Generating Biomedical Fact-Checking Reports with RL-Enhanced Agentic Search

Jiongxiao Wang[1] **Dingli Ma**[2] **Chaoqun Ni**[1]
[1]University of Wisconsin–Madison; [2]University of Washington

Abstract

Automated fact-checking is essential for ensuring the reliability of public health information, yet the biomedical domain poses unique challenges. Validating biomedical claims requires rigorous interpretation of scientific literature, assessment of retrieved evidence, and comprehensive justification toward the conclusion. Although Large Language Models (LLMs) enhanced by Retrieval-Augmented Generation (RAG) and agentic search perform automated fact-checking in a retrieve-then-verify paradigm, current methods still output isolated prediction labels, lacking explanatory depth and offers limited utility for human understanding. To bridge this gap, we introduce an LLM-based agent named BioCheck Agent that generates structured biomedical fact-checking reports with agentic search. Rather than merely outputting supported or refuted labels, our agent synthesizes final conclusions with retrieved evidence and rigorous analysis. To ensure domain-specific accuracy, BioCheck Agent exclusively searches high-quality scientific literature in PubMed, utilizing advanced Boolean search operators. Recognizing that direct prompting often results in hallucinations and low-quality reports, especially for lightweight open-source models, we further propose the Evidence-Grounded Group Relative Policy Optimization (EG-GRPO) to perform reinforcement learning on BioCheck Agent with a task-specific reward that incentivizes advanced search behavior and high-quality evidence retrieval while penalizing hallucinations. Our experimental results show that compared to the base model Qwen3.5-4B, BioCheck Agent with EG-GRPO improves label prediction accuracy on SciFact by 9.95%. Furthermore, it achieves a 3.7% higher evidence quality score and a 19.63% lower evidence hallucination rate, demonstrating its ability to generate biomedical fact-checking reports with improved accuracy and quality.

1 Introduction

Health misinformation has become a major challenge for public health, especially during health crises when people must navigate false or misleading claims, information overload, uncertainty, and rapidly changing scientific evidence. The World Health Organization described this problem as an "infodemic" and emphasized that managing misinformation is part of controlling public health emergencies[1]. This challenge is especially consequential in biomedical contexts, where inaccurate claims may shape vaccination decisions, treatment choices, adherence to public health guidance, trust in medical institutions, and the use of unsafe remedies. Prior research shows that health misinformation can produce measurable harms. During the COVID-19 pandemic, misinformation about prevention and treatment circulated widely across countries and was linked to serious real-world consequences, including hospitalizations and deaths[2]. Experimental studies further show that exposure to COVID-19 vaccine misinformation can reduce vaccination intent[3], while cross-national survey research finds that susceptibility to misinformation is associated with lower self-reported compliance with public health guidance and lower willingness to be vaccinated[4]. Together, these

Preprint.

findings suggest that biomedical misinformation is not merely a communication problem; it can affect health behavior, undermine public trust, and weaken evidence-based public health interventions.

Automated fact checking offers a promising way to improve the reliability of online health information, but biomedical fact checking remains particularly difficult. Public health fact checking often requires domain expertise and assessment against credible evidence, and prior work has argued that this setting requires not only veracity prediction but also explanation generation [5]. Health claims also require careful interpretation of biomedical evidence, including the strength and certainty of available studies, which is why recent medical fact checking datasets incorporate evidence levels in addition to labels such as supported, refuted, and not enough information [6]. More broadly, evidence based medical reasoning requires attention to the population or patient problem, intervention, comparison, and outcome when formulating clinical questions and searching for evidence [7]. For this reason, biomedical fact-checking should provide more than a final label. It should explain what evidence was retrieved, how that evidence relates to the claim, what uncertainty remains, and why a particular conclusion is justified.

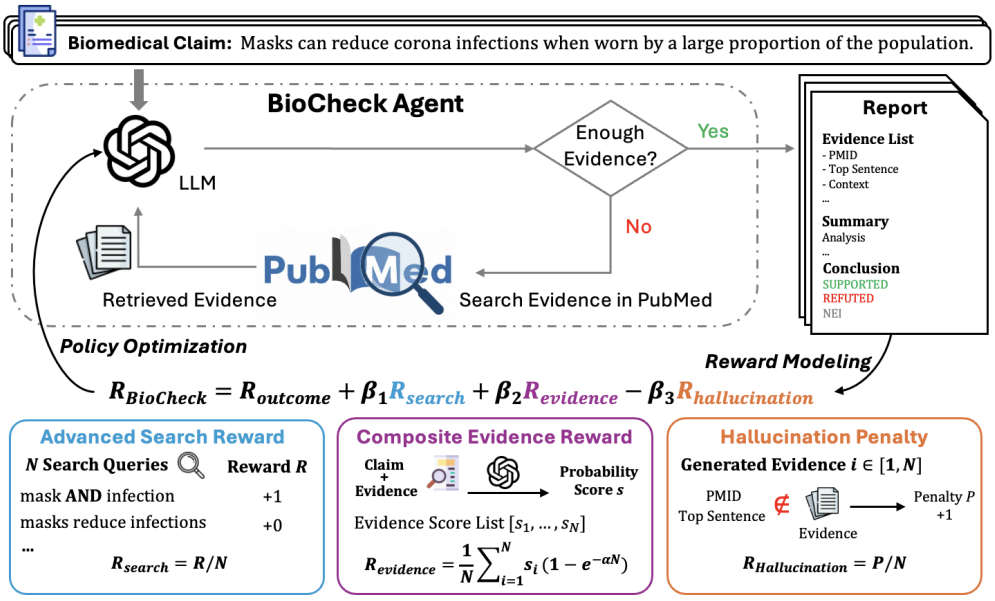


Figure 1: Illustration for BioCheck Agent and Reward Modeling for EG-GRPO.

With the advancement of Large Language Models (LLMs) and techniques like retrieval-augmented generation (RAG) [8] and agentic search [9], current automated fact-checking tasks have transitioned from text classification to a retrieve-then-verify paradigm [10, 11, 12]. This requires first retrieving external information from the open world and then verifying the claim based on those findings. Although these approaches have achieved high classification accuracy, their label-focused results and evaluation lack perfect reliability, making human experts still essential for robust decision-making where predicted labels fail to provide meaningful information.

To advance automated biomedical fact-checking beyond label prediction through comprehensive evidence analysis, we propose **BioCheck Agent**, a novel LLM-based agent for structured biomedical fact-checking report generation. BioCheck Agent generates comprehensive reports by outlining relevant evidence, writing analytical summaries and making the final conclusion with predicted labels. Tailored for the biomedical domain, our agent executes agentic searches within the PubMed database via its API, which retrieves the most up-to-date literature, thus offering superior evidence quality compared to static databases or generalized search engines. Besides, our agent is equipped with advanced academic search strategies, automatically decomposing complex claims into core entities and linking them with Boolean operators ('AND', 'OR') to construct highly accurate search queries.

To further enhance the performance and scalability of BioCheck Agent beyond direct prompting, we propose Evidence-Grounded Group Relative Policy Optimization (EG-GRPO) for end-to-end reinforcement learning (RL). Building upon GRPO [13], which has proven effective across various

agentic tasks [14, 15], EG-GRPO employs a task-specific reward design that emphasizes the quality and groundedness of evidence in the generated reports. Specifically, alongside an outcome-based reward for ground-truth label alignment, we introduce three extra components: an Advanced Search Reward to encourage the use of Boolean search operations, a Composite Evidence Reward to balance the quantity and quality of the generated evidence, and a Hallucination Penalty to prevent the model from generating evidence that contradicts the retrieved literature. Then we apply EG-GRPO to train the lightweight Qwen3.5-4B model on the SciFact [16] training split to serve as the backbone model for BioCheck Agent. An illustrative overview for BioCheck Agent and the reward design of EG-GRPO is shown in Figure 1.

We evaluate the performance of BioCheck Agent with EG-GRPO on both the SciFact test split and the HealthFC [6] dataset. Across both datasets, BioCheck Agent optimized with EG-GRPO outperforms all baselines in label prediction metrics when using Qwen3.5-4B as the base model, even surpassing the commercial model GPT5.2. Compared to the base model prior to RL, EG-GRPO significantly improves prediction accuracy by 9.95% on SciFact and 7.03% on HealthFC. Furthermore, although directly applying the baseline GRPO yields performance comparable to EG-GRPO on SciFact, it shows limited improvement on HealthFC. This demonstrates the necessity of our task-specific reward design, where explicitly optimizing for fact-checking reports leads to better transferable performance. Beyond standard label prediction metrics, we introduce two additional metrics designed to provide fine-grained evaluations of the generated reports: Evidence Quality Score (EQS) and Evidence Hallucination Rate (EHR). Compared to the base model Qwen3.5-4B, applying EG-GRPO on BioCheck Agent clearly improves the quality of the generated biomedical fact-checking reports. Specifically, on the SciFact dataset, it yields a 3.7% improvement in EQS (reflecting better evidence retrieval) and a substantial 19.63% reduction in EHR (indicating fewer hallucinations). These results highlight the effectiveness of EG-GRPO in enabling BioCheck Agent to achieve state-of-the-art performance in both accurate label prediction and high-quality report generation. Ultimately, by mitigating hallucinations and grounding outputs in high-quality evidence, our BioCheck Agent represents a vital step toward assisting humans in discerning health misinformation through generated fact-checking reports.

