Scaling Down, Powering Up:
RLHF-Enhanced Small LLMs for Healthcare Misinformation Detection

Anonymous ACL submission

Abstract

Healthcare misinformation poses a critical threat to public well-being, necessitating detection systems that are both accurate and computationally efficient. While large language models (LLMs) have demonstrated strong performance in misinformation detection, their deployment is often constrained by high resource requirements. In this work, we investigate the effectiveness of smaller LLMs (360M–3.8B parameters) using a three-stage framework comprising standardized prompting, supervised fine-tuning (SFT), and reinforcement learning from human feedback (RLHF). We evaluate seven LLMs across two benchmark datasets—FakeHealth and ReCOVery—and compare them against four larger LLMs (14B–72B) and five transformer-based baselines. For the RLHF stage, we study three policy optimization methods: Binary Classifier Optimization (BCO), Contrastive Preference Optimization (CPO), and our enhanced variant, CPO**. Empirical results demonstrate that while SFT improves domain adaptation, CPO** consistently achieves the best F1 performance, enabling small LLMs to rival or even outperform significantly larger counterparts. Our findings highlight the potential of RLHF techniques to close the performance gap, offering a scalable and cost-effective solution for real-world healthcare misinformation detection.

1 Introduction

Misinformation has become a pervasive challenge in the digital age, influencing public opinion (Cacciatore, 2021), threatening political stability (Jerit and Zhao, 2020), and undermining decision-making across various domains (Fernandez and Alani, 2018). The rapid spread of false or misleading content—particularly via social media— has made robust misinformation detection a critical research priority (Aïmeur et al., 2023). The urgency of this research was underscored especially

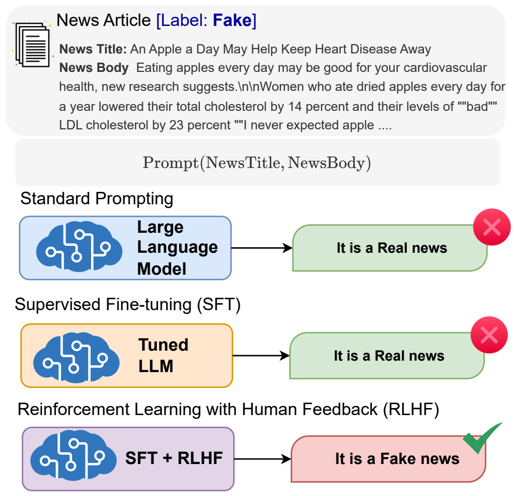

Figure 1: Role of RLHF with LLMs in misinformation detection. In this case, standard prompting the LLM fails to output a correct judgment of news veracity, and also a finetuned LLM; however, RLHF judges correctly.

during the COVID pandemic, when the widespread dissemination of false and misleading information undermined public trust, fueled vaccine skepticism, and in certain instances, pushed individuals toward extremist ideologies (Agbasiere, 2024).

Recent advances in natural language processing (NLP) have opened new avenues for combating misinformation using large language models (LLMs). These models, with their impressive linguistic capabilities, are increasingly being explored for their potential to judge the veracity of claims (Lucas et al., 2023; Huang and Sun, 2023; Wang et al., 2023; Irnawan et al., 2025). However, most prior research has focused on LLMs ranging from 70B to 340B parameters, overlooking smaller models (e.g., 1B to 3.5B) that are more practical for real-world deployment.

The challenge arises from the evolving, context-dependent nature of misinformation (Chen and Shu, 2024), especially in high-stakes fields like healthcare (Han et al., 2024), where claims often require domain knowledge and cultural understanding. This requires healthcare domain LLMs to be capable of understanding complex terminology and the cultural context of claims. This directs the research toward larger LLMs that can model language and capture intricate linguistic phenomena; they are also prone to hallucinations (Chen et al., 2024), or reflect training data biases (Chen and Shu, 2023), and more importantly, they are computationally expensive for deployment in sensitive settings (Wang et al., 2024). By contrast, smaller LLMs offer a path toward building reliable, cost-effective models that can scale in resource-constrained environments.

Healthcare systems are vulnerable to misinformation, especially during crises like pandemics, where healthcare practitioners may rely on online content to guide urgent decisions. If these sources are inaccurate, it can lead to harmful consequences in both clinical care and medical research. LLMs, with their ability to consume large volumes of knowledge, have the potential to act as an intermediate decision-support to help practitioners identify and navigate misleading content. Yet, this raises important questions: *How effective are LLMs at detecting misinformation?* And if they are effective, *how can we scale them down for practical use in real-world healthcare applications, given the computational costs of large models?*

While large-scale LLMs have gained significant attention for their performance, their practical deployment remains limited by resource demands (Prather et al., 2025). Our study takes a pragmatic approach: we explore the potential of smaller, more efficient LLMs and enhance them using reinforcement learning with human feedback (RLHF) (Ouyang et al., 2022). In particular, we contribute a refined variant of Contrastive Preference Optimization (CPO) (Xu et al., 2024), denoted CPO**, which introduces a log-based weighting mechanism to improve alignment with human preferences and factual accuracy. As illustrated in Figure 1, RLHF can yield more accurate judgments of news veracity than both standard prompting and SFT. To systematically investigate these challenges, we define three core research questions (RQs):

RQ1: How effective are smaller LLMs in detecting misinformation in healthcare? We evaluate *seven* small and *four* large LLMs (360M–72B parameters; see Table 1) on healthcare misinformation using a standardized prompting method. Our findings show that although small LLMs perform modestly under standardized prompting (e.g., Qwen2.5-0.5 achieves 45.8% F1 on FakeHealth), some models like LLaMA-3.2-1B outperform even larger models. This indicates that parameter count alone does not determine base performance.

LLM	Size
Qwen2.5 (Yang et al., 2024)	0.5B, 14B, 32B, 72B
Qwen3 (Yang et al., 2025)	0.6B
LLaMA-3 (Touvron et al., 2023)	1B, 70B
SmolLM2 (Allal et al., 2025)	360M, 1.7B
Falcon3 (Team, 2024)	3B
Phi-3.5-Mini (Abdin et al., 2024)	3.8B

Table 1: Summary of LLMs used.

RQ2: To what extent does supervised fine-tuning (SFT) enhance task-specific adaptation for misinformation detection? Building on the baseline, we apply SFT using a QLoRA-based parameter-efficient finetuning approach (Dettmers et al., 2023) to adapt the smaller models for domain-specific misinformation detection. Our findings show that while SFT boosts performance across most models (e.g., Falcon3 improves from 41.7% to 63.4% F1 on FakeHealth), gains vary widely. Some small models like Phi-3.5-Mini benefit significantly, while others like Qwen3 see limited improvement. Larger models, such as Qwen2.5-14B, show minimal gains, suggesting that architecture and pretraining (not just size) govern fine-tuning effectiveness.

