DRPG (DECOMPOSE, RETRIEVE, PLAN, GENERATE): AN AGENTIC FRAMEWORK FOR ACADEMIC REBUTTAL

Peixuan Han **Yingjie Yu** **Jingjun Xu** **Jiaxuan You**
University of Illinois Urbana-Champaign
ph16@illinois.edu, jiaxuan@illinois.edu

ABSTRACT

Despite the growing adoption of large language models (LLMs) in scientific research workflows, automated support for academic rebuttal, a crucial step in academic communication and peer review, remains largely underexplored. Existing approaches typically rely on off-the-shelf LLMs or simple pipelines, which struggle with long-context understanding and often fail to produce targeted and persuasive responses. In this paper, we propose **DRPG**, an agentic framework for automatic academic rebuttal generation that operates through four steps: **D**ecompose reviews into atomic concerns, **R**etrieve relevant evidence from the paper, **P**lan rebuttal strategies, and **G**enerate responses accordingly. Notably, the Planner in DRPG reaches over 98% accuracy in identifying the most feasible rebuttal direction. Experiments on data from top-tier conferences demonstrate that DRPG significantly outperforms existing rebuttal pipelines and achieves performance beyond the average human level using only an 8B model. Our analysis further demonstrates the effectiveness of the planner design and its value in providing multi-perspective and explainable suggestions. We also showed that DRPG works well in a more complex multi-round setting. These results highlight the effectiveness of DRPG and its potential to provide high-quality rebuttal content and support the scaling of academic discussions. We will release our code through GitHub.

1 INTRODUCTION

With the rapid advancement of large language models (LLMs), AI agents have become increasingly integrated into the human research workflow. In particular, they have begun to assist researchers across multiple stages of scientific discovery, including idea generation (Lu et al., 2024a), paper writing (Aydin et al., 2025), and peer review (Chang et al., 2025; Yu et al., 2024). Despite these advances, **AI for academic rebuttal**—the process in which authors and reviewers exchange feedback on a paper—remains largely underexplored.

As a critical stage of the research lifecycle, rebuttal plays an essential role in ensuring fair and objective evaluation of submissions. Moreover, as the research community continues to expand, particularly in fast-growing fields such as computer science, preparing thoughtful rebuttals has become increasingly time-consuming for conscientious authors. For instance, major conferences such as NeurIPS and ICLR received over 25,000 submissions in 2025, creating an urgent need for more efficient mechanisms to facilitate communication between authors and reviewers. Consequently, automated or assistive rebuttal agents can potentially reduce researchers' workload substantially, allowing them to focus more on innovative research.

Despite its potential benefits, automating academic rebuttal with LLMs is a challenging task. Rebuttal represents a unique adversarial, multi-agent scenario that requires diverse skills, including precise comprehension, persuasive argumentation, and domain-specific expertise. Prior work typically relies on off-the-shelf LLMs or ad-hoc pipelines to generate responses (Kirtani et al., 2025; Jin et al., 2024b); however, such approaches often yield suboptimal results for two main reasons. Firstly, academic papers are usually lengthy and information-dense. As revealed by the "Lost in the Middle" phenomenon (Liu et al., 2024), LLMs struggle to identify and extract the most relevant evidence from long contexts when responding to specific reviewer concerns. Secondly, effective rebuttals require well-structured and convincing arguments tailored to reviewers' critiques. Since LLMs are not

explicitly trained for persuasion, they tend to produce responses that are overly generic, excessively conciliatory or defensive, failing to directly and convincingly address reviewers' key concerns.

To overcome these limitations, we propose **DRPG**(Decompose, Retrieve, Plan, Generate), **a four-stage agentic framework designed to automatically generate high-quality academic rebuttals**. To mitigate long-context challenges, DRPG first employs a Decomposer to break a review into several "points" where each point is an atomic concern or confusion that needs to be addressed. For each point, a Retriever then selects the most relevant paragraphs from the paper, reducing the input length by over 75% while preserving critical evidence needed for rebuttal. To further enhance argument quality, we introduce a Planner that explicitly formulates rebuttal strategies before response generation. Inspired by planning techniques in structured debates (Wang et al., 2025a; Han et al., 2025), the Planner proposes multiple rebuttal perspectives and identifies the most promising one that is best supported by the paper content. Trained with compelling human rebuttal data, the planner can effectively identify the most supported perspective with an accuracy of over 98%.

Experiments conducted on data from top-tier conferences demonstrate that DRPG can effectively address reviewers' questions, outperforming existing rebuttal pipelines with around 40 points higher Elo score, which implies consistently higher win rates. In addition, DRPG surpasses average human performance using only an 8B model. Further analyses highlight the successful design of the Planner module, and show that it provides a multi-perspective and explainable signal that substantially improves rebuttal quality. Finally, we showed the impressive performance of DRPG on multi-round discussions and conducted a human study to validate our evaluation metrics. Overall, DRPG represents a promising exploration of integrating LLM agents into the peer-review process, with the potential to reshape how authors and reviewers communicate at scale.

2 Related Work

LLM for academic rebuttal. Recent advancements in AI have significantly impacted various stages of scientific discovery, including idea generation, experimentation, and paper writing Luo et al. (2025); Lu et al. (2024a); Schmidgall et al. (2025); Yuan et al. (2025). Among these stages, peer review plays a key role in ensuring the quality and credibility of research papers, and has been receiving increasing attention within the AI research community (Zhuang et al., 2025; Liang et al., 2024; Wei et al., 2025). Existing work in this domain has focused primarily on simulating the review process Yu et al. (2024); Bougie & Watanabe (2024) and training more effective reviewers Chang et al. (2025); Kirtani et al. (2025); Jin et al. (2024b). However, the rebuttal phase, which is vital for facilitating communication between authors and reviewers, has received relatively little attention. A few recent studies explore this area by collecting real-world rebuttal datasets (Zhang et al., 2025a; Kennard et al., 2022), using zero-shot LLMs to generate preliminary rebuttals (Kirtani et al., 2025; Jin et al., 2024b), and training rebuttal agents by designing pre-defined templates to address different questions (Purkayastha et al., 2023; Orbach et al., 2019). Based on these studies, we propose a more systematic and effective rebuttal agent in this work.

LLM for debate and persuasion. Similar to debate and persuasion (Rogiers et al., 2024), the objective of the rebuttal is to convince the reviewer to change their opinion. Researchers have utilized human strategies(Wang et al., 2019; Yang et al., 2019) and high-quality dialogues(Singh et al., 2024; Stengel-Eskin et al., 2024; Jin et al., 2024a; Furumai et al., 2024) to equip LLMs with strong persuasion capabilities. Recently, advanced agentic pipelines (Wang et al., 2025a; Hu et al., 2024; Wu et al., 2025) and Reinforcement learning algorithms Cheng & You (2025); Han et al. (2025) on debate has also been proposed. In the long term, we believe academic rebuttal is a promising domain in debate research, since rebuttal is fact-oriented, grounded in high-quality papers, and has substantial public data available.

