Misinformation Detection in Healthcare

Misinformation detection in healthcare has become a critical area of research and application, especially in the age of digital communication and social media. The rapid dissemination of false or misleading information about medical treatments, vaccines, and health practices can have severe consequences, including public health crises and erosion of trust in healthcare systems. Advanced technologies, such as natural language processing (NLP), machine learning (ML), and artificial intelligence (AI), play a pivotal role in identifying and combating misinformation. These tools analyze vast amounts of text, images, and videos to detect patterns indicative of false claims. For instance, algorithms can assess the credibility of sources, cross-reference claims with established medical literature, and flag potentially harmful content for further review.

Despite technological advancements, misinformation detection in healthcare faces several challenges. One significant issue is the dynamic and evolving nature of misinformation, which makes it difficult for automated systems to keep up. Additionally, the context-dependent nature of medical information requires sophisticated models capable of understanding domain-specific terminology and the cultural context of claims. Large Language Models (LLMs) like the LLaMA family have significant potential to enhance misinformation detection in healthcare by leveraging their advanced NLU and generation capabilities. LLMs are trained on vast datasets, enabling them to process complex, domain-specific terminology and understand various contexts, which is particularly valuable for detecting any forms of misinformation in medical content. By analyzing text from social media, articles, or forums, LLMs can identify inconsistencies, verify claims against trusted medical literature, and flag potentially misleading statements. Their ability to provide detailed explanations and references can also help healthcare professionals evaluate and counter misinformation more effectively.

However, a specific approach may not be completely suitable as it may not incorporate human evaluators' feedback into consideration. However, Reinforcement Learning with Human Feedback (RLHF) can significantly enhance the effectiveness of LLMs in detecting and combating misinformation in healthcare. RLHF involves fine-tuning LLMs using feedback from human evaluators, enabling the models to better align with human values and domain-specific requirements. In the context of healthcare misinformation detection, RLHF ensures that the model not only processes and analyzes information accurately but also prioritizes trustworthy sources and aligns its outputs with medical ethics and evidence-based practices during finetuning.

[1] Infodemics and health misinformation: a systematic review of reviews - PMC

Research Goals

Misinformation detection applications should be prominently updated in procedures, so using larger models is not optimal here as it may cost a lot in computational resources. Moreover, due to the task's complexity, it is possible to obtain lower performance with smaller LLMs, even with

fine-tuning. So, the aim of this work is to propose an RL technique capable of boosting smaller LLM performance with fewer computational resources so that it can be a practically viable solution in real-world applications.

The research questions (RQs) of this work are four-fold:

1. **RQ1:** How effective are LLMs in detecting misinformation in healthcare?
2. **RQ2:** To what extent does SFT enhance task-specific adaptation for misinformation detection in healthcare?
3. **RQ3:** How does RLHF influence the performance of LLMs in detecting misinformation in healthcare compared to SFT?
4. **RQ4:** What advancement in the RLHF paradigm can boost the effectiveness of LLMs for healthcare misinformation detection?

Datasets

Recommended datasets for this study are presented in the following table (the stats are presented using the raw datasets from this repository – https://github.com/ikr3-lab/health-misinformation/tree/master/Data/data_w_feature so they are presenting the claims that have been collected via the works):

	Real Claim	**Fake Claim**
ReCOVery [1, 2]	1364	665
FakeHealth [3, 4]	1386	706
FakeHealth Story	*1078*	*420*
FakeHealth Review	*308*	*286*

[1] https://github.com/apurvamulay/ReCOVery
[2] https://dl.acm.org/doi/10.1145/3340531.3412880
[3] https://arxiv.org/abs/2002.00837
[4] https://github.com/EnyanDai/FakeHealth

The FakeHealth dataset consists of two sets FakeHealth Story and Review which we combined to provide a more comprehensive dataset for further analysis of the work. For training and test sets, we used a ratio of 25% for test and 75% for train set splits. The following table presents the dataset statistics.