RQ3: How does RLHF influence the performance of LLMs in detecting misinformation compared to SFT? Beyond SFT, we explore three RLHF strategies—Binary Classifier Optimization (BCO) (Jung et al., 2024), Contrastive Preference Optimization (CPO), and our novel refinement, CPO**—to investigate their comparative and cumulative impact on misinformation detection. Our proposed CPO** introduces a log-based weighting mechanism that stabilizes learning and better aligns model outputs with human preferences and factual correctness. Our findings show that RLHF significantly outperforms SFT across models and datasets (e.g., Qwen2.5-0.5 sees a +46% F1 gain on ReCOVery), with CPO** achieving the highest improvements. Notably, smaller models fine-tuned with RLHF often match or exceed the performance of much larger models, indicating that alignment strategy (not scale) is key to strong performance.

This study bridges the gap between resource-intensive large-scale LLMs and the practical needs of real-world applications by systematically comparing small and large models for healthcare misinformation detection. We evaluate SFT for task adaptation and explore RLHF to boost reliability, aiming to develop effective, scalable NLP solutions for high-stakes domains like healthcare.

2 Related Work

Recent advancements in misinformation detection have leveraged LLMs to develop more refined techniques for identifying misinformation. One such approach involves fine-tuning models like BERT (Kaliyar et al., 2021; Qin and Zhang, 2024; Farokhian et al., 2024; Yu et al., 2025; Kumari et al., 2021) with additional deep learning layers. However, model performance may be hindered by its inability to adapt to evolving misinformation patterns (Allcott et al., 2019). A growing body of work has explored more sophisticated strategies to address these limitations, such as domain adaptation (Mao et al., 2024) and leveraging uncertainty resolution techniques (Orlovskiy et al., 2024) to mitigate the challenges posed by ambiguous or incomplete health-related misinformation.

An alternative line of research explores direct LLM-based misinformation detection, using models such as BART (Lewis et al., 2020), GPT-3.5 (Achiam et al., 2023), LLaMA-2 (Touvron et al., 2023), LLaMA-3 (AI@Meta, 2024), Palm-2 (Anil et al., 2023), and Dolly-2 (Conover et al., 2023) for fact-checking, claim verification, and misinformation generation (Pavlyshenko, 2023; Huang and Sun, 2023; Wang et al., 2023; Lucas et al., 2023; Lai et al., 2024; Li et al., 2025; Irnawan et al., 2025; Leite et al., 2025). Pavlyshenko (2023) found that larger LLaMA-2 models (13B or 70B) improved detection performance when trained on extensive datasets. Similarly, Huang and Sun (2023) demonstrated that GPT-3.5 achieved strong performance, though its effectiveness could be further enhanced by incorporating richer contextual information. However, reliance on LLMs for misinformation detection introduces biases inherent in model training, raising concerns about fairness and reliability (Li et al., 2025). Wang et al. (2023) showed that GPT-3.5 struggled with COVID-19 misinformation detection due to a lack of specialized domain knowledge, underscoring the importance of domain adaptation. Additionally, Lucas et al. (2023) explored LLMs as both disinformation generators and detectors, achieving promising results but facing challenges in hallucination control. To mitigate these limitations, Hu et al. (2024) compared fine-tuned smaller models like BERT against GPT-3.5 and introduced an Adaptive Rationale Guidance network that integrates LLM-generated rationales to assist BERT in detecting misinformation. Yet, hallucinations persist, as model-generated rationales sometimes introduce misleading patterns, increasing false positives and negatives (Li et al., 2025).

There is growing interest in enhancing transparency and explainability in health misinformation detection, with studies leveraging crowd intelligence (Yang et al., 2023) and interpretable frameworks to refine predictions (Liu et al., 2024; Banerjee et al., 2024). Despite these advances, a key research gap remains in aligning models with human judgment while ensuring efficiency (Upadhyay et al., 2024), particularly in health misinformation detection, where domain knowledge is crucial. Kamali et al. (2024) investigated persuasive writing strategies to improve classification using fine-tuned BERT-family models and leveraged GPT-based models for prompt engineering. While Zarharan et al. (2024) explored explainability in public health misinformation detection and concluded that despite GPT-4 excels, open-source models (e.g., Falcon-180B (Almazrouei et al., 2023), LLaMA-70B) can match or even surpass it in few-shot and parameter-efficient fine-tuning settings.

Unlike prior methods that primarily depend on fine-tuning or prompting, our approach leverages RLHF to systematically align smaller LLMs with human judgment. This strategy addresses key limitations identified in existing research, such as model hallucinations, domain knowledge gaps, and biases inherent in large pretrained models. By integrating RLHF, we enhance the adaptability, reliability, and factual accuracy of smaller, computationally efficient LLMs, thereby overcoming challenges that fine-tuning and zero-shot prompting alone struggle to resolve.

3 Methodology

In this study, we evaluate the effectiveness of smaller LLMs for detecting healthcare misinformation through a three-stage methodology: standardized prompting (**SP**), supervised fine-tuning (**SFT**) (Pareja et al., 2024), and reinforcement learning from human feedback (**RLHF**) (Kaelbling et al., 1996; Christiano et al., 2017). As shown in Figure 2, we begin with SP, where models as-

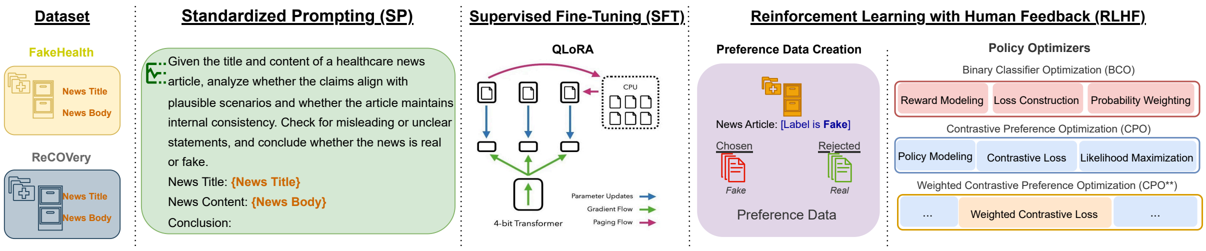

Figure 2: Illustration of the misinformation detection framework for healthcare.

sess healthcare news articles using both the title and full content to identify internal inconsistencies and flag potentially misleading claims. This step establishes a baseline for each model's zero-shot performance using two well-known datasets. In the second stage, we apply SFT to adapt models using labeled real and fake news articles. To prepare for reinforcement learning, we organize the training outputs into preference pairs—distinguishing between preferred (chosen) and undesirable (rejected) responses. Finally, we apply RLHF to align models more closely with human judgment, enhancing both factual accuracy and adherence to domain-specific values through policy optimization. Together, this pipeline enables a systematic evaluation of how prompting, fine-tuning, and RLHF can enhance misinformation detection in smaller, more efficient LLMs.