Planning in high-stakes decision making. In academic rebuttal, the author must form responses thoughtfully, as the rebuttal quality may affect the result of the submission, with few opportunities to make changes. Planning and selecting argument directions are crucial in such high-stakes settings. There are two prevalent ways of pruning ideas. Firstly, researchers simulate the consequences of adopting each idea, a method originating from Monte Carlo search trees (Coulom, 2006; Silver et al., 2016). This method is most commonly used in scenarios involving multi-agent interactions or external environments (Lu et al., 2024b; Shi et al., 2024; Weng et al., 2024; He et al., 2025). Secondly, researchers train selector or verifier networks to figure out the best candidate through supervised

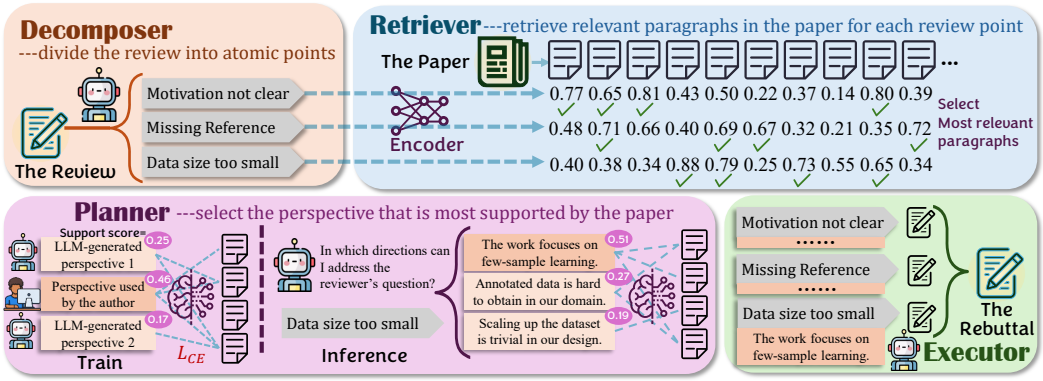

Figure 1: Overview of **DRPG**.

learning (Han et al., 2025; Wang et al., 2025b; Singh & Bali, 2024; Li et al., 2024; White et al., 2021; Lee et al., 2018) or reinforcement learning (Zhang et al., 2025b; Chen et al., 2025) when ground-truth can be obtained at scale. Following this line of work, we train a Planner to select rebuttal perspectives when designing DRPG.

3 METHOD

This section presents an overview of DRPG, an end-to-end framework designed to generate coherent and professional rebuttals based on a full-length, conference-level paper and its reviews. DRPG is composed of four core components—**Decomposer**, **Retriever**, **Planner**, and **Executor**—each of which is introduced in detail in the following subsections. The overall workflow of the framework is illustrated in Figure 1.

3.1 DECOMPOSER

Reviews of academic papers typically involve multiple aspects and viewpoints. To transform such multifaceted and complex feedback into manageable points (refer to Table 4 for an example) for downstream processing, we employ a **Decomposer** implemented using a large language model (LLM). The Decomposer identifies the weaknesses and questions raised by the reviewer, which are key elements that must be addressed in the rebuttal. As a result, the Decomposer divides the review into a set of independent, fine-grained points that will be addressed in the following modules.

3.2 RETRIEVER

Due to the substantial length of academic papers, the performance of rebuttal agents may be adversely affected by the "Lost in the Middle" phenomenon (Liu et al., 2024). To address this issue, we divide the paper into paragraphs and employ a **Retriever** to identify the most relevant paragraphs corresponding to each atomic point generated by the Decomposer. The Retriever is implemented using dense retrieval techniques, where a text encoder is used to embed both review points and paper paragraphs, and cosine similarity is applied to measure their relevance. Only the most relevant paragraphs are passed to the executor, thereby increasing information density while reducing content length for around 75% in practice.

3.3 PLANNER

In academic rebuttal scenarios, identifying an appropriate perspective from which to defend the authors' work is crucial. However, large language models (LLMs) aren't trained to conduct such deliberate planning, leading them to simply state specific details of the paper while overlooking the reviewer's underlying reasoning and value judgments. To address this limitation, we introduce a two-step **Planner** that guides LLMs to explicitly plan how to address review questions, making the communication between authors and reviewers more effective.

In the first step, an **idea proposer** generates several candidate perspectives based on a review point. The proposer is instructed to consider two high-level strategies: **clarification**, which identifies potential misunderstandings in the reviewer's comments, and **justification**, which argues that the reviewer's concern does not invalidate the paper's core contributions. Note that the paper content is intentionally withheld from the idea proposer to encourage creative and diverse perspective generation. As a result, some proposed perspectives may be infeasible or unsupported, which will be filtered out in the subsequent step.

In the second step, the Planner selects the most suitable perspective by evaluating its **supportive score** with respect to the paper's content. Concretely, the Planner is implemented using a text encoder (the same encoder as used in the Retriever) followed by a multi-layer perceptron (MLP). We first obtain vector representations for each candidate rebuttal perspective and each relevant paragraph of the paper using the encoder. These vectors are then concatenated and fed into the MLP to compute a score for each perspective–paragraph pair. The final score of a perspective is obtained by averaging its scores across all relevant paragraphs. Given a perspective "pers" and a set of paragraphs $p_{1..K}$ from the paper (where K denotes the number of relevant paragraphs), the supportive score $s(\text{pers}, p)$ is defined as[1]:

$$s(\text{pers}, p) = \frac{1}{K} \sum_{j=1}^{K} \mathbf{M}\big(\mathbf{E}(\text{pers}) \| \mathbf{E}(p_j)\big), \tag{1}$$

where $\mathbf{E}$ denotes the text encoder and $\mathbf{M}$ represents the MLP module in the Planner.

During training, we select rebuttals that lead to an increase in review scores. For each review point, we construct a candidate set consisting of five "synthetic" perspectives generated by the idea proposer and one "ground-truth" perspective extracted from the actual content. The Planner is optimized using a cross-entropy loss. Let the set of candidate perspectives be $I_{1..N}$, and gt denote the index of the ground-truth, the training loss is then defined as:

$$\mathcal{L}(gt) = -\log \frac{\exp\big(s(I_{gt}, p)\big)}{\sum_{i=1}^{N} \exp\big(s(I_i, p)\big)}. \tag{2}$$

During inference, we design a self-confidence mechanism to ensure the reliability of the selected perspective. A perspective is passed to the Executor only if its confidence score exceeds a predefined threshold T; otherwise, DRPG falls back to the setting without the Planner. The selected perspective and its confidence can be computed through the following formulas:

$$\text{ans} = \text{Argmax}_{i=1}^{N}\, s(I_i, p), \tag{3}$$

$$\text{conf}(\text{ans}) = \frac{\exp\big(s(I_{\text{ans}}, p)\big)}{\sum_{i=1}^{N} \exp\big(s(I_i, p)\big)}. \tag{4}$$

Algorithm 1 The procedure of DRPG[2].