	FakeHealth			**ReCOVery**		
	Real	***Fake***	***Total***	***Real***	***Fake***	***Total***
Train	1040	529	1569	1022	499	1521
Test	346	177	523	342	166	508

We observed several numbers of news from FakeHealth are not accessible (no longer it is available on the website so, we ignored them as we required news content in the datasets).

Standardized Prompting

For standardized prompting for testing LLMs, we adopted the following prompt templates:

```
Given the title and content of a healthcare news article, analyze whether the claims align with plausible scenarios and whether the article maintains internal consistency. Check for misleading or unclear statements, and conclude whether the news is real or fake.

News Title:
{TITLE}
News Content:
{NEWS}
Conclusion:
```

This prompt:
1. **Plausibility**: This focuses on whether the claims are realistic. The LLM can evaluate this based on patterns, known language structures, and the context of the article itself. Even without external sources, the LLM can assess whether the content aligns with typical healthcare scenarios.
2. **Internal Consistency**: Internal consistency is a key factor the LLM can check based on how well the arguments, claims, and facts within the text align with one another. Inconsistent statements or contradictions can easily indicate misinformation, which the model can identify from the text itself.
3. **Clarity**: The instruction to look for "misleading or unclear statements" encourages the model to focus on how well the content is communicated, which is within its capabilities. Vague or contradictory wording is an effective sign of unreliable information.
4. **Clear Decision**: The instruction to "conclude if the news is real or fake" ensures a direct and actionable outcome from the model's analysis, aligning with the query's intent to provide a definitive judgment.

Supervised Finetuning (SFT)

We used supervised fine-tuning (SFT) with QLoRA (Quantized Low-Rank Adaptation) as an optimal approach for fine-tuning LLMs while optimizing computational efficiency. QLoRA as it works with PEFT (Parameter Efficient Fine-Tuning) methods, is designed to allow you to fine-tune models with reduced memory usage by using quantized low-rank adaptation layers, which significantly reduce the computational overhead without compromising model performance.

Reinforcement Learning with Human Feedback (RLHF)

Reinforcement Learning with Human Feedback (RLHF) is a framework where the model is trained to optimize an objective function that reflects human preferences or judgments. We adopted the **Binary Classifier Optimization (BCO)** and **Conservative Policy Optimization (CPO)** for misinformation detection where BCO is a binary classifier trained to distinguish between "good" (chosen) and "bad" (rejected) completions for a given prompt. However, CPO at a high level, CPO trains models to avoid making correct but suboptimal predictions in classification tasks, ensuring that the model not only classifies accurately but also improves the quality of its predictions, aligning with the desired performance metrics.

However, BCO requires sampling that impacts the BCO's performance. Moreover, the CPO is a general approximation to Direct Preference Optimization (DPO) where recent research has shown that DPO may lead to biased results, such as models producing lengthy outputs, which affects the model's ability to follow instructions and reasoning (Disentangling Length from Quality in Direct Preference Optimization - ACL Anthology and Length Desensitization in Direct Preference Optimization | OpenReview).

Imitation Learning via Scaled KL-divergence-like Behavior Cloning (KL-BC)

The proposed imitation learning via KL-divergence-like behavior cloning (KL-BC) mimics the effect of Kullback-Leibler (KL) divergence in guiding a model's predictions toward a reference behavior. KL-divergence is commonly used in behavior cloning (BC) – https://arxiv.org/pdf/2212.02125 –, especially in RL and preference modeling, to match a policy to expert demonstrations. The goal is to ensure that the policy closely follows the demonstrated behavior encoded in the reference model. Given a policy model $\pi_\theta(a \mid s)$ and a reference model $\pi_{ref}(a \mid s)$, KL minimizes the following objective $DL_{KL}(\pi_{ref}(a \mid s) \,||\, \pi_\theta(a \mid s))$. Where $D_{KL}(P||Q) = \sum_x P(x)\, log\, \frac{P(x)}{Q(x)}$ with P as the policy being trained, Q is the expert policy, and the loss encourages P to imitate Q by penalizing deviations.