3.1 Standardized Prompting

For Standardized Prompting (SP), we designed a structured prompt to evaluate the plausibility, internal consistency, and clarity of healthcare news articles. As shown in Figure 2, each input $x = [x_{title}, x_{body}]$ is passed to the model using $SP(x) := Prompt(x_{title}, x_{body})$, instructing the LLM to analyze whether the claims are realistic, coherent, and unambiguous. The prompt guides the model to flag misleading or vague statements and ultimately classify the article as real or fake. The prompt explicitly guides the model to check for misleading or unclear statements before concluding whether the news is real or fake. This process hinges on four key aspects: (1) plausibility, by assessing alignment with known healthcare narratives (Tan et al., 2024); (2) internal consistency, by detecting contradictions or logical gaps (Dusmanu et al., 2017); (3) clarity, by identifying vague or ambiguous language (Guigon et al., 2024); and (4) decisiveness, requiring an explicit final judgment. This structured setup ensures consistent, interpretable assessments of model performance in misinformation detection.

3.2 Supervised Fine-Tuning

Supervised Fine-Tuning (SFT) with Quantized Low-Rank Adaptation (QLoRA) (Dettmers et al., 2023) was used to efficiently fine-tune LLMs while reducing memory and computational costs. QLoRA applies Parameter Efficient Fine-Tuning (PEFT) (Xu et al., 2023), using low-rank approximations of weight matrices, where $W \approx A \cdot B^T$, reducing the number of parameters while maintaining performance. The fine-tuning process is guided by a standardized prompt $(SP(x), y) : SP(x) \rightarrow y$, and optimized using the $\mathcal{L}$ loss function, defined as $\mathcal{L}(\theta) = \frac{1}{N} \sum_{i=1}^{N} [\hat{y}_i \log(y_i) + (1 - \hat{y}_i) \log(1 - y_i)]$ where $\hat{y}_i$ are the predictions, y_i are the true labels, and N is the total number of samples. This method ensures computational efficiency while maintaining model effectiveness in tasks like claim verification.

3.3 Reinforcement Learning

Preference Data Creation. To fine-tune the model using RLHF, we constructed a preference dataset designed to guide the policy optimization process. The dataset consists of structured interactions where the model receives prompts based on standardized templates and generates responses that are explicitly ranked for preference learning. We employed two formats for preference data collection, one for Binary Classifier Optimization (BCO) (Jung et al., 2024) and the other for Contrastive Preference Optimization (CPO) (Xu et al., 2024). In the BCO format, we constructed both positive and negative completions for each instance. Given a news x, the standardized prompt $SP(x)$ was structured. The model's response y was labeled as preferred (True) when aligned with the original ground-truth label and non-preferred (False) when intentionally flipped to the opposite class. The BCO formated data defined as:

$\mathcal{D}_{\text{BCO}} = \{(SP, y^+, L^+), (SP, y^-, L^-)\}$, where y^+ and y^- represent the correct and incorrect completions, and $L^+ = 1$, $L^- = 0$ denote preference labels. In contrast, the CPO format explicitly pairs the correct and incorrect completions under a chosen vs. rejected paradigm. Each sample contains a preferred response y^+ and a rejected response y^-, structured as $\mathcal{D}_{\text{CPO}} = \{(SP, y^+, y^-)\}$. This format allows the model to directly learn to differentiate between correct and incorrect outputs, refining its response ranking capabilities. The structured preference data thus enables fine-tuning through RLHF by optimizing the policy to maximize reward based human-aligned feedback signals.

Binary Classifier Optimization (BCO). The BCO framework (Jung et al., 2024) provides an efficient method for aligning LLMs using binary feedback signals rather than comparative preference-based ranking. In the context of healthcare misinformation detection, BCO enables models to learn directly from binary evaluations of content accuracy. Given a news data $\mathcal{D}_{\text{BCO}}$, the binary feedback enables the model to iteratively refine its classification function $f_\theta(x)$ by minimizing the binary cross-entropy loss $\mathcal{L}_{BCO}(\theta)$.

$$\begin{aligned}\mathcal{L}_{BCO}(\theta) = &- \mathbb{E}_{(x,y)\sim D^+} \left[\log \sigma(r_\theta(x,y) - \delta)\right] \\ &- \mathbb{E}_{(x,y)\sim D^-} \left[\frac{p_\psi(f=1|x)}{p_\psi(f=0|x)} \log \sigma(-(r_\theta(x,y) - \delta))\right]\end{aligned}$$

Where σ is the sigmoid function, $r_\theta(x,y)$ represents the reward function parameterized by θ, δ is a margin term, and $p_\psi(f=1|x)$ and $p_\psi(f=0|x)$ denote the probabilities of correct and incorrect classifications, respectively. By leveraging binary supervision, BCO enables efficient preference-based optimization, making it particularly useful for mitigating misinformation in healthcare.

Contrastive Preference Optimization (CPO). The CPO (Xu et al., 2024) directly optimizes policy preference rankings by leveraging contrastive learning. The model aims to distinguish between preferred (*chosen*) and less preferred (*rejected*) outputs, refining its response generation toward more accurate and reliable completions. This is particularly critical in misinformation detection, where incorrect classifications can lead to harmful consequences. Given a news data $\mathcal{D}_{CPO}(\theta)$, CPO objective is: $\mathcal{L}_{CPO}(\theta) = \min_\theta \underbrace{\mathcal{L}(\pi_\theta, U)}_{\mathcal{L}_{\text{prefer}}} \underbrace{-\mathbb{E}_{(x,y^+)\sim\mathcal{D}} \left[\log \pi_\theta(y^+ \mid x)\right]}_{\mathcal{L}_{\text{NLL}}}$,

where, $\mathcal{L}_{\text{NLL}}$ is the negative log-likelihood (NLL) (Rafailov et al., 2023) loss term that maximizes the likelihood of preferred outputs, and π_θ is the model's policy. Moreover, $\mathcal{L}(\pi_\theta, U)$ represents the behavior cloning (BC) regularizer (Hejna et al., 2024) using Kullback–Leibler (KL) divergence and is defined as: $\mathcal{L}(\pi_\theta; U) = -\mathbb{E}_{(x,y^+,y^-)\sim\mathcal{D}} \left[\log \sigma\left(\beta \log \pi_\theta(y^+|x) - \beta \log \pi_\theta(y^-|x)\right)\right]$, where $\sigma(\cdot)$ is the sigmoid function, ensuring the preference ranking is learned in a probabilistic manner, $\pi_\theta(y^+|x)$ and $\pi_\theta(y^-|x)$ represent the model's probability distribution over responses given the input $x := SP(x)$, and β is a scaling factor controlling the contrastive margin.