Require: Paper P, Review R, Decomposer $\mathbf{D}$, Encoder $\mathbf{E}$, Retrieved count K, Idea proposer $\mathbf{I}$, Planner $\mathbf{P}$, Threshold T, Executor $\mathbf{X}$

$r[1..N_r] \leftarrow \mathbf{D}(R)$ ▷ DECOMPOSE
$V_p[1..N_p] \leftarrow \mathbf{E}(P[1..N_p])$ ▷ RETRIEVE
$V_r[1..N_r] \leftarrow \mathbf{E}(r[1..N_r])$
for $i = 1$ to N_r **do** ▷ p means relevant paragraphs
  $sim[i, 1..N_p] \leftarrow V_r[i]^\top V_p[j],\ \forall j = 1..N_p$
  $p[i, 1..K] \leftarrow \text{TopK}(sim[i], K)$
end for
for $i = 1$ to N_r **do** ▷ PLAN
  $I[i, 1..N_I] \leftarrow \mathbf{I}(r[i])$ ▷ candidate ideas
  $s[i, 1..N_I] \leftarrow \mathbf{P}(I[i], p[i])$ ▷ supportive scores (Equation (1))
  $id \leftarrow \arg\max_j s[i, j]$
  $\text{conf} \leftarrow \exp(s[i, id]) \,/\, \sum_{j=1}^{N_I} \exp(s[i, j])$
  if conf $\geq T$ **then**
    $I_{select}[i] \leftarrow I[i, id]$
  else
    $I_{select}[i] \leftarrow \epsilon$ ▷ fallback when conf is low
  end if
end for
for $i = 1$ to N_r **do** ▷ GENERATE
  $res[i] \leftarrow \mathbf{X}(r[i], p[i], I_{select}[i])$
end for
RES $\leftarrow \|_{i=1}^{N_r} res[i]$ ▷ concatenate all responses
return RES

3.4 EXECUTOR

The Executor serves as the final stage of the rebuttal pipeline. Given the structured information produced by the preceding modules, the Executor generates a coherent and persuasive rebuttal

[1]The operator $\|$ means vector concatenation.

[2]In Algorithm 1, boldface symbols represent LLMs or networks, and normal symbols represent data variables. Variables starting with N are array sizes induced from LLM outputs.

paragraph for each individual review point. The Executor can be instantiated using either a general-purpose LLM or a model specialized for rebuttal generation.

In summary, DRPG is an agentic workflow designed to automatically generate high-quality rebuttals. By integrating four specialized components, DRPG addresses two key limitations of using a single LLM for rebuttal writing: the difficulty of effectively processing lengthy paper and review texts, and the tendency to produce generic, insufficiently targeted responses. The overall workflow of DRPG is illustrated in Algorithm 1.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Dataset. We conduct experiments on Re^2 (Zhang et al., 2025a), a large-scale dataset consisting of over 17k academic papers and approximately 60k corresponding reviews and rebuttals collected from 45 top-tier computer science conferences over 8 years, including ACL, ICLR, and NeurIPS. Data statistics are reported in Section B.1.

Models. We evaluate our method on four base LLMs spanning different families and sizes: Qwen3-8B (Yang et al., 2025), GPT-oss-20B (Agarwal et al., 2025), Mixtral-8x7B (Jiang et al., 2024), and LLaMa3.3-70B (Dubey et al., 2024). Among all settings, the Retriever is always implemented as BGE-M3 (Chen et al., 2024).

Baselines. We compare DRPG against both human-written rebuttals from the dataset (denoted as `REAL`) and four agentic baselines. As summarized in Table 1, the first three baselines, `Direct`, `Decomp`, and `DRG`, correspond to ablated versions of DRPG with specific components removed. In addition, `Jiu-Jitsu` (Purkayastha et al., 2023) serves as a strong baseline that generates perspectives using predefined templates based on question types, replacing the Planner module in our pipeline.

Metrics. We employ two LLM-based evaluation metrics to assess rebuttal quality[3]. First, we use GPT-4o as a pairwise comparator to rank rebuttals and compute an **Elo score** for each method, following standard practice in open-ended generation tasks (Chiang et al., 2024; Boubdir et al., 2024). Elo scores are estimated by maximum likelihood under a standard Bradley–Terry model, with a base rating of 1000. To validate the reliability of such comparison, we conduct a human study, the results of which are reported in Section 5.5. Second, we use reinforcement learning to train a judge model based on Qwen3-4B to simulate the reviewer's evaluation and scoring process after reading the rebuttal. The judge model gives **judge score** exactly identical with human reviewers on 71% of the test data. Additional details are provided in Section B.3.

Table 1: Components of different baselines.

Setting	Decomposer	Retriever	Planner	Executor
`Direct`	✗	✗	✗	✓
`Decomp`	✓	✗	✗	✓
`DRG`	✓	✓	✗	✓
`Jiu-Jitsu`	✓	✓	✓	✓
DRPG	✓	✓	✓	✓

Training and Inference Details. Detailed training configurations for the Planner are described in Section B.2. For retrieval, we set the number of retrieved paragraphs per review point to $K = 15$. During inference, the Planner applies a confidence threshold of $T = 0.8$. Under this setting, approximately 62% of review points are assigned a valid perspective. The remaining cases typically fall into two categories. The review point is either straightforward and thus doesn't require an explicit perspective, or it's heavily dependent on specific paper content, making it difficult to propose a valid perspective for the Planner.

4.2 MAIN RESULTS

Table 2 clearly shows that **an agentic workflow is crucial for producing high-quality rebuttals**: Directly prompting an LLM to respond to an entire review (`Direct`) consistently yields inferior performance. All agent-based variants achieve substantially higher scores than `Direct`, demonstrating the effectiveness of structured processing.

[3]Traditional n-gram-based metrics such as ROUGE and BLEU are not well-suited for evaluating the reasoning quality and coherence of rebuttals, and are therefore omitted.

Table 2: Performance of rebuttal agents, which shows **DRPG generates the most effective rebuttals across all settings**. The last 5 columns in the pairwise comparison section correspond to the results of comparing the different agent designs using the same base model. Elo scores are calculated within each base model.

Setting	Win Rate Against (%)						Elo Score	Judge Score
	REAL	Direct	Decomp	DRG	Jiu-Jitsu	DRPG		
REAL	-	-	-	-	-	-	-	5.72
Qwen3-8B (Direct)	47.81	-	40.59	37.68	42.48	34.47	936	5.63
Decomp	-	59.61	-	46.77	44.11	43.37	991	5.75
DRG	-	62.32	53.23	-	45.51	41.91	1004	5.75
Jiu-Jitsu	-	57.52	55.89	54.49	-	41.79	1013	5.72
DRPG	60.65	65.53	56.63	58.09	58.21	-	**1054**	**5.78**
GPT-oss-20B (Direct)	40.28	-	27.14	24.39	21.39	22.02	837	5.73
Decomp	-	72.86	-	44.03	47.89	42.57	1012	5.80
DRG	-	75.61	55.97	-	55.53	48.02	1054	5.75
Jiu-Jitsu	-	78.61	52.11	44.47	-	43.21	1029	5.72
DRPG	75.53	77.98	57.43	51.88	56.79	-	**1067**	**5.88**
Mixtral-8x7B (Direct)	49.04	-	17.71	12.06	12.15	13.18	738	5.63
Decomp	-	82.29	-	31.31	28.80	27.82	959	**5.68**
DRG	-	87.94	68.69	-	49.65	44.67	1088	5.66
Jiu-Jitsu	-	87.85	71.20	50.35	-	44.41	1093	5.65
DRPG	51.42	86.62	72.18	55.33	55.59	-	**1119**	**5.68**
LLaMa3.3-70B (Direct)	50.09	-	10.10	18.27	11.49	9.07	725	5.60
Decomp	-	89.90	-	42.78	46.74	40.61	1040	5.67
DRG	-	81.73	57.22	-	50.88	45.91	1065	**5.68**
Jiu-Jitsu	-	88.51	53.26	49.12	-	41.28	1058	5.67
DRPG	65.44	90.93	59.39	54.09	58.72	-	**1109**	**5.68**

Among all methods, **DRPG consistently outperforms the other variants** in pairwise comparisons and achieves the highest post-rebuttal scores in most settings. This finding suggests that each component of DRPG plays an important role in mitigating the inherent limitations of a single-LLM approach. Firstly, the Decomposer and Retriever break down complex reviews into atomic, focused points that can be easily handled, avoiding the shortcomings of excessively long contexts. Secondly, the Planner proposes and identifies an appropriate response direction for each review question. Building on these outputs, the Executor can generate high-quality, tailored responses. We show qualitative cases of DEPG's benefit in Section D.2.