2 Related Work

Automated Fact-Checking. Fact-checking is defined as the task of assessing the veracity of factual claims. Traditionally, fact-checking is performed by human experts in journalism or scientific domains. However, these processes are normally labor-intensive and time-consuming. With the development of Natural Language Processing (NLP), it has become possible to perform automated fact-checking [17, 18]. Early automated fact-checking mainly depended on supervised learning [19, 20, 21] to train classifiers for verification, and the evidence was often sourced from pre-annotations within the dataset [22, 16]. The emergence of LLMs and RAG has transitioned fact-checking to an open-world retrieve-and-verify paradigm. Rather than relying on static, pre-annotated evidence, current methods actively retrieve information from external sources like vector-based retrieval databases [11, 23] or structured knowledge graphs [24]. More recently, LLM-based agents have gained prominence by leveraging external tools, such as search engines, to perform more comprehensive and real-time evidence retrieval [12, 25, 26]. Specifically, biomedical fact-checking presents unique challenges that demand extensive domain expertise and rigorous verification, given its significant impact on public health, as evidenced during the COVID-19 pandemic [27]. To advance automated fact-checking in this field, numerous datasets and benchmarks have been established [16, 28, 29, 6, 30], alongside specialized methods designed to leverage high-authority, rigorous resources for verification [31, 23].

LLM-based Agents and Agentic Search. The advanced instruction-following and reasoning capabilities of LLMs now enable the automated execution of complex tasks as agents. Consequently, various frameworks, such as ReAct [32], Plan-and-Act [33], and CodeAct [34], have been proposed to build autonomous agents with any backbone LLMs. Given the growing real-world impact of agents, industry-level frameworks like LangGraph [35] and OpenHands [36] have also emerged. Besides, most LLMs nowadays are trained to execute predefined actions via a standard function-calling pipeline [37], rather than relying on structured text outputs like ReAct. As a critical application of LLM-based agents, agentic search integrates search engines with reasoning capabilities to enhance response quality [9, 38]. By grounding outputs in verifiable references, this approach significantly mitigates hallucinations [39]. Initially pioneered by agents such as Bing Chat [40], agentic search is

now widely employed to bolster performance across a broad spectrum of applications, from general benchmarks to domain-specific tasks [41].

Reinforcement Learning for Agents. Unlike general chatbot LLMs, whose performance mainly depends on human preferences [42], agent tasks often feature explicit answers, making reinforcement learning with verifiable rewards (RLVR) highly applicable. Specifically, Group Relative Policy Optimization (GRPO) [13] has been demonstrated as an effective algorithm and is widely applied to improve agent performance across various tasks [9, 43, 44, 45].

3 Method

This section provides a detailed description of our BioCheck Agent and the EG-GRPO method used to further improve the agent performance.

3.1 BioCheck Agent

BioCheck Agent is built as a standard search agent, iteratively invoking an LLM through LangChain to first reason and then call a predefined search tool to retrieve related information for verifying a given claim. Following each tool response, the LLM determines whether to continue searching or to generate a structured fact-checking report using the gathered evidence. To prevent the context window from exceeding its limits, we impose a hard constraint on the agent: once a maximum number of searches is reached, we inject a user instruction as the intervention. This prompts the LLM to terminate the search and immediately output its report based on the evidence collected thus far.

Agent Framework. We utilize LangGraph [35] to develop BioCheck Agent. LangGraph is a low-level orchestration framework designed for building stateful agents via graph-based structures. Within this framework, each node represents an action, such as invoking LLMs, executing tools, or incorporating human-in-the-loop interventions, while edges define the execution flow between nodes, often governed by conditional logic. Additionally, since LangGraph only serves as a framework to connect agent nodes, LangChain [46] remains essential for defining key nodes by invoking LLMs with customized tools through their standard function-calling capabilities.

Tool Definition. Given our focus on biomedical fact-checking, we designed the search tool named $pubmed\_search$ that utilizes the PubMed API [1] to perform query-based searches within PubMed database. The initial search retrieves a list of relevant PMIDs. Considering the long context length and inconsistent availability of full-text papers, our tool extracts only the titles and abstracts of these PMIDs. These elements are then concatenated and returned as the tool responses to provide retrieved context for fact-checking. In this paper, we limited the tool responses to a maximum of 5 papers per search and set the maximum number of iterative searches to 5.

Agent Output. The final output of BioCheck Agent is a structured **Report** consisting of four main sections: **Supporting Evidence**, **Refuting Evidence**, **Summary**, and **Conclusion**. The Supporting/Refuting Evidence section details the literature that either supports or refutes the claim, with each entry providing the PubMed ID (PMID), a verbatim Top Sentence quote from the abstract, and the relevant context. The Summary section outlines the overall reasons to support or refute the claim, accompanied by a consensus check and a final justification. Finally, the Conclusion section clearly presents the final verdict by outputting either 'SUPPORTED' or 'REFUTED' within '<answer>' tags. Details about the LangGraph implementation of BioCheck Agent with corresponding prompts are provided in Appendix A.

3.2 Evidence-Grounded GRPO for BioCheck Agent

In this section, we introduce Evidence-Grounded GRPO by first formalizing the task, detailing our task-specific reward design, and finally presenting the optimization objective.

3.2.1 Task Definition

We define our task of biomedical fact-checking reports generation with BioCheck Agent as a **Partially Observable Markov Decision Process (POMDP)** $(\mathcal{S}, \mathcal{A}, \mathcal{O}, \mathcal{T}, \mathcal{R})$ [47, 48], where $\mathcal{S}, \mathcal{A}, \mathcal{O}, \mathcal{T}, \mathcal{R}$

[1] https://www.ncbi.nlm.nih.gov/home/develop/api/

represents the state space, the action space, the observation space, the state transition function and the reward function, respectively. Given a biomedical claim c, the LLM-based BioCheck Agent initializes at state s_1, which comprises the claim paired with an instructional prompt for fact-checking. At each time step t, given the current state $s_t \in \mathcal{S}$, the agent generates an action $a_t \sim \pi_\theta(\cdot|s_t)$ based on the policy model π_θ to process the task. The action $a_t \in \mathcal{A}$ here includes both the reasoning process and the search tool calling with a generated query. Upon executing the action, the agent receives the tool response as the observation $o_t \in \mathcal{O}$ from the environment, and the system transitions to the next state $s_{t+1} = \mathcal{T}(s_t, a_t)$. After N interaction turns, the agent concludes the task by generating a fact-checking report as its final action a_N. A reward function $\mathcal{R}$ then evaluates this report to assign a final reward $R = \mathcal{R}(a_N)$.

3.2.2 Task-Specific Reward Modeling

EG-GRPO performs a task-specific reward design. The baseline approach of GRPO relies solely on a binary outcome-based reward, where $R = 1$ if the agent task is performed successfully and $R = 0$ otherwise. For BioCheck Agent, although an outcome-based reward, which evaluates whether the report's conclusion aligns with the ground truth label, can improve label prediction performance, generating high-quality biomedical fact-checking reports still requires more fine-grained feedback. Below, we detail the specific design for BioCheck reward R_{BioCheck}:

Outcome-based Reward. We retain the outcome-based reward, R_{outcome}, as a component of the final reward. Specifically, $R_{\text{outcome}} = 1.0$ if the final predicted label matches the ground truth provided by the dataset, and $R_{\text{outcome}} = 0.0$ otherwise.

Advanced Search Reward. To encourage BioCheck Agent to decompose the claim and perform advanced searches by connecting keywords with the Boolean operators 'AND' or 'OR', we introduce an additional reward, R_{search}, which is defined as the ratio of search queries containing Boolean operators to the total number of searches: $R_{\text{search}} = \frac{\text{Num of Searches with Boolean Operator}}{\text{Num of Searches}}$.

Composite Evidence Reward. One critical factor of a fact-checking report is the quality of the evidence it contains. For each piece of retrieved evidence, we apply an additional reward model to assign a confidence score s evaluating the probability that the evidence supports or refutes the claim. Specifically, we frame this claim verification task as a multiple-choice question, utilizing a lightweight LLM Qwen3.5-4B [49] to compute the token probability of selecting the supported or refuted answer based on the provided evidence. The prompt for the multiple-choice question and the corresponding implementation details to compute the evidence confidence score are included in Appendix B.