Weighted Contrastive Preference Optimization (CPO).** In the CPO, behavior cloning (BC) aligns a model's predictions with reference behavior, minimizing the divergence between the model's policy and the expert's demonstrations. The KL divergence measures this dissimilarity, guiding the model to emulate the expert's actions closely. The objective is to ensure that the model's policy approximates the expert's policy, promoting accurate and reliable outputs. To stabilize the optimization process and prevent overfitting, we introduce a log-based weight term $W = -\log(\exp(L) + \epsilon)$, where $L = \log(\pi_\theta(y^+|x)) - \log(\pi_\theta(y^-|x))$ is the difference between the log-probabilities of chosen and rejected responses, and ϵ is a small constant to avoid numerical issues. This term encourages the model to favor chosen responses over rejected ones by penalizing large deviations, akin to the behavior enforced by KL divergence. Additionally, the exponential function within the term serves as an unnormalized probability ratio, and applying the logarithm helps mitigate abrupt gradient fluctuations, leading to more stable training. By applying W to $\mathcal{L}(\pi_\theta; U)$ KL-based behavior cloning, we enhance preference learning through $\mathcal{L}(\pi_\theta; U) = \mathcal{L}(\pi_\theta; U) \times W$. This formulation combines the strengths of preference optimization and behavior cloning regularization, effectively addressing challenges related to biased outputs and sampling inefficiencies (Xu et al., 2024), promoting robustness in preference learning, while ensuring the model generates high-quality responses.

4 The Framework Evaluation

4.1 Experimental Setup

Experimental Datasets. We use two publicly available healthcare misinformation datasets in this study: FakeHealth(Dai et al., 2020) and ReCOVery(Zhou et al., 2020), both of which contain la-

	FakeHealth			ReCOVery		
	Real	Fake	Total	Real	Fake	Total
Train	1,040	529	1,569	1,022	499	1,521
Test	346	177	523	342	166	508

Table 2: Details of datasets.

	FakeHealth			ReCOVery		
	Prec	Rec	F1	Prec	Rec	F1
BERT	65.8	64.6	65.0	91.2	87.0	88.7
FakeNews	59.7	59.8	59.8	83.1	82.6	82.8
ALBERT	62.1	60.3	60.7	90.1	90.1	90.1
Flan-T5	33.0	50.0	39.8	84.6	54.5	49.2
ELECTRA	33.0	50.0	39.8	88.8	82.2	84.4

Table 3: Results of transformer-based models. The FakeNews model is refers to a domain-specific fine-tuned BERT (`https://huggingface.co/dhruvpal/fake-news-bert`).

beled real and fake health-related news articles and claims. The ReCOVery dataset focuses on COVID-19 misinformation, including 1,364 real and 665 fake claims sourced from fact-checking platforms and authoritative health agencies. It was constructed by analyzing content from 2,000 news publishers and selecting 60 with extreme credibility scores to ensure accurate labeling. The FakeHealth dataset contains two subsets: Story (1,078 real / 420 fake) and Review (308 real / 286 fake). We combine these subsets to form a more diverse and challenging benchmark. Articles no longer accessible online were excluded to maintain data consistency. For both datasets, we apply a 75%-25% train-test split. Detailed statistics are presented in Table 2.

Experimental Models. We evaluated five model variants to compare the effectiveness of prompting, SFT, and RL strategies: (1) **SP**, a prompting-only baseline using the base model without task-specific adaptation. (2) **SFT**, a supervised fine-tuning variant trained using QLoRA. (3) **+ BCO**, an SFT model further trained with RLHF using BCO. (4) **+ CPO**, an SFT model enhanced with RLHF using CPO. (5) **+ CPO****, an SFT model fine-tuned with RLHF using an improved CPO algorithm that incorporates a log-based weighted loss to better align with human preferences.

4.2 Results

RQ1: How effective are smaller LLMs in detecting misinformation in healthcare? Baseline evaluation using SP is represented in Table 4 for seven small LLMs and Table 5 for different larger LLMs, revealed considerable variability in the ability of smaller LLMs to detect misinformation.

	FakeHealth				ReCOVery			
	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1
Qwen2.5 (0.5B)								
SP	46.4	48.1	47.9	45.8	49.4	51.6	51.8	48.8
SFT	59.0	55.9	56.3	55.9	51.9	48.7	48.6	48.3
+ BCO	72.2	69.3	64.6	65.4	96.6	97.0	95.3	96.1
+ CPO	71.7	68.4	68.6	68.5	92.7	93.4	89.9	91.3
+ CPO**	73.0	69.9	69.9	69.9	95.0	94.7	94.0	94.3
Falcon3 (3B)								
SP	43.0	43.4	42.7	41.7	39.3	58.5	53.5	36.1
SFT	69.5	65.3	62.9	63.4	94.0	93.6	92.8	93.2
+ BCO	72.4	69.1	66.4	67.2	98.0	98.3	97.1	97.7
+ CPO	74.1	71.1	68.9	69.6	94.6	94.7	93.1	93.8
+ CPO**	74.9	72.1	69.7	70.5	95.6	95.0	95.0	95.0
LLaMA-3.2 (1B)								
SP	63.2	44.8	48.7	41.9	64.9	53.9	51.8	49.0
SFT	65.9	60.0	57.5	57.5	85.2	84.1	81.4	82.5
+ BCO	72.8	70.5	64.7	65.5	96.0	95.6	95.3	95.5
+ CPO	71.8	70.2	62.1	62.4	95.4	96.2	93.5	94.7
+ CPO**	74.1	72.0	66.9	67.9	96.0	95.5	95.5	95.5
Phi-3.5-Mini (3.8B)								
SP	64.8	42.0	49.2	40.3	70.0	73.3	55.1	51.1
SFT	71.3	67.6	66.3	66.7	93.5	92.2	93.1	92.6
+ BCO	70.1	66.1	64.0	64.6	97.8	98.0	96.9	97.5
+ CPO	76.0	73.3	73.3	73.3	95.4	95.6	94.0	94.7
+ CPO**	76.8	74.1	73.4	73.7	97.0	97.5	95.7	96.5
Qwen3 (0.6B)								
SP	66.1	58.1	50.1	40.3	68.1	83.9	55.2	42.7
SFT	65.9	33.0	49.8	39.7	67.1	33.6	49.8	40.1
+ BCO	66.5	83.2	50.5	41.0	85.4	88.0	78.9	81.5
+ CPO	74.7	71.7	69.6	70.3	95.6	95.4	94.6	95.0
+ CPO**	72.4	69.1	68.7	68.9	96.2	95.6	95.8	95.7
SmolLM2 (1.7B)								
SP	65.7	33.0	49.7	39.6	66.7	33.5	49.5	40.0
SFT	52.7	49.1	49.1	49.0	66.7	43.5	49.7	40.5
+ BCO	74.1	73.4	65.4	66.3	96.0	95.6	95.3	95.5
+ CPO	75.7	73.2	70.3	71.2	95.6	95.4	94.6	95.0
+ CPO**	76.4	74.1	71.1	72.1	96.8	96.7	96.1	96.3
SmolLM2 (360M)								
SP	60.8	44.0	47.6	42.6	66.3	58.2	54.6	53.4
SFT	65.2	54.5	51.2	45.7	67.3	58.8	50.6	42.4
+ BCO	67.4	62.2	55.9	54.3	95.0	95.5	93.2	94.2
+ CPO	72.0	68.5	66.6	67.2	94.4	93.9	93.4	93.6
+ CPO**	73.9	71.1	67.9	68.8	95.1	94.3	94.4	94.4