In our approach, we introduced the following log-based regularization term $-\ log\ (exp(\theta)\ +\ \epsilon)$ where the $\theta := log(P_{chosen}) - log(P_{rejected})$ is a difference between desirable (chosen) and undesirable (rejected) responses. The term itself exhibits behavior similar to KL divergence in several key ways. First, it encourages preference limitation, ensuring that the model learns to favor chosen responses over rejected ones by penalizing large deviations, much like KL divergence does when aligning probability distributions. Second, it promotes smooth behavior cloning, as the term $exp(\theta)$ functions as an unnormalized probability ratio, and applying the logarithm mitigates abrupt gradient fluctuations, leading to more stable optimization. Lastly, the regularization effect is evident in the $-\ log\ (exp(\theta)\ +\ \epsilon)$ term, which

prevents excessively high preference scores, thereby improving training stability and preventing overfitting to specific samples.

Building on this, we incorporate Contrastive Preference Optimization (CPO) with KL-BC. To enhance preference learning, we used a CPO a sigmoid loss with label smoothing, as follows, to guide the optimization process and learn the preferences effectively using CPO.

$$L_{sigmoid} = \; - \; log(\delta(\beta \,.\, \theta)) \,.\, (1 \; - \; sm) \; - \; log(\delta(- \; \beta \,.\, \theta)) . sm$$

Where sm represents the label smoothing term that adjusts the target label distribution, reducing the model's confidence in its predictions and making the model more robust to overfiting. Moreover, $\delta(x)$ is the sigmoid function and β is a scaling factor here.

The final loss function is determined as follows to address the limitations of BCO and CPO, particularly the challenges of sampling inefficiencies and biases in DPO.

$$L_{KL-BC} \; = \; L_{sigmoid} \; * \; (- \; log(exp(\theta) \;\; + \;\; \epsilon)$$

This loss function combines the sigmoid loss with a regularization term to stabilize the model's learning process. By incorporating both CPO's preference optimization and KL-BC's regularization, it effectively mitigates the issues of biased outputs and sampling impacts, promoting more stable and accurate preference learning. This approach ensures the model can generate high-quality responses while maintaining robustness in its performance.

Experimental Models:

Here, smaller LLMs are the key since the RL part plays the most important role in this study, and we might need computational resources in the RL part. Choosing smaller LLMs can boost development. Moreover, recently, researchers/practitioners have been looking for smaller LLMs so they can be deployable and usable in production-level applications. In this regard, RL can help with achieving better performance even with smaller LLMs.

The criteria in choosing proposed LLMs are: 1) they are state of the art, 2) they offer smaller versions and bigger versions, that bigger versions are working well, but we believe that our proposed RLHF can boost their performance, and lastly, 3) The important note here is that the chosen LLMs are the latest versions. So, overall, we aim to propose RL specific model that can boost LLMs' performances in finetuning for misinformation tasks in social media. With respect to this, the following LLMs are chosen:

- LLaMA-3.2-1B-Instruct: https://huggingface.co/meta-llama/Llama-3.2-1B-Instruct
- Falcon3-3B-Instruct: https://huggingface.co/tiiuae/Falcon3-3B-Instruct
- Qwen2.5-0.5B-Instruct: https://huggingface.co/Qwen/Qwen2.5-0.5B-Instruct
- Phi-3.5-Mini-Instruct: https://huggingface.co/microsoft/Phi-3.5-mini-instruct

Results On Test

- The values in blue represent the top performances, while the orange values indicate the second-best performances. The Macro averages are reported. Results are in %.
- ** = refers to the modified CPO RL.