Normally, multiple pieces of evidence are included for each report. To ensure both the quantity and quality of the retrieved evidence, we propose the Composite Evidence Reward. Formally, given a list of claim evidence pairs $(\mathcal{C}, \mathcal{E}) = [(c_1, e_1), ..., (c_N, e_N)]$, a label $y \in \{\text{supported}, \text{refuted}\}$, and a reward model $\mathcal{L}$, we compute the probability score as $s_k = L(y|c_k, e_k)$. Then we compute the Composite Evidence Reward as

$$R_{\text{CER}}((\mathcal{C}, \mathcal{E}), y) = \frac{1}{N} \sum_{i=1}^{N} s_i (1 - e^{-\alpha N}) \tag{1}$$

where the α coefficient serves as a scaling factor that controls the saturation rate: a larger value would cause the function to saturate more rapidly, meaning fewer pieces of evidence would be required to reach full reward. Here we set $\alpha = 1.0$ by default.

The Composite Evidence Reward is computed for all abstracts retrieved by PMIDs. To ensure that the generated sentences serve as meaningful evidence, we only count the probability score when the corresponding abstract's probability score exceeds 0.5; otherwise, we set $s = 0.0$. Notably, within each fact-checking report, the agent is allowed to retain evidence that contradicts its final decision, and the corresponding evidence scores are also included. This design incentivizes the model to critically retrieve evidence from both sides of a claim. Given the abstract evidence list $\mathcal{E}_{\text{supported}}, \mathcal{E}_{\text{refuted}}$, the evidence reward is defined as

$$R_{\text{evidence}} = R_{\text{CER}}((\mathcal{C}, \mathcal{E}_{\text{supported}}), \text{supported}) + R_{\text{CER}}((\mathcal{C}, \mathcal{E}_{\text{refuted}}), \text{refuted}) \tag{2}$$

Hallucination Penalty. The report generation process is susceptible to hallucinations, such as the fabrication of PMIDs or Top Sentences in the generated evidence list. To mitigate this, we implement

a rule-based hallucination detection mechanism that penalizes hallucinated evidence by assigning it a penalty score. We detect the evidence as a hallucination under two specific conditions: (1) if a provided PMID fails the Exact Match criterion within the retrieved PMID list; (2) if the generated Top Sentence yields a ROUGE-L precision score of less than 0.85 against the source abstract. Then we compute the accumulated hallucination penalty as the ratio of detected hallucinations to the total number of evidence: $R_{\text{hallucination}} = \frac{\text{Num of Hallucination}}{\text{Num of Total Evidence}}$.

Overall, the final BioCheck reward is presented in the following formulation:

$$R_{\text{BioCheck}} = R_{\text{outcome}} + \beta_1 R_{\text{search}} + \beta_2 R_{\text{evidence}} - \beta_3 R_{\text{hallucination}} \tag{3}$$

Here, β_1, β_2 and β_3 are hyperparameters that control the weight of the extra rewards. We set $\beta_1 = 0.5$ and $\beta_2 = \beta_3 = 1.0$ by default.

3.2.3 EG-GRPO Formulation

EG-GRPO shares the same objective function as the baseline GRPO [13]. Specifically, for each claim c from the training claim set C, we first sample a group of trajectories $\{\tau_1, ..., \tau_G\}$ from the old policy $\pi_{\theta_{old}}$, with each trajectory $\tau_i = \{s_{i,1}, a_{i,1}, ..., s_{i,|\tau_i|}, a_{i,|\tau_i|}\}$ includes $|\tau_i|$ states and generated actions. Then, the policy model is optimized by maximizing the following objective function:

$$\begin{aligned}\mathcal{J}_{\text{EG-GRPO}}(\theta) = \mathbb{E}[c \sim C, \{\tau_i\}_{i=1}^{G} \sim \pi_{\theta_{old}}(\tau|c)] \\ \frac{1}{G}\sum_{i=1}^{G}\frac{1}{|\tau_i|}\sum_{j=1}^{|\tau_i|}\Big(\frac{1}{|a_{i,j}|}\sum_{t=1}^{|a_{i,j}|}\{\min[w_{i,j,t}\hat{A}_{i,t}, \text{clip}\left(w_{i,j,t}, 1-\epsilon, 1+\epsilon\right)\hat{A}_{i,t}] - \beta\mathbb{D}_{\text{KL}}\}\Big),\end{aligned} \tag{4}$$

where $w_{i,j,t} = \frac{\pi_\theta(a_{i,j,t}|s_{i,j},a_{i,j,<t})}{\pi_{\theta_{old}}(a_{i,j,t}|s_{i,j},a_{i,j,<t})}$ is the importance sampling term, ϵ is the clip ratio and β is the ratio of KL penalty between policy model π_θ and reference model π_{ref}. $\hat{A}_{i,t}$ is the advantage calculated based on the relative rewards of the output reports inside each group. Given rewards $\mathbf{r} = \{R_1, R_2, ..., R_G\}$ of the output reports under the same group, the advantage is defined as $\hat{A}_{i,t} = \frac{R_i - \text{mean}(\mathbf{r})}{\text{std}(\mathbf{r})}$.

4 Experiment

In this section, we detail the experiments conducted to evaluate BioCheck Agent and EG-GRPO. Specifically, we outline the experimental settings, introduce the baseline methods, and discuss the final results with additional ablation study.

4.1 Experimental Settings

Datasets. We utilize two datasets for our experiments: SciFact [16] and HealthFC [6]. Because only SciFact provides explicit training, validation, and test splits, we used it for both training and evaluation, while using HealthFC only for evaluation. Furthermore, although both datasets include a 'Not Enough Information' (NEI) label, its meaning differs between the two datasets. In SciFact, NEI indicates an absence of evidence to support or refute a claim specifically within the provided corpus. Due to the limited size of this corpus, it does not necessarily imply a lack of information in real world. Conversely, in HealthFC, where the data is collected from human written fact-checking articles, NEI means that there is no existing evidence about the claim in real word, or that the available evidence is inherently conflicting. For simplicity, our default experiments perform training and evaluation only on examples with binary labels 'supported' and 'refuted'. Finally, because the claims in HealthFC are originally formatted as questions, we utilized an advanced large language model (GPT-5.2) to convert them into declarative statements, ensuring structural alignment with SciFact.

Base Models. We evaluate BioCheck Agent across different LLM backbones, including the proprietary GPT-5.2 [50] and the open-source Qwen3.5-4B [49]. For the EG-GRPO, we specifically perform the training on Qwen3.5-4B with SciFact training split.

Training Details. We implement EG-GRPO with the verl RL training library [51]. To adapt our agent for training, we integrate verl with the LangGraph framework to perform the rollout process. We apply EG-GRPO on the SciFact training set for 10 epochs with a learning rate of 1e-6. At least

4xNVIDIA A100 80GB GPUs are required to perform the training. During training, we perform evaluation on the validation set every 5 steps and select the checkpoint with the highest average reward as our final model. Additional hyperparameter details and training curves are provided in Appendix C.

Task Evaluation. We evaluate BioCheck Agent on SciFact test set and HealthFC with various metrics for both label prediction and report generation tasks. (1) **Label Prediction:** Following previous fact-checking works [23, 26], we evaluate label prediction by computing the accuracy (Acc), macro precision (Prec), macro recall (Rec), and macro F1 score (F1). For BioCheck Agent, we extract the predicted labels 'SUPPORTED', 'REFUTED', or 'NOT ENOUGH INFO' from the Conclusion section of the generated report. Then, these labels are evaluated against the ground-truth labels provided in the dataset. (2) **Report Generation:** To evaluate report generation, we propose two novel metrics: Evidence Quality Score (EQS) and Evidence Hallucination Rate (EHR). The EQS is calculated by averaging the mean score of the evidence within each report that aligns with the report's conclusion, across all examples. The EHR is defined as the ratio of hallucinated instances to the total number of generated reports. Higher EQS and lower EHR values indicate greater report quality.

Baseline Methods. We evaluate the performance of BioCheck Agent compared with the following baselines: (1) **No Retrieval**: A naive baseline directly applies an LLM for label prediction without any external information. (2) **CER** [23]: LLM classification with context provided by embedding-based retrieval. (3) **FIRE** [26]: An LLM-based agent equipped with Google Search. (4) **PMSearch Agent**: An LLM-based LangGraph agent shares the same architecture as BioCheck Agent, but outputs only the predicted label after conducting a search, rather than a comprehensive fact-checking report. (5) **BioCheck Agent with GRPO**: For comparison with EG-GRPO, we also implement a baseline GRPO on BioCheck Agent without the task-specific reward design.