Table 4: Experimental results of LLMs. The blue color represents the best performance, while orange represents the second-best performance.

Overall Effectiveness of Small LLMs. According to the Table 4, while smaller models demonstrated some capacity to distinguish between factual and misleading content, their raw performance was limited. For example, Qwen2.5 (0.5B) achieved an F1 of 45.8 on FakeHealth and 48.8 on ReCOVery—comparable to or better than some larger models (see Table 5). Interestingly, LLaMA-3.2 (1B) showed competitive performance despite its modest size, with F1 scores of 41.9% (FakeHealth) and 49.0% (ReCOVery), respectively. This may be attributed to its extended larger context window, which enables better comprehension of long-form content. Performance did not scale linearly with model size. For instance, Falcon3 (3B) lagged significantly in both datasets, suggesting that architectural differences and pretraining quality are just as important as parameter count.

Domain-Specific Precision Analysis. According to

	FakeHealth				ReCOVery			
	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1
Previous Works								
GPT-3.5	-	-	-	-	-	96.4	93.9	95.0
LLaMA-3 (8B)	-	-	-	-	-	96.1	94.5	95.2
Standardized Prompting of Larger LLMs								
Qwen2.5 (32B)	66.3	61.8	50.6	41.9	77.1	77.9	67.8	69.3
LLaMA-3.3 (70B)	66.3	63.2	50.5	41.4	75.9	78.6	65.1	66.1
Qwen2.5 (72B)	65.9	49.7	49.9	40.2	76.5	77.0	67.1	68.5
Qwen2.5 (14B)								
SP	66.1	58.1	50.4	41.3	75.1	72.0	68.7	69.7
SFT	66.1	33.0	50.0	39.8	72.8	75.1	60.1	59.4
+ BCO	76.4	74.0	71.4	72.6	98.0	98.2	97.2	97.7
+ CPO	76.0	73.5	74.6	73.9	97.8	98.2	96.8	97.5
+ CPO**	78.3	75.9	75.1	75.4	97.8	97.9	97.1	97.5

Table 5: Larger LLMs experimental results. The GPT-3.5 LLM has been explored by Wang et al. (2023), and LLaMA-3 (8B) by Irnawan et al. (2025).

the precision scores in the medical domain, small LLMs generally underperform compared to larger LLMs. For instance, the best precision obtained by Qwen3 (0.6B) on FakeHealth was 58.1%, whereas LLaMA-3.3 (70B) achieved 63.2%. In contrast, on ReCOVery, Qwen3 (0.6B) outperformed LLaMA-3.3 by approximately 3.5%, achieving 83.9%.

These findings suggest that while smaller LLMs can detect misinformation, they struggle with context-dependent claims. This highlights the importance of domain adaptation strategies to improve their effectiveness, even for larger LLMs.

RQ2: To what extent does SFT enhance task-specific adaptation for misinformation detection? To assess SFT's impact, we analyze how different models adapt to domain-specific misinformation detection after fine-tuning.

Smaller LLMs. As we can see in Table 4, Falcon3, which performed poorly in SP (F1 of 41.7% on FakeHealth and 36.1% on ReCOVery), shows the most dramatic improvement with SFT, reaching F1 of 63.4% on FakeHealth and 93.2% on ReCOVery. Moreover, LLaMA-3.2, Phi-3.5-Mini, and SmolLM2, which already exhibited moderate performance in SP, also see notable improvements. Phi-3.5-Mini achieves F1 of 66.7% on FakeHealth and 92.6% on ReCOVery after fine-tuning, indicating that SFT helps these models leverage their parameter size and context length more effectively. However, Qwen3 (0.6B) struggles in terms of precision and F1-scores in both datasets, showing that even with SFT, some models may not benefit substantially. Similarly, SmolLM2 (360M)—the smallest model—improves from 42.6% to 45.7% on FakeHealth, but fails to gain on ReCOVery.

Larger LLMs. Unfortunately, a similar improvement pattern was not observed with the larger LLM Qwen2.5 (14B) after SFT (see Table 5). This suggests that SFT may provide diminishing returns for some large models, potentially due to their pretraining objectives, parameter saturation, or optimization difficulties during fine-tuning. However, alternative fine-tuning strategies, such as those proposed by Wang et al. (2023) for GPT-3.5 and by Irnawan et al. (2025) for LLaMA-3-8B, substantially outperform both SP and standard SFT approaches. This suggests that conventional SFT alone may not fully unlock the potential of LLMs, and size alone is not the main driver of model performances.

Transformer-Based Models. When comparing across architectures, transformer-based models generally benefited from SFT (see Table 3). However, even these models did not surpass Phi-3.5-Mini (3.8B), indicating that fine-tuning effectiveness depends not just on architecture or scale, but also on pretraining quality and task alignment. Notably, traditional transformers like BERT struggled with longer inputs, often exceeding their token limits. In contrast, LLMs with extended context windows handled such content more effectively, underscoring key limitations of earlier transformer models in real-world misinformation detection.

In summary, SFT proves to be a valuable adaptation method, especially for smaller and underperforming models, though its effectiveness is highly dependent on model architecture, pretraining strategy, and task alignment. This hinders the need for more advanced fine-tuning approaches to further boost performance across models.