Although Jiu-Jitsu adopts a pipeline structure similar to DRPG, its performance is consistently lower due to the limitations of its Planner. Specifically, the Jiu-Jitsu Planner selects from a fixed set of canonical rebuttal templates, which often results in generic or impractical perspectives[4]. In contrast, DRPG employs a content-aware Planner that selects perspectives based on the paper's content, leading to more specific and persuasive rebuttals.

5 Analysis

5.1 Ablation Study on Planner Design

The Planner is built around an MLP-based scoring function that selects an effective rebuttal perspective by modeling the **supportive relationship** between candidate perspectives and paper paragraphs that are relevant to the review point. In this section, we further analyze its design by comparing it with three alternative variants: 1) no-paper, where the MLP scores each perspective independently without taking any paper content as input; 2) full-paper, where all paragraphs of the paper are used instead of the K relevant paragraphs selected by the Retriever; and

Table 3: Comparison of different planner designs.

Setting	Train loss	Test acc (%)
no-paper	0.5881	61.55
full-paper	0.2393	86.41
encoder	N/A	45.44
Our Planner	**0.0914**	**98.64**

[4]Typical examples include statements such as "We agree some observations have been made in previous work, but there are critical differences" or "We will gladly provide the trained networks on request."

3) `encoder`, a training-free setting which uses vector similarity scores between perspectives and paragraphs as scores.

Table 3 reports the training loss and test accuracy of different planner designs. Our Planner successfully identifies the perspective adopted in successful human rebuttal with an accuracy of **98.64%**, substantially outperforming all alternatives. These results lead to the following observations:

• Incorporating paper content is essential for effective planning. Scoring perspectives in isolation (`no-paper`) results in poor performance.

• Including the Retriever as a preprocessing step significantly improves performance. Compared to `full-paper`, using only relevant paragraphs makes the Planner focus on content directly related to the review point, preventing irrelevant paragraphs from dominating the aggregation in Equation (1).

• Simply relying on encoder similarity without learning (`encoder`) is insufficient. This suggests that the relationship between a rebuttal perspective and its supporting evidence is more nuanced than surface-level relevance, and requires a learned module to capture.

5.2 Analyse on Two Types of Rebuttal Perspectives

As introduced in Section 3.3, the Planner considers two types of rebuttal perspectives: **Clarification** and **Justification**. Clarification aims to correct factual inaccuracies or misunderstandings in the review, whereas Justification seeks to defend the paper's methodology or contributions when the reviewer's comments are factually correct but potentially based on debatable evaluation criteria. Table 4 presents an illustrative example of these two perspective types for the same review point.

Table 4: An example of candidate perspectives generated by the Planner.

Example Atomic Point in Real-world Review: The proposed method performs much worse than HiNet in terms of the extraction accuracy of the secret-in-image hiding.

Perspective by DRPG	**Type**
PSNR may not be the most suitable metric for evaluating the extraction accuracy in the image hiding task.	Justification
Our performance with obfuscating is actually better than HiNet's.	Clarification
Our method works well on secret-in-network hiding, a task much more challenging than image hiding.	Justification

To analyze the effect of perspective choice, we conduct an ablation study that restricts the Planner to a single perspective type. Specifically, instead of allowing the Planner to select the most supported perspective, we force the Executor to respond using only Clarification or only Justification. We denote these two variants as DRPG-C and DRPG-J, respectively. These settings represent two extreme rebuttal strategies: one that focuses exclusively on factual correctness, and the other that emphasizes significance and contribution.

Table 5: Performance of DRPG with restricted perspective types (**C**larification or **J**ustification). We use LLaMa3.3-70B as the base model in this experiment.

Setting	Win Rate Against (%)				**Elo Score**	**Judge Score**	**Ratio of Clarification**
	DRG	DRPG-C	DRPG-J	DRPG			
DRG	-	49.62	64.96	45.91	1018	5.75	-
DRPG-C	50.38	-	65.95	41.49	1014	5.72	100%
DRPG-J	35.04	34.05	-	34.02	915	5.65	0%
DRPG	54.09	58.51	65.98	-	**1051**	**5.78**	66.26%

As shown in Table 5, both variants underperform the full DRPG, and even lag behind the DRG baseline, which does not even include a Planner. This result highlights that relying solely on a single perspective type weakens rebuttal quality. **Effective academic rebuttals require a balanced use of both clarification and justification.** DRPG adapts between these two strategies depending on the review context: it applies clarification when addressing technical misunderstandings, and justification when responding to critiques based on subjective or questionable evaluation standards.

5.3 Interpreting Planner Scores

This section presents an interpretability approach that reveals the explainability advantages of DRPG in the Planner's decision-making. By examining the Planner's scores for each perspective–paragraph pair individually, we can gain insight into why a particular perspective is selected.

Figure 2 illustrates the Planner's scores for the example shown in Table 4. Through supervised training, the Planner learns to capture claim–evidence relationships between candidate perspectives and paper content, rather than relying solely on surface-level semantic similarity. For example, perspective 3 is most strongly supported by paragraph 3, which explicitly states that SinGAN performs better on challenging tasks. Paragraph 4 also receives a relatively high score, as it discusses embedding richer information, which can serve as auxiliary evidence when constructing the rebuttal from the angle of "hard tasks". Such fine-grained score analysis not only improves the transparency of the Planner's decision-making process but also provides a useful structural guide for human authors when composing or refining rebuttals.

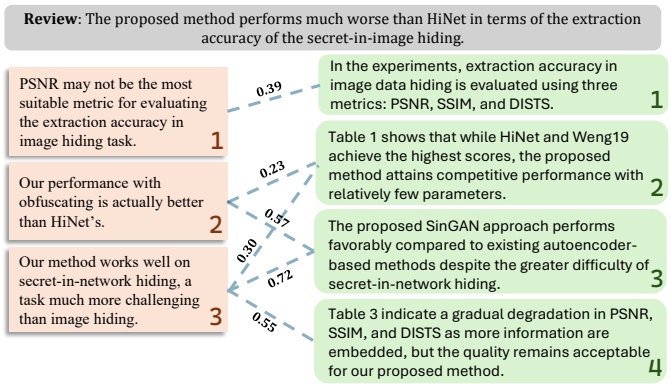

Figure 2: An example to illustrate how the planner evaluates different perspectives. Scores presented are normalized using a sigmoid function, and only scores ≥ 0.2 are displayed.

5.4 Multi-round Discussion with DRPG

Previous experiments focus on single-round rebuttals, however, in real conference review processes, authors and reviewers sometimes engage in multiple rounds of discussion, during which evaluations of the work become more informed and objective. To better reflect this setting, we design an experiment that simulates multi-round reviewer–author interactions and evaluates the rebuttal agent's performance accordingly.

The interaction proceeds in a round-by-round manner, alternating between the DRPG and the judge model trained via reinforcement learning (see Section B.3). In each round, the judge model first summarizes and evaluates the current rebuttal in its chain-of-thought (CoT), and then outputs a final score. We extract this CoT content as a proxy for the reviewer's follow-up feedback and treat it as the "new review" for the next round. The DRPG then generates a subsequent rebuttal in response. Repeating this process enables us to simulate multi-round discussions between reviewers and authors.