While all baselines are evaluated on the label prediction task, existing literature lacks established benchmarks for evaluating fact-checking report generation. Therefore, our evaluation on report generation focuses primarily on comparing the base model against those optimized with baseline GRPO and EG-GRPO. Further implementation details for all baselines are provided in Appendix D.

4.2 Main Results

Label Prediction. We first present the Label Prediction results for both the SciFact and HealthFC in Table 1. As shown in the table, BioCheck Agent optimized with EG-GRPO consistently outperforms all baselines under the Qwen3.5-4B base model, including training-free agents and BioCheck Agent trained with baseline GRPO. Notably, applying EG-GRPO to the lightweight, open-source Qwen3.5-4B model enables it to surpass even the proprietary GPT-5.2 model, achieving an 8.9% improvement in prediction accuracy on SciFact. Furthermore, although not explicitly trained on the HealthFC dataset, our model still exhibits improved performance. While the baseline GRPO also improves BioCheck Agent performance on SciFact, it yields limited accuracy gains and a worse F1 score on HealthFC. This demonstrates that relying solely on an outcome-based reward causes the model behavior to easily overfit to the training dataset. In contrast, our EG-GRPO, with its evidence-aware reward design, achieves better transferability in label prediction by optimizing the report quality.

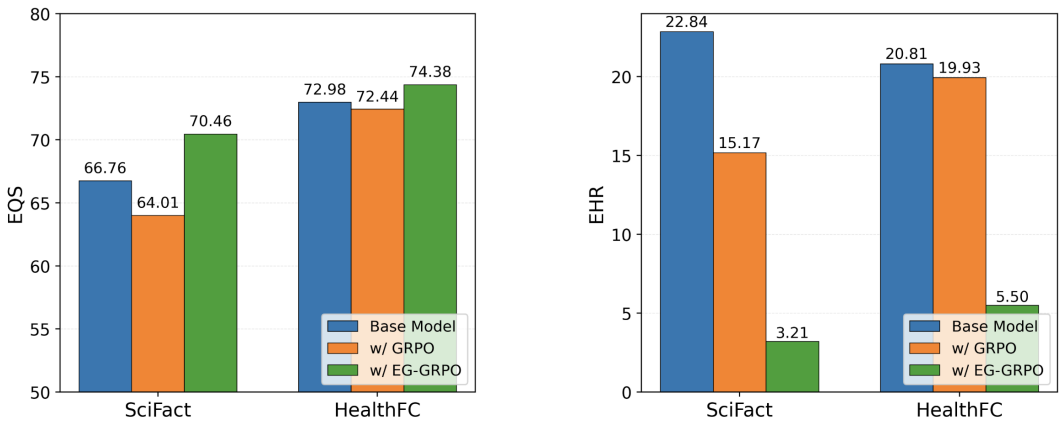
When comparing the results among training-free agents, we found that BioCheck Agent shows only marginal improvements in accuracy and F1 score compared to the PMSearch Agent on GPT-5.2, and exhibits even worse performance under Qwen3.5-4B. This is likely because report generation introduces additional complexity and requires advanced model capabilities to execute effectively. We also noted that the baseline method CER, which applies embedding-based retrieval within the PubMed database and re-ranking before justification, achieves high performance on SciFact but performs poorly on HealthFC. This discrepancy likely occurs because SciFact claims are sourced directly from PubMed, making embedding-based retrieval highly effective in-domain, whereas it struggles to process the more general, out-of-domain health claims in HealthFC. Furthermore, when comparing the two base models, we observe the most substantial performance disparity in the No Retrieval setting. This demonstrates that while GPT-5.2 possesses superior internal parametric knowledge, the introduction of external evidence retrieval effectively bridges this gap.

Report Generation. For the report generation task, we evaluate the Evidence Quality Score (EQS) and Evidence Hallucination Rate (EHR) respectively in Figure 2a and Figure 2b on both the SciFact and HealthFC datasets with the base model Qwen3.5-4B. As shown in the figures, compared to

Table 1: Label prediction performance of BioCheck Agent against various baselines. For each Base Model, the highest value is bolded and the second highest is underlined.

Base Model	Method	SciFact				HealthFC			
		Acc	Prec	Rec	F1	Acc	Prec	Rec	F1
GPT5.2	No Retrieval	66.49	84.41	65.47	71.74	60.86	**84.22**	53.68	59.52
	CER [23]	78.53	**96.25**	78.73	**86.47**	54.13	82.95	47.01	53.00
	FIRE [26]	**82.72**	85.59	**83.54**	82.57	**74.01**	72.58	**73.01**	72.76
	PMSearch Agent	78.01	92.12	78.24	84.43	70.95	77.36	67.79	71.78
	BioCheck Agent	81.15	88.98	81.57	84.68	72.48	75.34	71.47	**73.34**
Qwen3.5-4B	No Retrieval	59.16	72.11	58.66	64.45	50.15	64.56	44.56	50.59
	CER [23]	75.92	88.86	75.84	81.80	52.60	73.88	47.76	56.23
	FIRE [26]	81.15	82.33	81.69	81.11	71.87	70.50	70.37	70.43
	PMSearch Agent	81.15	90.01	81.09	85.32	70.64	76.29	66.48	69.90
	BioCheck Agent	80.10	89.22	80.40	84.30	70.34	74.91	66.99	70.28
	w/ GRPO	89.01	90.27	88.76	89.38	72.17	74.45	65.73	66.10
	w/ **EG-GRPO**	**90.05**	**91.02**	**90.17**	**90.51**	**77.37**	**79.51**	**72.38**	**73.59**

BioCheck Agent with the base model and baseline GRPO, BioCheck Agent optimized with EG-GRPO generates biomedical fact-checking reports with a higher EQS and a significantly reduced EHR. This reflects higher average evidence quality with fewer hallucinations, demonstrating the effectiveness of EG-GRPO in improving the overall quality of biomedical fact-checking reports. Additionally, we find that BioCheck Agent with baseline GRPO results in both lower EQS and EHR. This demonstrates that without the specific reward design used in EG-GRPO, baseline GRPO still impacts report generation quality when optimizing the outcome-based reward. However, lacking specific guidance, it yields only a marginal reduction in EHR while suffering from an even lower EQS. Combined with the results in Table 1, we demonstrate that our EG-GRPO can improve BioCheck Agent in both label prediction and report generation performance.



(a) (b)

Figure 2: Sub-figures (a) and (b) evaluate the EQS (%) and EHR (%), respectively, for the generated fact-checking reports across various models.

Qualitative Analysis of Generated Reports. To provide a better understanding of the quality of generated reports, Appendix E includes two sample reports: one generated by the base model (Qwen3.5-4B) and another by the model optimized with EG-GRPO. A comparison of these two examples reveals a clear difference in quality. The base model cited two mismatched PMIDs and retrieved irrelevant evidence, ultimately leading to an erroneous conclusion. In contrast, BioCheck Agent optimized by EG-GRPO retrieved evidence with much higher quality. After a critical analysis of the context, it finally made the correct judgment.

4.3 Ablation Study

Evidence Confidence Score with Alternative Judge LLM. In our default setting, we use the same Qwen3.5-4B model as the judge LLM to compute the evidence confidence score during training with EG-GRPO and for the EQS evaluation. To demonstrate that the improvements in EQS do not rely on the specific judge LLM used during training, we perform additional evaluations using an alternative LLM, GPT-4o [52] to compute the evidence confidence score. GPT-4o is currently the most advanced GPT model that provides logprob outputs. The prompt and implementation are identical to those used for the Qwen3.5-4B model, as detailed in Appendix B. We present the EQS results using GPT-4o as the judge in Table 2. The highest EQS among the various agent models is highlighted in bold.

Table 2: EQS (%) evaluation with alternative judge LLM.

Agent Model	Qwen3.5-4B		GPT-4o	
	SciFact	HealthFC	SciFact	HealthFC
Qwen3.5-4B	66.76	72.98	64.73	73.83
w/ GRPO	64.01	72.44	63.60	71.65
w/ EG-GRPO	**70.46**	**74.38**	**69.92**	**75.84**

As shown in the table, although evaluating with GPT-4o yields different EQS values, the overall trend remains consistent: BioCheck Agent optimized with EG-GRPO continues to outperform the base model and the baseline GRPO in generating fact-checking reports with higher evidence quality.

5 Conclusion

In this paper, we mitigate the pervasive issue of health misinformation by advancing automated biomedical fact-checking from a traditional text classification task to comprehensive report generation. Rather than merely outputting an isolated prediction label, fact-checking reports provide contextualized information, serving as a much better reference for human decision-makers. To achieve this, we propose BioCheck Agent to generate structured biomedical fact-checking reports by performing agentic search within the PubMed database. To further enhance the agent's performance, we propose an end-to-end RL method named Evidence-Grounded GRPO (EG-GRPO), which utilizes a task-specific reward to optimize the groundedness and quality of the generated evidence in reports. Comprehensive experiments demonstrate the efficacy of BioCheck Agent optimized with EG-GRPO, achieving both accurate label prediction and high-quality report generation for biomedical fact-checking, making it possible to assist humans in discerning health misinformation.