RQ3: How does RLHF influence the performance of LLMs in detecting misinformation compared to SFT? To answer this RQ, the empirical evaluation of small LLMs presented in Table 4 and larger LLMs shown in Table 5 for +BCO, +CPO, and +CPO** strategies.

Overall RLHF Benefits Across LLMs. RLHF consistently outperforms SFT across both FakeHealth and ReCOVery datasets. According to Table 4, for all tested models, adding RLHF led to notable F1 gains. For instance, with the Qwen2.5 (0.5B) model, the CPO** approach achieved a +14% F1 increase (from 55.9% to 69.9%) on FakeHealth, and an impressive +46% gain (from 48.3% to 94.3%) on ReCOVery compared to SFT. As an another example, the most smallest LLM, the SmolLM2 (360M) model, using the CPO** achieved a +23% F1 increase (from 45.7% to 68.8%) on Fakehealth, and 52% gain (42.4% to 94.4%) on ReCOVery compared to SFT. These improvements evident that RLHF enhances LLM performances, especially for underperforming SFT models. A similar pattern was also observed in larger LLMs. As shown in Table 5, Qwen2.5 (14B) improves from 39.8% (SFT) to 75.4% (CPO**) on FakeHealth. It also reaches 97.5% F1 on ReCOVery, outperforming GPT-3.5 (95.0%) and LLaMA-3-8B (95.2%) from previous works.

*Superiority of CPO***. Among the RLHF methods, CPO** outperformed both BCO and CPO across nearly all models and datasets. In small LLMs, as we can see in Table 4, the Phi-3.5-Mini (3.8B), the CPO** attained the highest F1 score of 73.7% for FakeHealth and 96.5% for ReCOVery datasets. The next LLM that stood out in the small LLM category is SmolLM2 (1.7B), with F1-score of 72.1% with CPO** on FakeHealth and 96.3% on ReCOVery. Across 8 LLMs evaluated on 2 datasets each (yielding a total of 16 experiments), the CPO** method achieved the highest F1 score in 10 out of the 16 cases, showcasing that the log-based weighting mechanism can facilitate more stable training and better alignment.

Performance Convergence Across Model Sizes. RLHF—particularly CPO—dramatically narrows the performance gap between small and large models. For example, SmolLM2 (1.7B) with CPO achieves 96.3% F1 on ReCOVery, surpassing LLaMA-3 (8B) (95.2%) and even Qwen2.5 (14B) (97.5), which is just a 1.2% difference despite an 8x size gap. Likewise, smaller models such as Qwen3 (0.6B) and LLaMA-3.2 (1B), when fine-tuned with CPO**, outperform standard prompting of LLaMA-3 (70B) on FakeHealth and ReCOVery. These results show that model scale alone no longer guarantees superior performance. With high-quality alignment strategies like BCO, CPO, or CPO**, smaller models become competitive alternatives—offering strong task performance with significantly lower computational cost and better deployment feasibility.

Our findings highlight RLHF as a powerful tool for improving misinformation detection across LLM scales. It not only boosts individual performance but also bridges the capability divide between small and large models, making small LLMs viable alternatives for real-world deployment.

5 Discussions

5-Fold Cross-Validation. Table 6 summarizes the 5-fold cross-validation results for Phi-3.5-Mini fine-tuned with CPO**. The model demonstrates strong and consistent performance, particularly on the ReCOVery dataset, achieving 97.0% F1 score, with an $F1_{fake}$ of 96.0%. While performance on FakeHealth is comparatively lower (F1: 71.6, $F1_{fake}$: 61.9), the results still reflect meaningful gains from CPO**. Overall, these findings highlight the reliability and robustness of our RL-based fine-tuning approach for enhancing misinformation detection in smaller LLMs.

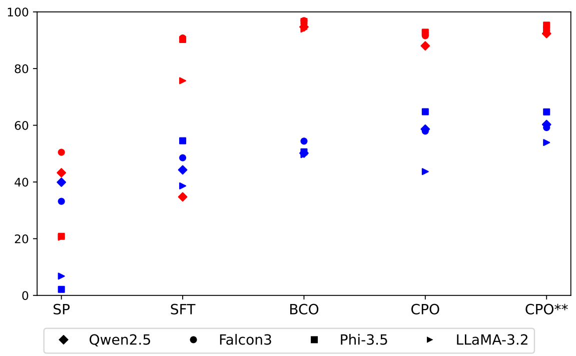

Figure 3: $F1_{\text{fake}}$ for ReCOVery and FakeHealth datasets.

	Acc	Prec	Rec	F1
FakeHealth	75.0	72.0	71.4	71.6
ReCOVery	97.4	97.5	96.5	97.0

Table 6: 5-Fold Cross Validations using SFT + CPO** for Phi-3.5-Mini LLM.

RLHF Impact on Fake Claims. The Figure 3 shows the F1 scores for the "fake" class across LLMs and setups on both datasets. CPO** consistently outperforms other methods, especially on ReCOVery, highlighting its strength in fake news detection. While BCO performs stably across models, CPO shows more variance. The log-based weighting in CPO** helps stabilize RLHF fine-tuning, enhancing both performance and robustness over BCO. Although SFT improves over SP, it remains highly variable in detecting "fake" claims.

6 Conclusion

We showed that smaller LLMs, when enhanced with RL can effectively detect healthcare misinformation. Evaluated on FakeHealth and ReCOVery, these models outperformed SFT and approached larger-model performance with lower computational cost. This suggests a promising path for efficient, real-world misinformation detection.

Limitations

While our study demonstrates that small and mid-sized LLMs, enhanced through RLHF, can achieve competitive performance in healthcare misinformation detection, it is not without limitations. First, our evaluation is restricted to two healthcare-specific datasets—FakeHealth and ReCOVery. Although these benchmarks are well-established, they do not reflect the full diversity of misinformation found in broader domains such as finance, politics, or climate science. Future work should incorporate additional datasets to assess cross-domain generalizability. Second, while we applied a standardized prompting strategy across models, we did not conduct extensive prompt engineering or instruction tuning. This may have limited the performance ceiling, particularly for models in the SP stage. As prompt sensitivity can significantly impact LLM behavior—especially in smaller architectures—more systematic prompt optimization could yield further improvements. Third, due to the computational demands of large-scale experimentation, we focused on three RLHF strategies—BCO, CPO, and our proposed CPO**—that are well-suited for factual alignment. However, other RLHF variants such as DPO (Rafailov et al., 2023), PPO (Schulman et al., 2017), or reward modeling could offer additional insights and performance gains if explored in future work. Finally, our study is limited to a binary misinformation classification setting. While this is a practical and common formulation, fine-grained misinformation detection—e.g., categorizing claims by severity, intent, or harm—could provide richer insights and more actionable outputs. We hope this empirical study lays the groundwork for such future directions by demonstrating the viability of small LLMs in high-stakes domains like healthcare.