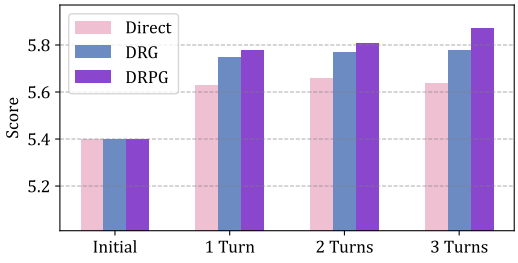

Figure 3: Performance of different rebuttal agents in multi-round discussions. **DRPG addresses follow-up questions better and delivers greater gains compared with the baselines.**

Figure 3 illustrates the performance of different workflows over 3 rounds of discussion. As the number of interaction turns increases, the advantage of DRPG becomes increasingly pronounced. While the judge scores of baseline methods quickly plateau after the first round, DRPG continues to achieve consistent improvements in subsequent rounds. This trend suggests that DRPG is better equipped to incorporate feedback from earlier interactions and to respond effectively to follow-

up questions raised by the reviewer, enabling reviewers to develop a more complete understanding of the paper's technical contributions.

5.5 Human Study on Pairwise Comparison

In this section, we conduct a human study to validate the LLM pairwise comparison in Section 4. Specifically, we sample 20 reviews[5] and their corresponding rebuttals from DRG and DRPG (the base model is LLaMa-3.3-70B in both settings). The reviews are selected so that exactly 10 DRG rebuttals are preferred by gpt-4o, and 10 DRPG rebuttals are preferred. 3 experts in computer science then independently judge which rebuttal is more effective.

In Table 6, we report the agreement among human annotators, as well as the alignment between human judgments and GPT-4o. The results show that the three human experts exhibit highly consistent preferences, and that their evaluations demonstrate a substantial level of agreement with the LLM. Moreover, qualitative analysis indicates that GPT-4o relies on evaluation criteria similar to those used by human reviewers, further supporting its validity as a high-quality proxy for assessing rebuttal quality. Detailed results for 20 reviewed cases are provided in Section D.1, and examples are provided in Section D.2.

Table 6: Human study results.

Setting	Align Rate (%)	κ Score
Among Human Annotators	70	Fleiss's $\kappa = 0.598$
Human Majority v.s. LLM	75	Cohen's $\kappa = 0.500$

6 Conclusion

In this work, we investigate the largely overlooked problem of academic rebuttal automation and present **DRPG**, an agentic framework designed to generate grounded, coherent, and convincing rebuttal responses. DRPG consists of four components: Decomposer, Retriever, Planner, and Executor. By decomposing reviewer feedback, retrieving targeted evidence from long papers, and explicitly planning rebuttal strategies, DRPG addresses key challenges that limit the effectiveness of off-the-shelf LLMs in this setting. Experimental results on top-tier conference data demonstrate that our approach consistently outperforms existing rebuttal methods and achieves strong performance even with a compact model. Beyond empirical gains, our analysis also shows that structured planning provides an interpretable and multi-perspective signal that meaningfully improves rebuttal quality. As the research community continues to grow, we believe agentic systems like DRPG have the potential to help improve the quality and efficiency of scholarly discussions, thereby supporting the continued development of the academic community.

Limitation

This work aims to design an academic rebuttal agent that generates fluent, grounded, and convincing rebuttal arguments. While DRPG shows strong performance, it primarily focuses on clarifying paper content and defending existing contributions, and isn't capable of conducting new experiments. In practice, additional experimental results can sometimes help address reviewers' concerns during rebuttal. An interesting future direction is to integrate DRPG with AI Scientist systems to support experimental supplementation and achieve complete automation of rebuttal process.

Ethical Considerations

This work focuses on automating the academic rebuttal process. While language agents have the potential to significantly assist authors during rebuttal, they also entail inherent risks, particularly those related to hallucinations. Therefore, outputs generated by DRPG (as well as other rebuttal agents) should be carefully reviewed and verified by the authors before submission or release.

[5]Due to the substantial length of each review and rebuttal, we limit the human evaluation to 20 samples to ensure high-quality expert judgments.

REFERENCES

Sandhini Agarwal, Lama Ahmad, Jason Ai, Sam Altman, Andy Applebaum, Edwin Arbus, Rahul K Arora, Yu Bai, Bowen Baker, Haiming Bao, et al. gpt-oss-120b & gpt-oss-20b model card. *arXiv preprint arXiv:2508.10925*, 2025.

Omer Aydin, Enis Karaarslan, Fatih Safa Erenay, and Nebojsa Bacanin. Generative ai in academic writing: A comparison of deepseek, qwen, chatgpt, gemini, llama, mistral, and gemma. *arXiv*, abs/2503.04765, 2025. doi: 10.48550/arXiv.2503.04765. URL https://arxiv.org/abs/2503.04765. [Online; accessed 11-February-2025].

Meriem Boubdir, Edward Kim, Beyza Ermis, Sara Hooker, and Marzieh Fadaee. Elo uncovered: Robustness and best practices in language model evaluation. *Advances in Neural Information Processing Systems*, 37:106135–106161, 2024.

Nicolas Bougie and Narimasa Watanabe. Generative adversarial reviews: When llms become the critic. *arXiv preprint arXiv:2412.10415*, 2024.

Yuan Chang, Ziyue Li, Hengyuan Zhang, Yuanbo Kong, Yanru Wu, Hayden Kwok-Hay So, Zhijiang Guo, Liya Zhu, and Ngai Wong. Treereview: A dynamic tree of questions framework for deep and efficient llm-based scientific peer review. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 15662–15693, 2025.

Jianlv Chen, Shitao Xiao, Peitian Zhang, Kun Luo, Defu Lian, and Zheng Liu. Bge m3-embedding: Multi-lingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation. *arXiv preprint arXiv:2402.03216*, 2024.

Xiusi Chen, Gaotang Li, Ziqi Wang, Bowen Jin, Cheng Qian, Yu Wang, Hongru Wang, Yu Zhang, Denghui Zhang, Tong Zhang, et al. Rm-r1: Reward modeling as reasoning. *arXiv preprint arXiv:2505.02387*, 2025.

Zirui Cheng and Jiaxuan You. Towards strategic persuasion with language models. *arXiv preprint arXiv:2509.22989*, 2025.

Wei-Lin Chiang, Lianmin Zheng, Ying Sheng, Anastasios Nikolas Angelopoulos, Tianle Li, Dacheng Li, Banghua Zhu, Hao Zhang, Michael Jordan, Joseph E Gonzalez, et al. Chatbot arena: An open platform for evaluating llms by human preference. In *Forty-first International Conference on Machine Learning*, 2024.

Rémi Coulom. Efficient selectivity and backup operators in monte-carlo tree search. In *International conference on computers and games*, pp. 72–83. Springer, 2006.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. The llama 3 herd of models. *arXiv e-prints*, pp. arXiv–2407, 2024.

Kazuaki Furumai, Roberto Legaspi, Julio Vizcarra, Yudai Yamazaki, Yasutaka Nishimura, Sina J Semnani, Kazushi Ikeda, Weiyan Shi, and Monica S Lam. Zero-shot persuasive chatbots with llm-generated strategies and information retrieval. *arXiv preprint arXiv:2407.03585*, 2024.

Peixuan Han, Zijia Liu, and Jiaxuan You. Tomap: Training opponent-aware llm persuaders with theory of mind. *arXiv preprint arXiv:2505.22961*, 2025.