6 Limitation

While BioCheck Agent provides an effective automated approach for biomedical fact-checking, its current scope is primarily limited to isolated atomic claims. In real-world scenarios, health misinformation is rarely presented in such a clean format. Instead, it is normally embedded within complex narratives across diverse platforms (e.g., social media posts, news articles) and often relies on multi-modal contexts, such as images and videos. Extending BioCheck Agent to verify claims within these environments remains a significant challenge for future work.

A second limitation concerns the lack of high-quality data for training our agent. In fact, the SciFact training set contains only around 500 labeled examples and lacks Not Enough Info (NEI) cases that frequently occur in the wild. Furthermore, because SciFact is constructed from scientific claims extracted from PubMed abstracts, there is a notable distribution shift when applying the agent model to general public health. This gap largely explains why BioCheck Agent presents a comparably modest performance improvement on the HealthFC dataset compared to SciFact.

References

[1] NGOZI COMFORT Omojunikanbi. Public relations and effective communication during a global health crisis: Combating disinformation, misinformation, and fake news on covid-19. *Journal of Communication and Media Research*, 14(1):64–71, 2022.

[2] Md Saiful Islam, Tonmoy Sarkar, Sazzad Hossain Khan, Abu-Hena Mostofa Kamal, SM Murshid Hasan, Alamgir Kabir, Dalia Yeasmin, Mohammad Ariful Islam, Kamal Ibne Amin Chowdhury, Kazi Selim Anwar, et al. Covid-19–related infodemic and its impact on public health: A global social media analysis. *The American journal of tropical medicine and hygiene*, 103(4):1621, 2020.

[3] Sahil Loomba, Alexandre De Figueiredo, Simon J Piatek, Kristen De Graaf, and Heidi J Larson. Measuring the impact of covid-19 vaccine misinformation on vaccination intent in the uk and usa. *Nature human behaviour*, 5(3):337–348, 2021.

[4] Jon Roozenbeek, Claudia R Schneider, Sarah Dryhurst, John Kerr, Alexandra LJ Freeman, Gabriel Recchia, Anne Marthe Van Der Bles, and Sander Van Der Linden. Susceptibility to misinformation about covid-19 around the world. *Royal Society open science*, 7(10), 2020.

[5] Neema Kotonya and Francesca Toni. Explainable automated fact-checking for public health claims. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7740–7754, 2020.

[6] Juraj Vladika, Phillip Schneider, and Florian Matthes. Healthfc: Verifying health claims with evidence-based medical fact-checking. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 8095–8107, 2024.

[7] Connie Schardt, Martha B Adams, Thomas Owens, Sheri Keitz, and Paul Fontelo. Utilization of the pico framework to improve searching pubmed for clinical questions. *BMC medical informatics and decision making*, 7(1):16, 2007.

[8] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems*, 33:9459–9474, 2020.

[9] Bowen Jin, Hansi Zeng, Zhenrui Yue, Jinsung Yoon, Sercan Arik, Dong Wang, Hamed Zamani, and Jiawei Han. Search-r1: Training llms to reason and leverage search engines with reinforcement learning. *arXiv preprint arXiv:2503.09516*, 2025.

[10] Jiho Kim, Sungjin Park, Yeonsu Kwon, Yohan Jo, James Thorne, and Edward Choi. Factkg: Fact verification via reasoning on knowledge graphs. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 16190–16206, 2023.

[11] Ronit Singal, Pransh Patwa, Parth Patwa, Aman Chadha, and Amitava Das. Evidence-backed fact checking using rag and few-shot in-context learning with llms. In *Proceedings of the Seventh Fact Extraction and VERification Workshop (FEVER)*, pages 91–98, 2024.

[12] Yuxia Wang, Revanth Gangi Reddy, Zain Muhammad Mujahid, Arnav Arora, Aleksandr Rubashevskii, Jiahui Geng, Osama Mohammed Afzal, Liangming Pan, Nadav Borenstein, Aditya Pillai, et al. Factcheck-bench: Fine-grained evaluation benchmark for automatic fact-checkers. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 14199–14230, 2024.

[13] Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

[14] Lang Feng, Zhenghai Xue, Tingcong Liu, and Bo An. Group-in-group policy optimization for llm agent training. *arXiv preprint arXiv:2505.10978*, 2025.

[15] Hanchen Zhang, Xiao Liu, Bowen Lv, Xueqiao Sun, Bohao Jing, Iat Long Iong, Zhenyu Hou, Zehan Qi, Hanyu Lai, Yifan Xu, et al. Agentrl: Scaling agentic reinforcement learning with a multi-turn, multi-task framework. *arXiv preprint arXiv:2510.04206*, 2025.

[16] David Wadden, Shanchuan Lin, Kyle Lo, Lucy Lu Wang, Madeleine van Zuylen, Arman Cohan, and Hannaneh Hajishirzi. Fact or fiction: Verifying scientific claims. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7534–7550, 2020.

[17] Xia Zeng, Amani S Abumansour, and Arkaitz Zubiaga. Automated fact-checking: A survey. *Language and Linguistics Compass*, 15(10):e12438, 2021.

[18] Zhijiang Guo, Michael Schlichtkrull, and Andreas Vlachos. A survey on automated fact-checking. *Transactions of the association for computational linguistics*, 10:178–206, 2022.

[19] Andreas Vlachos and Sebastian Riedel. Fact checking: Task definition and dataset construction. In *Proceedings of the ACL 2014 workshop on language technologies and computational social science*, pages 18–22, 2014.

[20] Nayeon Lee, Belinda Z Li, Sinong Wang, Wen-tau Yih, Hao Ma, and Madian Khabsa. Language models as fact checkers? In *Proceedings of the third workshop on fact extraction and verification (FEVER)*, pages 36–41, 2020.

[21] David Wadden, Kyle Lo, Lucy Lu Wang, Arman Cohan, Iz Beltagy, and Hannaneh Hajishirzi. Multivers: Improving scientific claim verification with weak supervision and full-document context. In *Findings of the association for computational linguistics: NAACL 2022*, pages 61–76, 2022.

[22] James Thorne, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal. Fever: a large-scale dataset for fact extraction and verification. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 809–819, 2018.

[23] Mariano Barone, Antonio Romano, Giuseppe Riccio, Marco Postiglione, and Vincenzo Moscato. Combining evidence and reasoning for biomedical fact-checking. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1087–1097, 2025.

[24] Jiho Kim, Yeonsu Kwon, Yohan Jo, and Edward Choi. Kg-gpt: A general framework for reasoning on knowledge graphs using large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 9410–9421, 2023.

[25] Jerry Wei, Chengrun Yang, Xinying Song, Yifeng Lu, Nathan Hu, Jie Huang, Dustin Tran, Daiyi Peng, Ruibo Liu, Da Huang, et al. Long-form factuality in large language models. *Advances in Neural Information Processing Systems*, 37:80756–80827, 2024.

[26] Zhuohan Xie, Rui Xing, Yuxia Wang, Jiahui Geng, Hasan Iqbal, Dhruv Sahnan, Iryna Gurevych, and Preslav Nakov. Fire: fact-checking with iterative retrieval and verification. In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 2901–2914, 2025.

[27] Oberiri Destiny Apuke and Bahiyah Omar. Fake news and covid-19: modelling the predictors of fake news sharing among social media users. *Telematics and informatics*, 56:101475, 2021.

[28] Mourad Sarrouti, Asma Ben Abacha, Yassine M'rabet, and Dina Demner-Fushman. Evidence-based fact-checking of health-related claims. In *Findings of the association for computational linguistics: EMNLP 2021*, pages 3499–3512, 2021.

[29] Isabelle Mohr, Amelie Wührl, and Roman Klinger. Covert: A corpus of fact-checked biomedical covid-19 tweets. In *Proceedings of the thirteenth language resources and evaluation conference*, pages 244–257, 2022.

[30] Caralyn Reisle, Cameron J Grisdale, Kilannin Krysiak, Arpad M Danos, Mariam Khanfar, Erin Pleasance, Jason Saliba, Melika Hanos, Nilan V Patel, Asmita Jain, et al. Evaluating language models for biomedical fact-checking: a benchmark dataset for cancer variant interpretation verification. *bioRxiv*, 2025.

[31] Hao Liu, Ali Soroush, Jordan G Nestor, Elizabeth Park, Betina Idnay, Yilu Fang, Jane Pan, Stan Liao, Marguerite Bernard, Yifan Peng, et al. Retrieval augmented scientific claim verification. *JAMIA open*, 7(1):ooae021, 2024.

[32] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik R Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. In *The eleventh international conference on learning representations*, 2022.