References

Marah Abdin, Jyoti Aneja, Hany Awadalla, Ahmed Awadallah, Ammar Ahmad Awan, Nguyen Bach, Amit Bahree, Arash Bakhtiari, Jianmin Bao, Harkirat Behl, and 1 others. 2024. Phi-3 technical report: A highly capable language model locally on your phone. *arXiv preprint arXiv:2404.14219*.

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, and 1 others. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Chinyere L. Agbasiere. 2024. From viral rumors to fact-checked information: The influence of social media platforms and fact-checking organizations on public trust during the covid-19 crisis. *ProQuest Dissertations and Theses*, page 111. Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2024-12-12.

AI@Meta. 2024. Llama 3 model card.

Esma Aïmeur, Sabrine Amri, and Gilles Brassard. 2023. Fake news, disinformation and misinformation in social media: a review. *Social Network Analysis and Mining*, 13(1):30.

Loubna Ben Allal, Anton Lozhkov, Elie Bakouch, Gabriel Martín Blázquez, Guilherme Penedo, Lewis Tunstall, Andrés Marafioti, Hynek Kydlíček, Agustín Piqueres Lajarín, Vaibhav Srivastav, and 1 others. 2025. Smollm2: When smol goes big–data-centric training of a small language model. *arXiv preprint arXiv:2502.02737*.

Hunt Allcott, Matthew Gentzkow, and Chuan Yu. 2019. Trends in the diffusion of misinformation on social media. *Research & politics*, 6(2):2053168019848554.

Ebtesam Almazrouei, Hamza Alobeidli, Abdulaziz Alshamsi, Alessandro Cappelli, Ruxandra Cojocaru, Mérouane Debbah, Étienne Goffinet, Daniel Hesslow, Julien Launay, Quentin Malartic, and 1 others. 2023. The falcon series of open language models. *arXiv preprint arXiv:2311.16867*.

Rohan Anil, Andrew M Dai, Orhan Firat, Melvin Johnson, Dmitry Lepikhin, Alexandre Passos, Siamak Shakeri, Emanuel Taropa, Paige Bailey, Zhifeng Chen, and 1 others. 2023. Palm 2 technical report. *arXiv preprint arXiv:2305.10403*.

Tanushree Banerjee, Richard Zhu, Runzhe Yang, and Karthik Narasimhan. 2024. Llms are superior feedback providers: Bootstrapping reasoning for lie detection with self-generated feedback. *arXiv preprint arXiv:2408.13915*.

Michael A Cacciatore. 2021. Misinformation and public opinion of science and health: Approaches, findings, and future directions. *Proceedings of the National Academy of Sciences*, 118(15):e1912437117.

Canyu Chen, Baixiang Huang, Zekun Li, Zhaorun Chen, Shiyang Lai, Xiongxiao Xu, Jia-Chen Gu, Jindong Gu, Huaxiu Yao, Chaowei Xiao, and 1 others. 2024. Can editing llms inject harm? *arXiv preprint arXiv:2407.20224*.

Canyu Chen and Kai Shu. 2023. Can llm-generated misinformation be detected? *arXiv preprint arXiv:2309.13788*.

Canyu Chen and Kai Shu. 2024. Combating misinformation in the age of llms: Opportunities and challenges. *AI Magazine*, 45(3):354–368.

Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. 2017. Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, 30.

Mike Conover, Matt Hayes, Ankit Mathur, Jianwei Xie, Jun Wan, Sam Shah, Ali Ghodsi, Patrick Wendell, Matei Zaharia, and Reynold Xin. 2023. Free dolly: Introducing the world's first truly open instruction-tuned llm.

Enyan Dai, Yiwei Sun, and Suhang Wang. 2020. Ginger cannot cure cancer: Battling fake health news with a comprehensive data repository. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 853–862.

Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. 2023. Qlora: Efficient finetuning of quantized llms. *Advances in neural information processing systems*, 36:10088–10115.

Mihai Dusmanu, Elena Cabrio, and Serena Villata. 2017. Argument mining on twitter: Arguments, facts and sources. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2317–2322.

Mahmood Farokhian, Vahid Rafe, and Hadi Veisi. 2024. Fake news detection using dual bert deep neural networks. *Multimedia Tools and Applications*, 83(15):43831–43848.

Miriam Fernandez and Harith Alani. 2018. Online misinformation: Challenges and future directions. In *Companion proceedings of the the web conference 2018*, pages 595–602.

Valentin Guigon, Marie Claire Villeval, and Jean-Claude Dreher. 2024. Metacognition biases information seeking in assessing ambiguous news. *Communications Psychology*, 2(1):122.

Tianyu Han, Sven Nebelung, Firas Khader, Tianci Wang, Gustav Müller-Franzes, Christiane Kuhl, Sebastian Försch, Jens Kleesiek, Christoph Haarburger, Keno K Bressem, and 1 others. 2024. Medical large language models are susceptible to targeted misinformation attacks. *NPJ digital medicine*, 7(1):288.

Joey Hejna, Rafael Rafailov, Harshit Sikchi, Chelsea Finn, Scott Niekum, W Bradley Knox, and Dorsa Sadigh. 2024. Contrastive preference learning: Learning from human feedback without reinforcement learning. In *The Twelfth International Conference on Learning Representations*.

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.

Yue Huang and Lichao Sun. 2023. Fakegpt: fake news generation, explanation and detection of large language models. *arXiv preprint arXiv:2310.05046*.

Bassamtiano Renaufalgi Irnawan, Sheng Xu, Noriko Tomuro, Fumiyo Fukumoto, and Yoshimi Suzuki. 2025. Claim veracity assessment for explainable fake news detection. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 4011–4029, Abu Dhabi, UAE. Association for Computational Linguistics.

Jennifer Jerit and Yangzi Zhao. 2020. Political misinformation. *Annual Review of Political Science*, 23(1):77–94.

Seungjae Jung, Gunsoo Han, Daniel Wontae Nam, and Kyoung-Woon On. 2024. Binary classifier optimization for large language model alignment. *arXiv preprint arXiv:2404.04656*.

Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore. 1996. Reinforcement learning: A survey. *Journal of artificial intelligence research*, 4:237–285.

Rohit Kumar Kaliyar, Anurag Goswami, and Pratik Narang. 2021. Fakebert: Fake news detection in social media with a bert-based deep learning approach. *Multimedia Tools Appl.*, 80(8):11765–11788.

Danial Kamali, Joseph D. Romain, Huiyi Liu, Wei Peng, Jingbo Meng, and Parisa Kordjamshidi. 2024. Using persuasive writing strategies to explain and detect health misinformation. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 17285–17309, Torino, Italia. ELRA and ICCL.