Model	FakeHealth					ReCOVery				
	Accuracy	Precision	Recall	F1-Score	F1-Fake	Accuracy	Precision	Recall	F1-Score	F1-Fake
Qwen2.5-0.5B-Instruct										
Standardized Prompting	46.46	48.14	47.94	45.81	39.91	49.40	51.68	51.88	48.80	43.26
SFT	59.08	55.97	56.38	55.97	44.27	51.96	48.79	48.67	48.37	34.75
SFT + BCO	72.27	69.30	64.69	65.48	50.17	96.65	97.04	95.34	96.12	94.70
SFT + CPO	71.70	68.47	68.67	68.57	*58.65*	92.71	93.44	89.94	91.39	88.02
SFT + CPO**	73.04	69.91	69.96	69.93	60.28	95.07	94.73	94.01	94.36	92.35
Falcon3-3B-Instruct										
Standardized Prompting	43.02	43.47	42.72	41.75	33.18	39.37	58.56	53.57	36.15	50.48
SFT	69.59	65.36	62.94	63.48	48.54	94.09	93.66	92.82	93.22	90.79
SFT + BCO	72.46	69.10	66.49	67.23	54.41	**98.03**	**98.39**	**97.14**	**97.73**	**96.91**
SFT + CPO	74.18	71.12	68.90	69.66	57.94	94.68	94.72	93.10	93.85	91.58
SFT + CPO**	74.95	72.16	69.75	70.56	59.19	95.66	95.07	95.07	95.07	93.37
LLaMA-3.2-1B-Instruct										
Standardized Prompting	63.28	44.86	48.79	41.96	6.79	64.96	53.93	51.81	49.03	20.53
SFT	65.96	60.00	57.58	57.53	38.62	85.23	84.19	81.43	82.55	75.72
SFT + BCO	72.84	70.53	64.71	65.52	49.64	96.06	95.65	95.37	95.51	93.93
SFT + CPO	71.89	70.26	62.19	62.47	43.67	95.47	96.20	93.53	94.70	92.69
SFT + CPO**	74.18	72.01	66.96	67.99	53.92	96.06	95.52	95.52	95.52	93.97
Phi-3.5-Mini-Instruct										
Standardized Prompting	64.81	42.00	49.26	40.34	2.12	70.07	73.31	55.14	51.19	20.83
SFT	71.31	67.65	66.31	66.79	54.54	93.50	92.27	93.16	92.69	90.26
SFT + BCO	70.17	66.14	64.07	64.63	50.63	97.83	98.08	96.99	97.51	96.61
SFT + CPO	76.09	73.31	73.38	73.34	**64.78**	95.47	95.65	94.00	94.76	92.83
SFT + CPO**	**76.86**	**74.19**	**73.40**	**73.75**	64.72	97.04	97.51	95.79	96.58	95.32

RQ1: How effective are LLMs in detecting misinformation in healthcare?

This was **Step 1–Test LLMs Capabilities** of the work which we tested via standardized prompting.

Step 1–Test LLMs Capabilities. This step aims to test LLMs' capabilities for misinformation detection, which plays a foundational step in further evaluations. The proposed prompting technique and LLMs of this study are represented as follows:

- Prompting: Zero-shot Prompting Scenario.
- LLMs:
    - LLaMA-3.2-1B-Instruct: https://huggingface.co/meta-llama/Llama-3.2-1B-Instruct
    - Falcon3-3B-Instruct: https://huggingface.co/tiiuae/Falcon3-3B-Instruct
    - Qwen2.5-0.5B-Instruct: https://huggingface.co/Qwen/Qwen2.5-0.5B-Instruct
    - Phi-3.5-Mini-Instruct: https://huggingface.co/microsoft/Phi-3.5-mini-instruct

RQ2: To what extent does SFT enhance task-specific adaptation for misinformation detection in healthcare?

This was **Step 2–Supervised Finetuning** of the work which we tested via SFT.

Step 2–Supervised Finetuning. Once we get an idea of how capable LLMs are, we can do the supervised finetuning to show how much improvement we have made with finetuning. For the finetuning here, we can use QLoRA-based finetuning of the best performer LLM from Step 1.

RQ3: How does RLHF influence the performance of LLMs in detecting misinformation in healthcare compared to SFT?

This was **Step 3–RLHF for Misinformation Detection** of the work which we tested via SFT + BCO and SFT + CPO models.

Step 3–RLHF for Misinformation Detection. Here, we can show that LLM generally performs better with RLHF. We can use an RL Policy (e.g., BCO, CPO, or PPO) to see whether RL performs better in a misinformation task

RQ4: What advancement in the RLHF paradigm can boost the effectiveness of LLMs for healthcare misinformation detection?