[33] Lutfi Eren Erdogan, Nicholas Lee, Sehoon Kim, Suhong Moon, Hiroki Furuta, Gopala Anumanchipalli, Kurt Keutzer, and Amir Gholami. Plan-and-act: Improving planning of agents for long-horizon tasks. In *International Conference on Machine Learning*, pages 15419–15462. PMLR, 2025.

[34] Xingyao Wang, Yangyi Chen, Lifan Yuan, Yizhe Zhang, Yunzhu Li, Hao Peng, and Heng Ji. Executable code actions elicit better llm agents. In *Forty-first International Conference on Machine Learning*, 2024.

[35] langchain ai. langgraph, 2026.

[36] Xingyao Wang, Boxuan Li, Yufan Song, Frank F Xu, Xiangru Tang, Mingchen Zhuge, Jiayi Pan, Yueqi Song, Bowen Li, Jaskirat Singh, et al. Openhands: An open platform for ai software developers as generalist agents. *arXiv preprint arXiv:2407.16741*, 2024.

[37] OpenAI. Function calling and other api updates, 2023. `https://openai.com/index/function-calling-and-other-api-updates/`.

[38] Xiaoxi Li, Guanting Dong, Jiajie Jin, Yuyao Zhang, Yujia Zhou, Yutao Zhu, Peitian Zhang, and Zhicheng Dou. Search-o1: Agentic search-enhanced large reasoning models. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 5420–5438, 2025.

[39] Yihan Li, Xiyuan Fu, Ghanshyam Verma, Paul Buitelaar, and Mingming Liu. Mitigating hallucination in large language models (llms): An application-oriented survey on rag, reasoning, and agentic systems. *arXiv preprint arXiv:2510.24476*, 2025.

[40] Dominique Kelly, Yimin Chen, Sarah E Cornwell, Nicole S Delellis, Alex Mayhew, Sodiq Onaolapo, and Victoria L Rubin. Bing chat: The future of search engines? *Proceedings of the Association for Information Science and Technology*, 60(1):1007–1009, 2023.

[41] OpenAI. Introducing deep research, 2025. `https://openai.com/index/introducing-deep-research/`.

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

[43] Zhepei Wei, Wenlin Yao, Yao Liu, Weizhi Zhang, Qin Lu, Liang Qiu, Changlong Yu, Puyang Xu, Chao Zhang, Bing Yin, et al. Webagent-r1: Training web agents via end-to-end multi-turn reinforcement learning. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 7920–7939, 2025.

[44] Jiongxiao Wang, Qiaojing Yan, Yawei Wang, Yijun Tian, Soumya Smruti Mishra, Zhichao Xu, Megha Gandhi, Panpan Xu, and Lin Lee Cheong. Reinforcement learning for self-improving agent with skill library. *arXiv preprint arXiv:2512.17102*, 2025.

[45] Kaituo Feng, Kaixiong Gong, Bohao Li, Zonghao Guo, Yibing Wang, Tianshuo Peng, Junfei Wu, Xiaoying Zhang, Benyou Wang, and Xiangyu Yue. Video-r1: Reinforcing video reasoning in mllms. *arXiv preprint arXiv:2503.21776*, 2025.

[46] langchain ai. langchain, 2026.

[47] Zhiheng Xi, Yiwen Ding, Wenxiang Chen, Boyang Hong, Honglin Guo, Junzhe Wang, Dingwen Yang, Chenyang Liao, Xin Guo, Wei He, Songyang Gao, Lu Chen, Rui Zheng, Yicheng Zou, Tao Gui, Qi Zhang, Xipeng Qiu, Xuanjing Huang, Zuxuan Wu, and Yu-Gang Jiang. Agentgym: Evolving large language model-based agents across diverse environments, 2024.

[48] Zhiheng Xi, Jixuan Huang, Chenyang Liao, Baodai Huang, Honglin Guo, Jiaqi Liu, Rui Zheng, Junjie Ye, Jiazheng Zhang, Wenxiang Chen, et al. Agentgym-rl: Training llm agents for long-horizon decision making through multi-turn reinforcement learning. *arXiv preprint arXiv:2509.08755*, 2025.

[49] QwenTeam. Qwen3.5: Towards native multimodal agents, 2026. `https://qwen.ai/blog?id=qwen3.5`.

[50] OpenAI. Introducing gpt-5.2, 2025. `https://openai.com/index/introducing-gpt-5-2/`.

[51] Guangming Sheng, Chi Zhang, Zilingfeng Ye, Xibin Wu, Wang Zhang, Ru Zhang, Yanghua Peng, Haibin Lin, and Chuan Wu. Hybridflow: A flexible and efficient rlhf framework. In *Proceedings of the Twentieth European Conference on Computer Systems*, pages 1279–1297, 2025.

[52] Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, et al. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*, 2024.

[53] Stephen Robertson and Hugo Zaragoza. *The probabilistic relevance framework: BM25 and beyond*, volume 4. Now Publishers Inc, 2009.

A Implementation Details of BioCheck Agent

This section presents how we use the LangGraph framework to build BioCheck Agent.

A.1 LangGraph

A demonstration figure for the structure of our LangGraph agent is presented in Figure 3.

Based on the LangGraph structure shown in the figure, BioCheck Agent starts at an 'LLM' node that processes the initial prompts. A conditional edge then directs the flow: if tool calls are detected in LLM outputs, it triggers the 'Tool' node; if not, the process ends. The 'Tool' node then executes the generated tool calls in LLM outputs and tracks the number of tool calls. If the number exceeds a maximum tool call limit, the flow is directed to an 'Intervention' node where an extra user prompt is injected to force a conclusion. If under the limit, the agent cycles back to the LLM for further iteration.

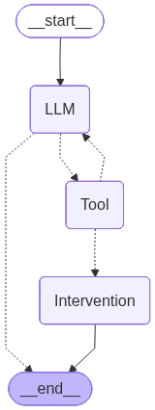


Figure 3: LangGraph structure of BioCheck Agent.

A.2 Prompt

We outline the prompts utilized in BioCheck Agent, including a system prompt that defines the agent's role and task; an initial user prompt containing the core task instructions; and an intervention prompt (injected as a user message), which triggers when the maximum search limit is reached to terminate searching and force the generation of the final fact-checking report.

System Prompt

> You are a biomedical expert. Your goal is to analyze a provided claim with the PubMed database by using the 'search_pubmed' tool. Based on your search results, synthesize your findings into a brief report and determine the most appropriate label: supported or refuted.

Initial User Prompt to start the agent. The red-highlighted text indicates a placeholder for the claim undergoing fact-checking.

> Follow this strict protocol to verify the claim:
>
> **AGENT PROTOCOL**
> 1. Analyze & Retrieve: Analyze the claim to identify Core Entities (with expanded aliases/MeSH terms), the Relation, and the Context. Construct a comprehensive Boolean query using advanced operators (AND/OR) connecting these keywords, and use the 'search_pubmed' tool to execute your initial search.
> 2. Iterate with Adaptive Strategies: If the search yields insufficient evidence, iteratively refine your query and continue searching until you collect sufficient evidence or hit the maximum number of searches (5). Apply the following strategies dynamically during your refinement process:
> - Constraint Relaxation: Drop Context or Relation limits if you get 0 hits or too little evidence.
> - Adaptive Incorporation: Extract newly discovered synonyms, related pathways, or methodologies from abstracts to formulate more precise follow-up queries.
> - Active Refutation: Actively search for contradictory evidence (e.g., append "fails to", "opposite", or "artifact") if you find 0 supporting hits or are stuck in a one-sided confirmation loop.
> 3. Final Report & Conclusion: Once you have gathered sufficient evidence to make a definitive

conclusion, you MUST output your findings STRICTLY in the format below:

REPORT
Supporting Evidence:
[For each supporting article, list:]
- PMID: [e.g., 12345678]
- Top Sentence: [Direct, verbatim quote from the abstract enclosed in double quotation marks to support the claim]
- Context: [Note any specific conditions, e.g., "Only in murine models"]
(If none found, write: "No supporting evidence found.")

Refuting Evidence:
[For each refuting article, list:]
- PMID: [e.g., 87654321]
- Top Sentence: [Direct, verbatim quote from the abstract enclosed in double quotation marks to support the claim]
- Context: [Note any specific conditions]
(If none found, write: "No refuting evidence found.")

Summary
- Reasons to Support: [2-3 sentences analyzing and summarizing findings to support the claim]
- Reasons to Refute: [2-3 sentences analyzing and summarizing findings to refute the claim]
- Consensus & Context Check: [Evaluate if the claim's scope matches the evidence (e.g., is it a universal truth or a special case?)]
- Final Justification: [1-2 sentences explaining the final decision]

Conclusion
[Output STRICTLY ONE of the tags below on a new line]
<answer>SUPPORTED</answer>
OR
<answer>REFUTED</answer>

Please verify this claim now: CLAIM

Intervention Prompt

Attention: You have reached the maximum step limit. Do not call any more tools. Please provide the final report with conclusion based on the information collected so far. If the current information is insufficient to make the conclusion, please respond with '<answer>NOT ENOUGH INFORMATION</answer>' as your final conclusion after the report.

