language models by incorporating deep-learning-based plug-in modules, injecting both predictions and confidence levels into prompts for each test sample.
- **ARG** (Hu et al., 2024) introduces adaptive rationale guidance to advance deep-learning-based fake news detection, achieving state-of-the-art results.
- **LLM-GAN** (Wang et al., 2024b) utilizes prompting mechanisms to enable LLMs to act as both a Generator and a Detector, effectively addressing challenges in explainable fake news detection while improving both prediction accuracy and explanation quality.

A.2 Evaluation Metrics

For comparison purposes, we employ the macro F1 score (mac F1) and accuracy (Acc.) as our primary evaluation metrics. Additionally, to further analyze the models' performance in distinguishing between real and fake news, we report the F1 scores for each class separately (F1-real/F1-fake).

A.3 Dataset-Specific Topic Distributions

In our method, the purpose of introducing topic classification is to ensure that the Reinforced Sampling (RS) process will not bias towards any particular topics. Specifically, the topic classification model in our framework categorizes news into eight distinct topics. The details are shown in Table 4.

Table 4: Distribution of Topics Across Different Datasets

Topic	Weibo21	GossipCop	Twitter
Sports	234	376	542
Arts, Culture and Entertainment	864	684	1,567
Business	1,235	745	1,432
Health and Wellness	675	596	1,567
Lifestyle and Fusion	351	404	1,345
Science and Technology	468	489	2,456
Politics	764	269	2,453
Crime	1,623	321	2,560

A.4 Performance Evaluation of LLMs on Diverse Datasets

In addition to evaluating our method on Weibo21 and GossipCop, we have also tested its performance on the Twitter dataset (Boididou et al., 2015) and using another foundational LLM, Llama3-8B. The experimental results are summarized in the table below. From Tab. 5, our method demonstrates promising performance on the Twitter dataset and when using Llama3-8B. This highlights the generalizability of our approach. When using Llama3-8B as the foundational LLM, our method can still outperform the baselines by a large margin.

Table 5: Performance Comparison of Foundational LLMs on Different Datasets

Dataset	Foundational LLM	macF1	Acc	F1-real	F1-fake
Twitter	ChatGLM-6B	0.852	0.843	0.832	0.773
	Llama3-8B	0.821	0.831	0.821	0.762
Weibo21	ChatGLM-6B	0.880	0.899	0.875	0.884
	Llama3-8B	0.852	0.843	0.832	0.853
GossipCop	ChatGLM-6B	0.914	0.925	0.945	0.883
	Llama3-8B	0.889	0.896	0.904	0.871

A.5 Examples of Multi-Perspective News Generation

An Example from the Weibo21 Dataset (translated by GPT-4o)

Original News: Is ChaCha Sunflower Seeds Scamming Us? In the U.S., ChaCha sunflower seeds cost $1.99 per box, which is approximately 12.4 RMB per box. Each box contains 18 packs, meaning each pack costs about 0.68 RMB. In Chinese supermarkets, the same pack costs 6.9 RMB. This means that ChaCha sunflower seeds in U.S. supermarkets are 10 times cheaper than in China. Americans earn ten times the salary of Chinese people but buy products at one-tenth of the price. That's a 100-fold difference in purchasing power!

– –

Generated News

1. Rewritten News: In the U.S., the price of ChaCha sunflower seeds is $1.99, which is approximately 12.4 RMB. Each box contains 18 packs, making the price per pack $0.068. In Chinese supermarkets, the same pack costs 6.9 RMB. In other words, the price of ChaCha sunflower seeds in the U.S. is one-tenth of the price in China. However, Americans earn ten times the income of Chinese consumers, yet they purchase products at only one-tenth of the price in China. This results in a 100-fold difference in purchasing power.

2. Expanded News: In recent years, the controversy surrounding "Is ChaCha Sunflower Seeds Scamming Us?" has been widely discussed. In the U.S., ChaCha sunflower seeds are priced at $1.99, which is astonishingly low compared to China, where a single pack costs 6.9 RMB—ten times the price difference. This raises the question: Americans earn ten times the income of Chinese consumers, yet their purchasing power for the same product is only one-tenth. Is such a disparity reasonable?

In fact, this phenomenon is not unique. In recent years, many foreign goods have been priced higher in the Chinese market than in their country of origin. For instance, clothing, electronics, and daily necessities are often more expensive in China than abroad. The reasons behind this include not only tariffs but also transportation costs, brand premiums, and import taxes, among other factors.

However, from the consumer's perspective, whether such price differences are justified remains an open question. Take ChaCha sunflower seeds as an example: although they cost ten times more in China than in U.S. supermarkets, this does not necessarily mean that Chinese consumers are being exploited. Perhaps they are simply paying for better quality and taste. After all, price and quality are two crucial factors for any consumer.