Haorui He, Yupeng Li, Dacheng Wen, Yang Chen, Reynold Cheng, Donglong Chen, and Francis Lau. Debating truth: Debate-driven claim verification with multiple large language model agents. *arXiv preprint arXiv:2507.19090*, 2025.

Zhe Hu, Hou Pong Chan, Jing Li, and Yu Yin. Debate-to-write: A persona-driven multi-agent framework for diverse argument generation. In *International Conference on Computational Linguistics*, 2024. URL https://api.semanticscholar.org/CorpusId:270845910.

Albert Q Jiang, Alexandre Sablayrolles, Antoine Roux, Arthur Mensch, Blanche Savary, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Emma Bou Hanna, Florian Bressand, et al. Mixtral of experts. *arXiv preprint arXiv:2401.04088*, 2024.

Chuhao Jin, Kening Ren, Lingzhen Kong, Xiting Wang, Ruihua Song, and Huan Chen. Persuading across diverse domains: a dataset and persuasion large language model. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, pp. 1678–1706, 2024a.

Yiqiao Jin, Qinlin Zhao, Yiyang Wang, Hao Chen, Kaijie Zhu, Yijia Xiao, and Jindong Wang. Agentreview: Exploring peer review dynamics with llm agents. *arXiv preprint arXiv:2406.12708*, 2024b.

Neha Nayak Kennard, Tim O'Gorman, Rajarshi Das, Akshay Sharma, Chhandak Bagchi, Matthew Clinton, Pranay Kumar Yelugam, Hamed Zamani, and Andrew McCallum. Disapere: A dataset for discourse structure in peer review discussions. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1234–1249, 2022.

Chavvi Kirtani, Madhav Krishan Garg, Tejash Prasad, Tanmay Singhal, Murari Mandal, and Dhruv Kumar. Revieweval: An evaluation framework for ai-generated reviews. *arXiv e-prints*, pp. arXiv–2502, 2025.

Sang-Woo Lee, Yu-Jung Heo, and Byoung-Tak Zhang. Answerer in questioner's mind: Information theoretic approach to goal-oriented visual dialog. *Advances in neural information processing systems*, 31, 2018.

Zixuan Li, Lizi Liao, and Tat-Seng Chua. Learning to ask critical questions for assisting product search. *arXiv preprint arXiv:2403.02754*, 2024.

Weixin Liang, Yuhui Zhang, Hancheng Cao, Binglu Wang, Daisy Yi Ding, Xinyu Yang, Kailas Vodrahalli, Siyu He, Daniel Scott Smith, Yian Yin, et al. Can large language models provide useful feedback on research papers? a large-scale empirical analysis. *NEJM AI*, 1(8):AIoa2400196, 2024.

Nelson F Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. Lost in the middle: How language models use long contexts. *Transactions of the Association for Computational Linguistics*, 12:157–173, 2024.

Chris Lu, Cong Lu, Robert Tjarko Lange, Jakob Foerster, Jeff Clune, and David Ha. The ai scientist: Towards fully automated open-ended scientific discovery, 2024a. URL https://arxiv.org/abs/2408.06292.

Li-Chun Lu, Shou-Jen Chen, Tsung-Min Pai, Chan-Hung Yu, Hung-yi Lee, and Shao-Hua Sun. Llm discussion: Enhancing the creativity of large language models via discussion framework and role-play. *arXiv preprint arXiv:2405.06373*, 2024b.

Ziming Luo, Zonglin Yang, Zexin Xu, Wei Yang, and Xinya Du. Llm4sr: A survey on large language models for scientific research. *arXiv preprint arXiv:2501.04306*, 2025.

Matan Orbach, Yonatan Bilu, Ariel Gera, Yoav Kantor, Lena Dankin, Tamar Lavee, Lili Kotlerman, Shachar Mirkin, Michal Jacovi, Ranit Aharonov, et al. A dataset of general-purpose rebuttal. *arXiv preprint arXiv:1909.00393*, 2019.

Sukannya Purkayastha, Anne Lauscher, and Iryna Gurevych. Exploring jiu-jitsu argumentation for writing peer review rebuttals. *arXiv preprint arXiv:2311.03998*, 2023.

Alexander Rogiers, Sander Noels, Maarten Buyl, and Tijl De Bie. Persuasion with large language models: a survey. *arXiv preprint arXiv:2411.06837*, 2024.

Samuel Schmidgall, Yusheng Su, Ze Wang, Ximeng Sun, Jialian Wu, Xiaodong Yu, Jiang Liu, Zicheng Liu, and Emad Barsoum. Agent laboratory: Using llm agents as research assistants. *arXiv preprint arXiv:2501.04227*, 2025.

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.

Li Shi, Houjiang Liu, Yian Wong, Utkarsh Mujumdar, Dan Zhang, Jacek Gwizdka, and Matthew Lease. Argumentative experience: Reducing confirmation bias on controversial issues through llm-generated multi-persona debates. *arXiv preprint arXiv:2412.04629*, 2024.

David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484–489, 2016.

Gaurav Singh and Kavitesh Kumar Bali. Enhancing decision-making in optimization through llm-assisted inference: A neural networks perspective. In *2024 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–7. IEEE, 2024.

Somesh Singh, Yaman K Singla, Harini SI, and Balaji Krishnamurthy. Measuring and improving persuasiveness of large language models. *arXiv preprint arXiv:2410.02653*, 2024.

Elias Stengel-Eskin, Peter Hase, and Mohit Bansal. Teaching models to balance resisting and accepting persuasion. *arXiv preprint arXiv:2410.14596*, 2024.

Danqing Wang, Zhuorui Ye, Xinran Zhao, Fei Fang, and Lei Li. Strategic planning and rationalizing on trees make llms better debaters. *arXiv preprint arXiv:2505.14886*, 2025a.

Fuyu Wang, Jiangtong Li, Kun Zhu, and Changjun Jiang. Inspiredebate: Multi-dimensional subjective-objective evaluation-guided reasoning and optimization for debating. *arXiv preprint arXiv:2506.18102*, 2025b.

Xuewei Wang, Weiyan Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu. Persuasion for good: Towards a personalized persuasive dialogue system for social good. *arXiv preprint arXiv:1906.06725*, 2019.

Qiyao Wei, Samuel Holt, Jing Yang, Markus Wulfmeier, and Mihaela van der Schaar. The ai imperative: Scaling high-quality peer review in machine learning. *arXiv preprint arXiv:2506.08134*, 2025.

Yixuan Weng, Minjun Zhu, Guangsheng Bao, Hongbo Zhang, Jindong Wang, Yue Zhang, and Linyi Yang. Cycleresearcher: Improving automated research via automated review. *arXiv preprint arXiv:2411.00816*, 2024.

Julia White, Gabriel Poesia, Robert Hawkins, Dorsa Sadigh, and Noah Goodman. Open-domain clarification question generation without question examples. *arXiv preprint arXiv:2110.09779*, 2021.

Jibang Wu, Chenghao Yang, Simon Mahns, Yi Wu, Chaoqi Wang, Hao Zhu, Fei Fang, and Haifeng Xu. Ai realtor: Towards grounded persuasive language generation for automated copywriting. In *unknown*, 2025. URL https://api.semanticscholar.org/CorpusId:276575033.

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

Diyi Yang, Jiaao Chen, Zichao Yang, Dan Jurafsky, and Eduard Hovy. Let's make your request more persuasive: Modeling persuasive strategies via semi-supervised neural nets on crowdfunding platforms. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 3620–3630, 2019.