B Computing the Evidence Confidence Score

To compute the Composite Evidence Reward and evaluate the Evidence Quality Score, we use an LLM to calculate the evidence confidence score that the provided evidence supports or refutes the claim. Since this is a classification task rather than text generation, we frame it as a multiple-choice problem and compute the probability of each choice's token. Specifically, we prompt the LLM to output a choice from 'A: supported', 'B: refuted', or 'C: not enough information' based on the provided evidence and claim. Instead of answer generation, we only consider the next token's logprob outputs z and compute the probability score $s = \exp(z)$ for the corresponding token 'A', 'B', or 'C'.

The prompt for the LLM is shown below. Here, the red-highlighted 'EVIDENCE' and 'CLAIM' serve as placeholders.

Given the abstract and the claim, decide whether the claim is supported, refuted by the abstract, or if there is not enough information to make a conclusion.

Abstract: EVIDENCE

Claim: CLAIM

Return exactly one choice for the claim verification result: 'A: supported', 'B: refuted', or 'C: not enough information'.

C Training Details

This section includes the training details of applying EG-GRPO on BioCheck Agent.

Regarding the reward modeling hyperparameters, we utilize the default values defined in Section 3.2.2. Specifically, we set the saturation factor $\alpha = 1.0$ for R_{CER}. For the Biocheck Reward, formulated as $R_{\text{outcome}} + \beta_1 R_{\text{search}} + \beta_2 R_{\text{evidence}} - \beta_3 R_{\text{hallucination}}$, we set $\beta_1 = 0.5$ and $\beta_2 = \beta_3 = 1.0$.

The key hyperparameters configured for the verl reinforcement learning library are detailed in the following Table 3.

Table 3: Hyperparameters for EG-GRPO.

Hyperparameter	**Value**
total epochs	10
global batch size	64
mini batch size	64
learning rate	1e-6
KL loss coefficient	0.03
max response length	16000
clip ratio low	0.2
clip ratio high	0.28
num of agents in group	8

The entire training process has 70 steps, with the corresponding training curve depicted in Figure 4. We select the checkpoint at step 60 as our final agent model, as it achieves the highest average reward on the validation set.

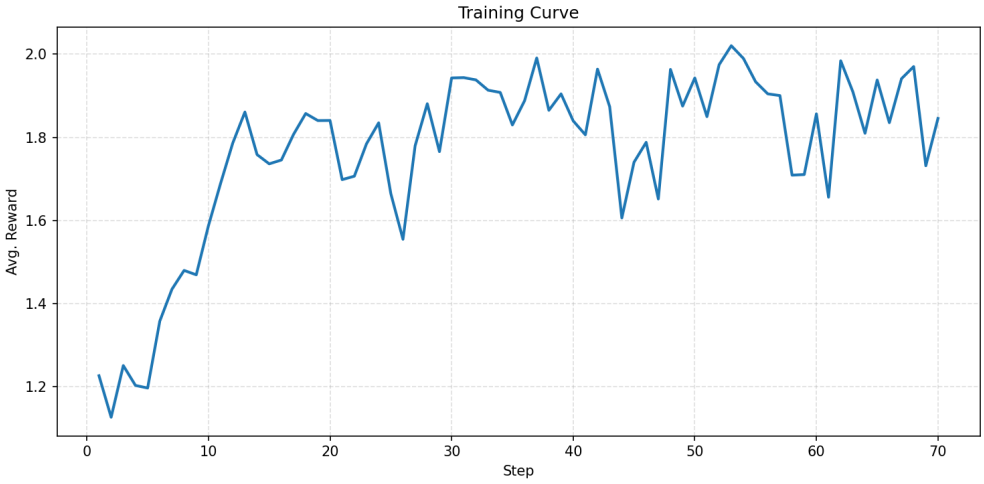


Figure 4: Training Curve for EG-GRPO.

D Baseline Methods

In this section, we detail the implementation of the baseline methods shown in Table 1.

No Retrieval. As a naive baseline, we prompt the LLM to fact-check the claim using only its internal knowledge. The corresponding system and user prompts are provided below, with the placeholder for the claim highlighted in red.

System Prompt

> You are a biomedical expert. Your goal is to analyze a provided claim based on your biomedical expertise and classify it using one of three labels.

User Prompt

> To verify the claim, directly provide your classification result exactly as '<answer>SUPPORTED</answer>', '<answer>REFUTED</answer>', or '<answer>NOT ENOUGH INFORMATION</answer>'.
>
> Please verify this claim now: CLAIM

CER [23]. We implement Combining Evidence and Reasoning (CER) as a retrieval-augmented biomedical fact-checking baseline following an evidence–reasoning–classification pipeline. For each input claim, we retrieve candidate evidence from an external PubMed abstract corpus using a sparse BM25 retriever [53]. The retrieval index is built over PubMed abstracts with standard text preprocessing, including lower-casing, stemming, stop-word removal, and normalization of punctuation, special characters, and acronyms.

Given a claim, the retriever first returns the top 20 PubMed abstracts. Each retrieved abstract is then segmented into sentences, and candidate evidence sentences are reranked using a biomedical sentence encoder. We encode the claim and candidate sentences in the same embedding space and select the top 3 sentences from each retrieved abstract according to cosine similarity. This step yields an evidence set of up to 60 PubMed sentences per claim.

The original CER framework prompts an LLM with retrieved scientific evidence for evidence-grounded reasoning and then trains a supervised classifier to produce the final veracity prediction. Our implementation follows the same retrieval and reasoning design but replaces the trained classifier with an LLM for zero-shot decision. In the first stage, the LLM is prompted as a biomedical reasoning model and is provided only with the retrieved PubMed evidence. It generates a preliminary decision together with a concise justification grounded in the retrieved evidence. We follow the prompt design of the original CER framework for this reasoning stage and refer readers to [23] for the full prompt specification. In the second stage, the LLM receives the claim, retrieved evidence, preliminary decision, and generated justification to output the final decision. In summary, compared with the original CER framework, our implementation preserves the evidence retrieval and reasoning components but uses an LLM, rather than a trained classifier, to make the final veracity decision.

FIRE [26]. We use Fact-checking with Iterative Retrieval and Verification (FIRE) as an iterative retrieval-and-verification baseline for atomic claims. Unlike fixed-depth retrieval pipelines, FIRE allows the verifier LLM to adaptively determine whether the currently available evidence is sufficient for a final factuality judgment or whether additional evidence should be retrieved. At each iteration, the model receives the claim together with the accumulated evidence snippets and either returns a final decision or issues a new Google search query. We follow the original FIRE prompting protocol for iterative verification and query generation and refer readers to [26] for the full system and user prompts.

When additional evidence is requested, we retrieve web evidence through Google Search using the Serper API and collect the top 3 results for each query. The evidence pool is constructed from available search-result snippets, including answer-box content, knowledge-graph descriptions, and organic-result snippets. Thus, FIRE differs from corpus-based baselines in that it does not rely on a

fixed offline evidence collection; instead, it verifies claims against evidence retrieved from the live web.

The iterative retrieval-and-verification loop runs for at most 5 steps, with up to 10 retries allowed when the model output does not match the required format. To reduce redundant retrieval, the loop can terminate early if the model repeatedly generates semantically similar queries or retrieves highly similar evidence. Semantic similarity is computed using `all-MiniLM-L6-v2` sentence embeddings with cosine similarity, with a threshold of 0.9. If the search budget is exhausted before a final decision is produced, FIRE performs a final verification step using all accumulated evidence. The resulting prediction is mapped to a binary factuality label.

PMSearch Agent. To ensure a fair comparison within the same LangGraph framework, we developed a baseline agent named PMSearch Agent. This baseline shares the same LangGraph structure as BioCheck Agent. It also utilizes the 'pubmed_search' tool to retrieve 5 relevant papers per search, with a maximum limit of 5 search iterations. The system prompt, initial user prompt, and intervention prompt for the PMSearch Agent are presented below.

System Prompt

> You are a biomedical expert. Your goal is to determine if a provided claim is supported or refuted with the PubMed database. To access the PubMed database, you can use the 'search_pubmed' tool to retrieve relevant papers.

Initial User Prompt with claim placeholder highlighted.