Additionally, it is important to recognize that China's consumption power is steadily rising. As the economy grows and living standards improve, Chinese consumers' purchasing ability is also strengthening. Against this backdrop, price differences may continue to widen.

Overall, this issue cannot be judged in absolute terms. While respecting consumer choices, it is also necessary to regulate market pricing reasonably to protect consumer rights. At the same time, businesses should be encouraged to focus more on product quality and cost-effectiveness, ultimately providing consumers with better products and services.

3. Disguised News: Domestic Report: According to reports, a snack called ChaCha Sunflower Seeds is sold in the U.S. for only $1.99 per box, which is approximately 12.4 RMB. Each box contains 18 packs, making the price per pack around 6.8 cents. In Chinese supermarkets, however, the same pack costs 6.9 RMB. In other words, the price of this sunflower seed product in China is ten times higher than in the U.S.

However, while Americans earn ten times the income of Chinese consumers, they are purchasing the same product at only one-tenth of the price. This results in a staggering 100-fold difference in purchasing power.

An Example from the GossipCop Dataset

Original News: In February, Alicia Silverstone and Christopher Jarecki announced they were divorcing after 20 years together. The couple has a six - year - old son together, Bear Blu. (Photo : Frazer Harrison, Getty Images) Alicia Silverstone has filed for divorce from her husband of nearly 13 years, actor and musician Christopher Jarecki. The divorce papers were filed in Los Angeles County Superior Court on Friday, according to the Associated Press. The Clueless star, 41, had separated from Jarecki, 47, in February after more than 20 years together as a couple. At the time the couple said in a statement that "they still deeply love and respect each other and remain very close friends." The papers state the couple will share custody of their 7 - year - old son, Bear Blue"

LLM-generated News

1. Rewritten News: Alicia Silverstone has officially filed for divorce from her husband, Christopher Jarecki, after nearly 13 years of marriage. The filing took place in Los Angeles County Superior Court on Friday, as reported by the Associated Press. Silverstone, 41, and Jarecki, 47, had announced their separation in February following more than two decades together as a couple. The pair share a 7-year-old son, Bear Blu, and the divorce papers indicate they will share custody of him.

The Clueless star and her husband released a statement at the time of their separation, stating that they "still deeply love and respect each other and remain very close friends."

2. Expanded News: Actress Alicia Silverstone has officially filed for divorce from her husband, Christopher Jarecki, after nearly 13 years of marriage. The filing took place in Los Angeles County Superior Court on Friday, according to the Associated Press. This decision comes months after the couple announced their separation in February, marking the end of a relationship that spanned over two decades.

Silverstone, best known for her iconic role in the 1995 film Clueless, had been with Jarecki, an actor and musician, for more than 20 years, including nearly 13 years as a married couple. The two tied the knot in 2005 during an intimate lakeside ceremony in Lake Tahoe after dating for eight years. Together, they share a 7-year-old son, Bear Blu, who was born in 2011. The divorce documents confirm that Silverstone and Jarecki plan to share custody of their son.

In February, when the couple first announced their separation, they issued a joint statement addressing their decision to part ways. The statement emphasized the mutual respect and care they still hold for one another, saying, "They still deeply love and respect each other and remain very close friends." Despite their separation, they expressed their intention to continue co-parenting their child in a loving and supportive manner.

Silverstone, now 41, and Jarecki, 47, had often been seen as a low-profile yet strong Hollywood couple. Throughout their relationship, they shared a passion for healthy, eco-friendly living and veganism. Silverstone is a longtime advocate for veganism and sustainability, having authored books like The Kind Diet and The Kind Mama, while Jarecki is known for his involvement in the indie music scene, including his time as the lead vocalist for the punk band S.T.U.N.

The announcement of their split came as a surprise to fans, as the couple had largely kept their personal lives private and avoided public controversy. Their decision to remain friends and co-parent their son reflects their commitment to maintaining a positive relationship despite their separation.

For now, both Silverstone and Jarecki appear focused on moving forward while ensuring the well-being of their son, Bear Blu. Silverstone continues to remain active in her acting career and advocacy work, while Jarecki is pursuing his music and other creative endeavors. The divorce marks the end of a significant chapter in their lives, but their statement suggests they remain unified in their dedication to their family.

3. Disguised News: According to reports, Alicia Silverstone has filed for divorce from her husband of nearly 13 years, Christopher Jarecki, in Los Angeles County Superior Court. The couple, who had been together for over two decades, announced their separation earlier this year. In a joint statement at the time, they emphasized that they "still deeply love and respect each other" and remain close friends. The pair share a 7-year-old son, Bear Blu, and have agreed to share custody as part of their ongoing commitment to his upbringing.