This was **Step 4–Proposed Novel RL Technique** of the work which we tested via proposing SFT + CPO + KL-BC model.

Step 4–Proposed Novel RL Technique. Once, we show how zero-shot, supervised fine-tuning, and RLHF work, here we will propose modifications to the RLHF mechanism to improve its performance. The key idea here is to apply multiple loss schema to the Policy so it can benefit from different learners' paradigms (learning from multiple losses).

Ablation Studies

Model	**FakeHealth**					**ReCOVery**				
	Accuracy	**Precision**	**Recall**	**F1-Score**	**F1-Fake**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**F1-Fake**
Standard prompting on all samples (train + test)										
Qwen2.5-0.5B-Instruct	46.84	47.67	47.41	45.83	38.42	52.48	53.84	54.33	51.62	45.16
Falcon3-3B-Instruct	44.16	43.66	42.92	42.34	32.09	39.62	58.13	53.55	36.56	50.50
LLaMA-3.2-1B-Instruct	59.56	47.48	48.45	46.14	19.27	61.11	51.50	51.11	50.31	27.14
Phi-3.5-Mini-Instruct	65.39	48.30	49.83	41.31	3.72	70.87	74.83	56.56	53.59	25.28
5-Fold Cross Validations (SFT + CPO**)										
Phi-3.5-Mini-Instruct	75.00	72.05	71.43	71.65	61.93	97.43	97.58	96.59	97.06	96.01
Transformer methods (finetuned on the train and evaluated on test)										
BERT is similar to [1, 7]	69.78	65.80	64.61	65.02	52.12	90.55	91.24	87.09	88.73	84.21
Fake News BERT	63.86	59.77	59.85	59.81	47.05	85.03	83.10	82.68	82.89	76.82
Comparison to the latest works!										
LLaMA-3-8B [8]	-	-	-	-	-	95.9	96.1	94.5	95.2	-
GPT-3.5 [9]	-	-	-	-	-	95.7	96.4	93.9	95.0	-

The standard prompting on all the dataset samples (train+test) showed the same challenge as we observed in the split test set. Even further cross-validation of one of the approaches (Phi-3.5-Mini-Instruct LLM using SFT + CPO** model) showed a similar behavior. This signifies the generalizability of such approach.

The finetuned BERT models on downstream tasks mostly fall short w.r.t the RLHF-based finetuning. A reason for this underperformance relies on the backbone of BERT variants which comes from limited input token limits (512 tokens), which often news articles are being considered as a long document. Another perspective of classification-based models is to finetune LLM on similar documents and then finetune it for the task on hand, however, the fake news BERT model which has been finetuned on a larger number of samples, showed that for specific tasks such as medical misinformation detection, this approach might not be suitable.

In an attempt to run a comparison of the work toward the latest works in the field. We observed that LLaMA-3-8B and GPT-3.5 were employed over the ReCOVery dataset, and as we can see that previous attempts toward using LLMs for misinformation detection in healthcare rely on using larger models as observed in works of [8] and [9], however, with the RLHF approach even the smaller LLM can surpass the need for larger models, when they are finetuned to align the human preferences with behavior cloning.

References:

[1] Bad Actor, Good Advisor: Exploring the Role of Large Language Models in Fake News Detection | Proceedings of the AAAI Conference on Artificial Intelligence
[2] Nice category of misinformation approaches ICTMCG/LLM-for-misinformation-research (good for storytelling of the paper – we will use all the papers of section 1 for related work!).
[3] https://dl.acm.org/doi/10.1007/s11042-020-10183-2
[5] https://arxiv.org/pdf/2310.05046
[6] https://aclanthology.org/2023.emnlp-main.883.pdf → This is the reason why we explored small LLMs, we aimed to study the capability of small LLMs for whatever they can achieve the larger LLM results or not, which we achieved and even surpassed!
[7] FakeBERT: Fake news detection in social media with a BERT-based deep learning approach | Multimedia Tools and Applications – no model is available for this!
[8] Claim veracity assessment for explainable fake news detection
[9] Fact-Checking COVID-19 News Claims with Scientific Evidence - ACL Anthology