Santoshi Kumari, Harshitha K Reddy, Chandan S Kulkarni, and Vanukuri Gowthami. 2021. Debunking health fake news with domain specific pre-trained model. *Global Transitions Proceedings*, 2(2):267–272.

Jianqiao Lai, Xinran Yang, Wenyue Luo, Linjiang Zhou, Langchen Li, Yongqi Wang, and Xiaochuan Shi. 2024. Rumorllm: A rumor large language model-based fake-news-detection data-augmentation approach. *Applied Sciences*, 14(8):3532.

João A. Leite, Olesya Razuvayevskaya, Kalina Bontcheva, and Carolina Scarton. 2025. Weakly supervised veracity classification with llm-predicted credibility signals. *EPJ Data Science*, 14(1):16.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.

Miaomiao Li, Hao Chen, Yang Wang, Tingyuan Zhu, Weijia Zhang, Kaijie Zhu, Kam-Fai Wong, and Jindong Wang. 2025. Understanding and mitigating the bias inheritance in llm-based data augmentation on downstream tasks. *arXiv preprint arXiv:2502.04419*.

Hui Liu, Wenya Wang, Haoru Li, and Haoliang Li. 2024. TELLER: A trustworthy framework for explainable, generalizable and controllable fake news detection. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 15556–15583, Bangkok, Thailand. Association for Computational Linguistics.

Jason Lucas, Adaku Uchendu, Michiharu Yamashita, Jooyoung Lee, Shaurya Rohatgi, and Dongwon Lee. 2023. Fighting fire with fire: The dual role of LLMs in crafting and detecting elusive disinformation. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 14279–14305, Singapore. Association for Computational Linguistics.

Minjia Mao, Xiaohang Zhao, and Xiao Fang. 2024. Early detection of misinformation for infodemic management: A domain adaptation approach. *arXiv preprint arXiv:2406.10238*.

Yury Orlovskiy, Camille Thibault, Anne Imouza, Jean-François Godbout, Reihaneh Rabbany, and Kellin Pelrine. 2024. Uncertainty resolution in misinformation detection. In *Proceedings of the 1st Workshop on Uncertainty-Aware NLP (UncertaiNLP 2024)*, pages 93–101, St Julians, Malta. Association for Computational Linguistics.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, and 1 others. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744.

Aldo Pareja, Nikhil Shivakumar Nayak, Hao Wang, Krishnateja Killamsetty, Shivchander Sudalairaj, Wenlong Zhao, Seungwook Han, Abhishek Bhandwaldar, Guangxuan Xu, Kai Xu, and 1 others. 2024. Unveiling the secret recipe: A guide for supervised finetuning small llms. *arXiv preprint arXiv:2412.13337*.

Bohdan M Pavlyshenko. 2023. Analysis of disinformation and fake news detection using fine-tuned large language model. *arXiv preprint arXiv:2309.04704*.

James Prather, Juho Leinonen, Natalie Kiesler, Jamie Gorson Benario, Sam Lau, Stephen MacNeil, Narges Norouzi, Simone Opel, Vee Pettit, Leo Porter, Brent N. Reeves, Jaromir Savelka, David H. Smith, Sven Strickroth, and Daniel Zingaro. 2025. Beyond the hype: A comprehensive review of current trends in generative ai research, teaching practices, and tools. In *2024 Working Group Reports on Innovation and Technology in Computer Science Education*, ITiCSE 2024, page 300–338, New York, NY, USA. Association for Computing Machinery.

Simeng Qin and Mingli Zhang. 2024. Boosting generalization of fine-tuning bert for fake news detection. *Information Processing & Management*, 61(4):103745.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, 36:53728–53741.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Fiona Anting Tan, Jay Desai, and Srinivasan H. Sengamedu. 2024. Enhancing fact verification with causal knowledge graphs and transformer-based retrieval for deductive reasoning. In *Proceedings of the Seventh Fact Extraction and VERification Workshop (FEVER)*, pages 151–169, Miami, Florida, USA. Association for Computational Linguistics.

Falcon-LLM Team. 2024. The falcon 3 family of open models.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, and 1 others. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

R Upadhyay and 1 others. 2024. Addressing the challenge of online health misinformation: Detection, retrieval, and explainability.

Gengyu Wang, Kate Harwood, Lawrence Chillrud, Amith Ananthram, Melanie Subbiah, and Kathleen McKeown. 2023. Check-COVID: Fact-checking COVID-19 news claims with scientific evidence. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 14114–14127, Toronto, Canada. Association for Computational Linguistics.

Jingwei Wang, Ziyue Zhu, Chunxiao Liu, Rong Li, and Xin Wu. 2024. Llm-enhanced multimodal detection of fake news. *PloS one*, 19(10):e0312240.

Haoran Xu, Amr Sharaf, Yunmo Chen, Weiting Tan, Lingfeng Shen, Benjamin Van Durme, Kenton Murray, and Young Jin Kim. 2024. Contrastive preference optimization: Pushing the boundaries of llm performance in machine translation. In *International Conference on Machine Learning*, pages 55204–55224. PMLR.

Lingling Xu, Haoran Xie, Si-Zhao Joe Qin, Xiaohui Tao, and Fu Lee Wang. 2023. Parameter-efficient fine-tuning methods for pretrained language models: A critical review and assessment. *arXiv preprint arXiv:2312.12148*.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, and 1 others. 2025. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.

An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, and 1 others. 2024. Qwen2. 5 technical report. *arXiv preprint arXiv:2412.15115*.

Chang Yang, Peng Zhang, Wenbo Qiao, Hui Gao, and Jiaming Zhao. 2023. Rumor detection on social media with crowd intelligence and chatgpt-assisted networks. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 5705–5717.

Kai Yu, Shiming Jiao, and Zhilong Ma. 2025. Fake news detection based on bert multi-domain and multimodal fusion network. *Computer Vision and Image Understanding*, page 104301.

Majid Zarharan, Pascal Wullschleger, Babak Behkam Kia, Mohammad Taher Pilehvar, and Jennifer Foster. 2024. Tell me why: Explainable public health fact-checking with large language models. In *Proceedings of the 4th Workshop on Trustworthy Natural Language Processing (TrustNLP 2024)*, pages 252–278, Mexico City, Mexico. Association for Computational Linguistics.

Xinyi Zhou, Apurva Mulay, Emilio Ferrara, and Reza Zafarani. 2020. Recovery: A multimodal repository for covid-19 news credibility research. In *Proceedings of the 29th ACM International Conference on Information Knowledge Management*, pages 3205–3212.