> Follow this strict protocol to verify the claim:
>
> **AGENT PROTOCOL**
> 1. Retrieve Evidence with Search Tool: Use the 'search_pubmed' tool with targeted search query to find papers that can verify the claim.
> 2. Iterate if Necessary: If the search yields no results (0 hits) or the search result is insufficient to make a conclusion, iteratively refine your query and continue searching until you reach the final conclusion or hit the maximum number of searches (5).
> 3. Final Decision: Once you have gathered sufficient evidence to make a definitive conclusion, provide a brief explanation followed strictly by '<answer>SUPPORTED</answer>' or '<answer>REFUTED</answer>' as your final conclusion.
>
> Please verify this claim now: CLAIM

Intervention Prompt

> Attention: You have reached the maximum step limit. Do not call any more tools. Please provide the best possible final answer based on the information collected so far. If the current information is insufficient to make a decision, please respond with '<answer>NOT ENOUGH INFORMATION</answer>'.

BioCheck Agent with GRPO. We compare our proposed EG-GRPO against a standard GRPO baseline applied to BioCheck Agent. The baseline GRPO relies solely on an outcome-based reward, evaluating whether the generated report's final conclusion matches the ground truth label. To ensure a fair comparison, all training parameters remain identical between the two approaches within the verl RL library. We apply GRPO to the SciFact training split, evaluating the average validation reward every 5 steps to identify the optimal checkpoint. For BioCheck Agent with GRPO, the highest average reward was achieved at step 60, which we consequently selected as our final agent model.

E Examples of Generated Fact-Checking Report

This section includes two examples of biomedical fact-checking reports generated by BioCheck Agent with the base model Qwen3.5-4B and the model optimized with EG-GRPO respectively. The

claim to verify is "Azo food dyes cause attention deficit disorder and hyperactivity in children." with the ground truth label "SUPPORTED". Red colored text within brackets indicates warnings about the mismatched PMIDs and low quality evidence.

Fact-Checking Report Generated by BioCheck Agent with the Base Model

REPORT
Supporting Evidence:
- PMID: 17825405
- Top Sentence: "Artificial colours or a sodium benzoate preservative (or both) in the diet result in increased hyperactivity in 3-year-old and 8/9-year-old children in the general population."
- Context: Randomized, double-blinded, placebo-controlled crossover trial in UK children

- PMID: 19882794
- Top Sentence: "A meta-analysis of 15 double-blind clinical trials that evaluated artificial food colouring in children already considered to be hyperactive showed an increase in their hyperactive behavior."
- Context: Meta-analysis of 15 double-blind RCTs across studies evaluating food colorings

- PMID: 33953077 [PMID 33953077 does not exist within the retrieved evidence list.]
- Top Sentence: "Forty (72.7%) demonstrated improved behaviour and 26 (47.3%) remained improved following liberalization of the diet... Two significant reactors were identified whose behavioural pattern featured extreme irritability, restlessness and sleep disturbance."
- Context: Double-blind crossover study on children with suspected hyperactivity

- PMID: 7169286 [Low evidence confidence score supporting the claim: $s = 0.09$.]
- Top Sentence: "The hyperactive children exhibited a greater degree of physiological activity following both the placebo and the drink containing food additives than did the control children. However, the magnitude of physiological changes in the hyperactive children were greater in response to ingestion of the additives than to placebo."
- Context: Randomized double-blind crossover study measuring heart rate and EEG in hyperactive children

- PMID: 6400044 [Low evidence confidence score supporting the claim: $s = 0.05$.]
- Top Sentence: "Only a subset of hyperactive children seem to be affected by food colors."
- Context: Review discussing individual sensitivity

Refuting Evidence:
- PMID: 8747098
- Top Sentence: "Numerous double-blind studies of the Feingold hypothesis have led to the rejection of the idea that this elimination diet has any benefit beyond the normal placebo effect."
- Context: Review of dietary elimination trials

- PMID: 24217026
- Top Sentence: "Capsule A containing AFCs and Capsule B containing sodium benzoate had no significant adverse effect compared with placebo in both behavior scores."
- Context: Randomized, double-blind, placebo-controlled trial in 130 Hong Kong children aged 8-9 years

- PMID: 6 [PMID 6 does not exist within the retrieved evidence list.]
- Top Sentence: "When the best probably blinded assessment was employed, effects remained significant for free fatty acid supplementation... but were substantially attenuated to nonsignificant levels for other treatments."
- Context: Meta-analysis of dietary and psychological ADHD treatments

Summary
- Reasons to Support: Multiple controlled studies (especially meta-analyses and RCTs) demonstrate associations between artificial food colorings and increased hyperactivity/ADHD symptoms in children. Physiological measurements show differential responses to additives versus placebo in hyperactive children.

- Reasons to Refute: Several high-quality double-blind placebo-controlled trials found no significant association between food colorings and behavior in general population children

(Hong Kong trial, Feingold diet review requiring blinded assessment). Some effects appear limited to subgroups of "sensitive" children.

- Consensus & Context Check: The evidence is conflicting and inconsistent across studies. The claim of a universal causal relationship is not supported - effects appear to be variable, potentially limited to certain age groups or sensitive individuals rather than a blanket effect.

- Final Justification: Given the contradictory evidence from multiple quality-controlled studies, the claim that azo food dyes universally cause ADHD/hyperactivity in children cannot be definitively supported. The relationship appears limited to certain populations, though some studies do show associations.

Conclusion
<answer>REFUTED</answer>

Fact-Checking Report Generated by BioCheck Agent with EG-GRPO

REPORT
Supporting Evidence:
- PMID: 38423749
- Top Sentence: "A systematic review has established that artificial azo dye food colourants (AFCs) and sodium benzoate preservative cause disturbed behaviour in children."
- Context: Systematic review of observational and animal model studies

- PMID: 15613992
- Top Sentence: "Meta-analytic modeling determined the overall effect size of AFCs on hyperactivity to be 0.283 (95% CI, 0.079 to 0.488), falling to 0.210 (95% CI, 0.007 to 0.414) when the smallest and lowest quality trials were excluded."
- Context: Analysis of 15 double-blind placebo-controlled trials; acknowledges publication bias

- PMID: 7965420
- Top Sentence: "Behavioral changes in irritability, restlessness, and sleep disturbance are associated with the ingestion of tartrazine in some children. A dose response effect was obtained."
- Context: Double-blind placebo-controlled study; significant reactions observed at all six dose levels tested (1-50mg tartrazine)

- PMID: 2857900
- Top Sentence: "Artificial colorants and preservatives were the commonest provoking substances, but no child was sensitive to these alone."
- Context: Oligoantigenic diet trial with 76 overactive children (62 improved)

Refuting Evidence:
- PMID: 20013441
- Top Sentence: "For both groups, mean intakes of the food additives among consumers only were far below the doses used in the previous study on hyperactivity. No child or teenager achieved the overall intakes used in the study linking food additives with hyperactivity."
- Context: Irish children and teenagers; dietary intake below threshold doses used in challenge studies

- PMID: 8277950
- Top Sentence: "Even when intake exceeds typical dietary levels, neither dietary sucrose nor aspartame affects children's behavior or cognitive function."
- Context: Double-blind trial in normal preschool and school-age children; no artificial food coloring in diets tested

Summary
- Reasons to Support: Multiple double-blind controlled studies and a meta-analysis demonstrate that a subgroup of children (approximately 23-89% depending on regimen) experience irritability, restlessness, and sleep disturbance when challenged with specific azo dyes like tartrazine at doses of 50mg+. A systematic review explicitly concludes that azo dye food colourants cause

disturbed behaviour in children, and neurological toxicity mechanisms are well-documented.

- Reasons to Refute: The behavioral effects are not universal—only children identified as sensitive/reactors demonstrate responses—while average dietary intakes in the general population fall below the threshold doses that produce detectable effects. Effects appear to be modulated by individual susceptibility, genetic factors, and co-exposure to other additives rather than being a direct universal causation.

- Consensus & Context Check: The claim's universal language ("cause") does not fully match the evidence, which shows a conditional association restricted to responsive children populations and at challenging doses above typical consumption levels. The relationship is best characterized as a risk factor for a subset of children rather than a universal cause for all cases of ADHD/hyperactivity.

- Final Justification: The accumulated evidence from meta-analyses and rigorous double-blind trials supports that azo food dyes can cause behavioral changes and hyperactivity-like symptoms in susceptible children, particularly at higher doses, but this is not a universal causal effect applicable to all children.

Conclusion
<answer>SUPPORTED</answer>

F Broader Impacts

The primary objective of BioCheck Agent is to mitigate the spread of health misinformation by verifying dubious claims. While our agent demonstrates promising capabilities, its accuracy in discerning all forms of misinformation is still not perfect. Therefore, we expect to use this agent as an assistive reference tool rather than a definitive authority. By providing users with comprehensive fact-checking reports, it is intended to aid human judgment. Over-reliance on our model for critical medical or health-related decision-making poses potential risks, and we strongly encourage users to consult professional medical advice.