Haofei Yu, Zhaochen Hong, Zirui Cheng, Kunlun Zhu, Keyang Xuan, Jinwei Yao, Tao Feng, and Jiaxuan You. Researchtown: Simulator of human research community. *arXiv preprint arXiv:2412.17767*, 2024.

Jiakang Yuan, Xiangchao Yan, Botian Shi, Tao Chen, Wanli Ouyang, Bo Zhang, Lei Bai, Yu Qiao, and Bowen Zhou. Dolphin: Closed-loop open-ended auto-research through thinking, practice, and feedback. *arXiv preprint arXiv:2501.03916*, 2025.

Daoze Zhang, Zhijian Bao, Sihang Du, Zhiyi Zhao, Kuangling Zhang, Dezheng Bao, and Yang Yang. Re2: A consistency-ensured dataset for full-stage peer review and multi-turn rebuttal discussions. *arXiv preprint arXiv:2505.07920*, 2025a.

Naifan Zhang, Ruihan Sun, Ruixi Su, Shiqi Ma, Shiya Zhang, Xianna Weng, Xiaofan Zhang, Yuhan Zhan, Yuyang Xu, Zhaohan Chen, et al. Echo-n1: Affective rl frontier. *arXiv preprint arXiv:2512.00344*, 2025b.

Zhenzhen Zhuang, Jiandong Chen, Hongfeng Xu, Yuwen Jiang, and Jialiang Lin. Large language models for automated scholarly paper review: A survey. *Information Fusion*, pp. 103332, 2025.

A PROMPTS

This section shows LLM prompts in the paper.

Prompts used in the DRPG pipeline are shown in Figures 5 to 8. Note that Figure 7 is used in the Executor in `Direct` setting, and Figure 7 is used in other settings. After the Executor responds to each individual point, they're merged together to form a complete rebuttal. Refer to Section D.2 for illustrative examples.

Prompts used in the evaluation are shown in Figures 9 and 10. To avoid the position bias, the order of the two rebuttals is randomly swapped during the comparison.

B DATA AND TRAINING DETAILS

B.1 DATA STATISTICS

Table 7: Dataset statistics.

(a) Size of the dataset.

Dataset	Papers	Reviews	Rebuttals
# Train	17,814	62,211	34,024
# Eval	600	2,097	1,092

(b) Score changes after rebuttal.

Dataset	Score Increased	Score Decreased	Score Unchanged
# Train	23,642	207	38,362
# Eval	745	5	1,347

We summarize the data statistics in Table 7a. Note that only a subset of official reviews are responded by the authors. Each official review includes a score reflecting paper quality, and the scores may be changed during the rebuttal process. However, some initial scores are missing in the dataset. To reconstruct the full rebuttal trajectory, we utilize GPT-oss-120B (Agarwal et al., 2025) with the prompt in Figure 11 to predict initial scores from the rebuttal text and the final scores.

Summary statistics of the rebuttal scores are shown in Table 7b. The resulting score distributions meet our expectations, and we further validated the predictions through human analysis of several randomly sampled examples.

B.2 PLANNER TRAINING

As described in Section 3.3, we construct the Planner's training data by combining five candidate perspectives generated by the idea proposer with one ground-truth perspective extracted from human-authored rebuttals. Using this strategy, we collect 50,000 review points to form the training set and an additional 5,000 review points for evaluation.

The Planner is implemented as an MLP with three hidden layers of sizes 2048, 1024, and 512, respectively. The input layer takes the concatenated vectors of a perspective–paragraph pair with a size of 2048, and the output layer produces a scalar supportive score. Training is conducted for 3 epochs using a batch size of 32 and a learning rate of 5×10^{-5}. Due to computational constraints, we freeze the parameters of the BGE-M3 encoder during training and update only the MLP module.

B.3 JUDGE MODEL TRAINING

The judge model is trained with Group Relative Policy Optimization (GRPO) Shao et al. (2024), and the base model is Qwen3-4B. As shown in Section A, the judge model is expected to take careful thinking before generating a final score. The reward is calculated as $r = 0.25^{|s-s_g|}$, where s is the judge model's answer, and s_g is the actual score. This means the model will receive a full reward when the predicted score matches exactly with the ground truth, and the reward decreases exponentially as the prediction gap increases. Figure 4 shows the hyperparameters during training.

Figure 4: GRPO training configuration for the judge model.

Hyperparameter	Value
Learning rate	1×10^{-6}
Rollout temperature	1.0
KL Coefficient	0.001
Train batch size	64
Mini batch size	32
Micro batch size	16
Training steps	200
Number of Rollouts	4

C THE JIU-JITSU BASELINE

We include `Jiu-Jitsu` (Purkayastha et al., 2023) as a planning baseline. Different from our Planner, the Jiu-Jitsu Planner generates perspectives through selecting a canonical rebuttal template.

Given an atomic review point r_i, `Jiu-Jitsu` maps the concern to a canonical rebuttal template via a retrieve-and-rank procedure. First, it converts r_i into a generated natural-language description d_i of the reviewer concern using a fine-tuned sequence-to-sequence language model, which abstracts away surface wording differences and summarizes the underlying issue. Second, it assigns d_i to the closest attitude root–theme cluster z_i (where the root corresponds to a reviewing aspect such as *Clarity* and the theme corresponds to a target paper section such as *Experiments*), which is associated with a cluster description d_{z_i}. Third, it uses the rebuttal action label a_i provided in the `Jiu-Jitsu` resources for each (d_i, z_i) to retrieve the relevant candidate rebuttals for ranking. Finally, `Jiu-Jitsu` scores each candidate $r \in R(z_i)$ using the cluster description d_{z_i} and action a_i, then selects the top-ranked candidate as the canonical rebuttal, which serves as the rebuttal perspective in our pipeline. Figure 12 shows an example of the procedure of *Jiu-Jitsu* baseline and the selected canonical rebuttal.

D ADDITIONAL EXPERIMENTS

D.1 DETAILS FOR HUMAN STUDY

Our human study is conducted through a web interface for human annotators to interact with, and the results are automatically collected. Figure 13 shows a screenshot of the webpage. We also provide the results for all 20 reviews in Table 8. In the human study, we shuffled the order of 20 reviews for each human annotator.

D.2 CASE STUDY

Figures 14 and 15 shows two examples comparing `Decomp`, `DRG` and DRPG. We also provide the comments from GPT-4o during the pairwise comparison, from which we can observe the value of the Retriever and Planner in providing more concrete rebuttals.

You are an experienced researcher in computer science. You have written a conference paper in the field of computer science or AI and received a review. You need to analyse the reviewer's comments. Specifically, identify and list all the weakness points or confusions raised by the reviewer.
- You may omit minor issues such as typos, but major comments should all be mentioned.
- Preferably, extract sentences or words directly from the review. Do not oversimplify the comments.

Below is an example of the expected output format:
[
"The paper introduced two modules, but lacks ablation study which includes only one of them.",
"What does the author mean by PPO? Further explain will be helpful.",
"The experimental results are only shown on 1 newly created environment."
]

Figure 5: System Prompt for the Decomposer

You are an experienced researcher in computer science. You have received a review on a research paper. Your task is to propose up to 5 perspectives to address this point in the rebuttal.
- The perspective should either show the reviewer's point wrong, or show that the work is valuable even though the review is correct. Specifically, You MUST consider the following two types of perspectives:
  - Clarification: The reviewer may have factual errors or misunderstood in the paper. For example, they may say something is missing when it's actually present in the paper, or say the methodology is wrong because of a misunderstanding.
  - Justification: Defend your choices and explain why the comment doesn't undermine your paper. For example, they may require an experiment which is unfeasible or unnecessary, or require empirical results for a theoretical paper.
- DO NOT propose suggestions or promises for future revision or future work.
- DO NOT mention specific locations in the paper since you won't be able to access it (e.g. "in section 3.2").

Below is an example of the expected output format:
Input: "The paper introduced two modules, but lacks ablation study which includes only one of them."
Output:
[
"Clarification: we have actually included such experiment in the paper.",
"Clarification: the two modules are dependent on each other and therefore cannot be separated.",
"Justification: the ablation study is not necessary as each module has been individually validated in prior work."
]

Figure 6: System Prompt for the Perspective Generator

You are an experienced researcher in computer science. You have written a conference paper in the field of computer science or AI and received a review. You need to write a rebuttal to address the reviewer's comments and convince them to increase their score.

Guidelines:
1. Be polite, concise, and professional. Make sure all responses are factual, respectful, and persuasive.
2. Address each comment point-by-point. It's recommended to format the main part of the rebuttal as: "Question: ...Response: ...". For each point:
3. For each point, you should respond with clear reasoning, and evidence from the original paper, and your professional knowledge.
   - If the comment has misunderstood the paper or missed some content, clarify the point. If not, defend your choices and explain why this comment doesn't undermine your paper.
   - DO NOT propose suggestions or promises for future revision or future work.
4. Be confident with your paper. Try your best to explain and validate your work, and rebut the concerns raised by the reviewer.
5. Your rebuttal should be concise and no more than 1000 words. You should directly generate a passage without additional comments or thoughts.

Figure 7: System Prompt for the Executor for a Whole Review

You are an experienced researcher in computer science. You have written a conference paper in the field of computer science or AI and received a review. You need to write a rebuttal to address the reviewer's comment and convince them to increase their score.

Guidelines:
1. Make sure your response is factual, respectful, and persuasive.
2. You should respond with clear reasoning, and evidence from the original paper, and your professional knowledge.
   - If the comment has misunderstood the paper or missed some content, clarify the point. If not, defend your choices and explain why this comment doesn't undermine your paper.
   - DO NOT propose suggestions or promises for future revision or future work.
3. Be confident with your paper. Try your best to explain and validate you work, and rebute the concerns raised by the reviewer.
4. Your rebuttal should be concise and no more than 200 words. You should directly generate a paragraph without additional comments or thoughts.

Figure 8: System Prompt for the Executor for Individual Review Points

You are an experienced academic paper reviewer. You will receive a response from the authors addressing your review comments. Your task is to evaluate the response and decide whether to adjust your original score for the paper.
The scoring rubric is from 1 - 10 scale. Certain scores correspond to the following meanings:
- 1: The paper has serious flaws, lacks novelty, or is clearly unsuitable for acceptance.
- 3: The paper has significant weaknesses or insufficient contributions.
- 6: Top 25% of all submissions. The paper is slightly above the acceptance threshold, with generally solid work, but some limitations.
- 8: Top 10% of all submissions. The paper has a good-quality paper with clear contributions and well-supported results.
- 10: Top 5% of all submissions. The paper makes exceptional contributions and is recommended for spotlight or oral presentation.

You should focus on the following criteria when assessing the author's response:
- 1. Does the author's response validates their work with clear arguments and coherent logic?
- 2. Does the author provide sufficient evidence or reasoning to support their claims?
- 3. Is the author's response consistent with the content of the original paper?

In addition, please keep in mind that the goal of the response is to CONVINCE the reviewer about the paper, instead of SUGGESTIONS for future work or ADMITTING weakness.
- DO NOT consider suggestions, promises, or impacts for future work and revisions when evaluating the responses. Focus on this paper alone.
- DO NOT consider tones or emotional appeals, as long as the content is professional. Focus on the logic and reasoning.

Then, you should decide whether to change your score based on the author's response.
- You should be confident with your original review in most cases. You may increase your score only if the author provides sufficient reasoning that addresses your comments.
- Do not increase your score based on minor corrections (e.g. typos) or promises on future revisions.
- If the original score is low, you should be more lenient in increasing the score. If the original score is high, you should hold a higher standard.
- In most cases, the score change will be small. Large changes, like 2 points, should be rare and well-justified.

As a conclusion, output "My final score is X" where X is your final score (an integer between 1 and 10).

Figure 9: System Prompt for the Rebuttal Judge

You are an experienced academic paper reviewer. You will receive a review of an academic paper in computer science, and two responses from the authors. Your task is to evaluate the responses and decide which response is better.
The response may address the reviewer's several comments. You should compare the responses to each comment individually. When comparing the responses, you can refer to the following criteria:
- 1. Does the author's response validate their work with clear arguments and coherent logic?
- 2. Does the author provide sufficient evidence or reasoning to support their claims?
- 3. Is the author's response consistent with the content of the original paper?

In addition, please keep in mind that the author isn't allowed to revise the paper afterwards. That is, the goal of the response is to CONVINCE the reviewer about the paper, instead of SUGGESTIONS for future work or ADMITTING weakness.
- DO NOT consider suggestions, promises, or impacts for future work and revisions when evaluating the responses. Focus on this paper alone.
- DO NOT consider tones or emotional appeals, as long as the content is professional. Focus on the logic and reasoning.

Please give concrete evidence while being concise. DO NOT repeat or simply summarize the responses' content or similarities; focus on their differences and YOUR ANALYSIS. Output "I think response X (1 or 2) is better" or "I think two responses are similar in quality" at the end of your answer.

Figure 10: System Prompt for Comparing Two Rebuttals

System:
You will be given a reviewer–author discussion text and the paper's final score. Based only on the discussion text and the final score, predict the paper's initial (pre-discussion) review score.
Strictly output a single valid JSON object and nothing else. The JSON must contain only these two fields:
{
"opinion": "In 2–6 concise sentences, explain your analysis and list the main evidence/signals that support your prediction. If information is insufficient or contradictory, note that uncertainty here",
"initial_score": "initial score as an integer from 1 to 10"
}
Hard rules (must follow):
1. **Output only the JSON object** — no extra commentary, no code fences outside the JSON, no explanations.
2. 'initial_score' must be an integer between 1 and 10.
3. 'opinion' must mention 2–4 clear signals or events from the discussion and explain how they affect the score estimate.
4. Do not invent facts outside the provided discussion text. Avoid hallucination.
5. If the discussion is ambiguous or contradictory, state that in 'opinion' and then give the most likely integer prediction.
Usually, the reviewer is confident with their review, which means they only raise or decrease scores where there is sufficient evidence.

User:
Discussion text: {discussion_text }
Final score: {final_score } / 10

Figure 11: System Prompt to Predict the Initial Review Score

Review point r_i: "The experimental setup is unclear. Please specify the hyperparameters and training details."
Generated description d_i: "Missing experimental details and unclear description of training settings."
Predicted root–theme cluster z_i: Root=*Clarity*, Theme=*Experiments*
Action label a_i: `rebuttal_concede-criticism`
Selected canonical rebuttal template c_i (perspective): "We apologize for the unclear description of the experimental settings. We will revise the paper to include the missing hyperparameters and training details."

Figure 12: An example of `Jiu-Jitsu` procedure to generate rebuttal perspective for a review point.

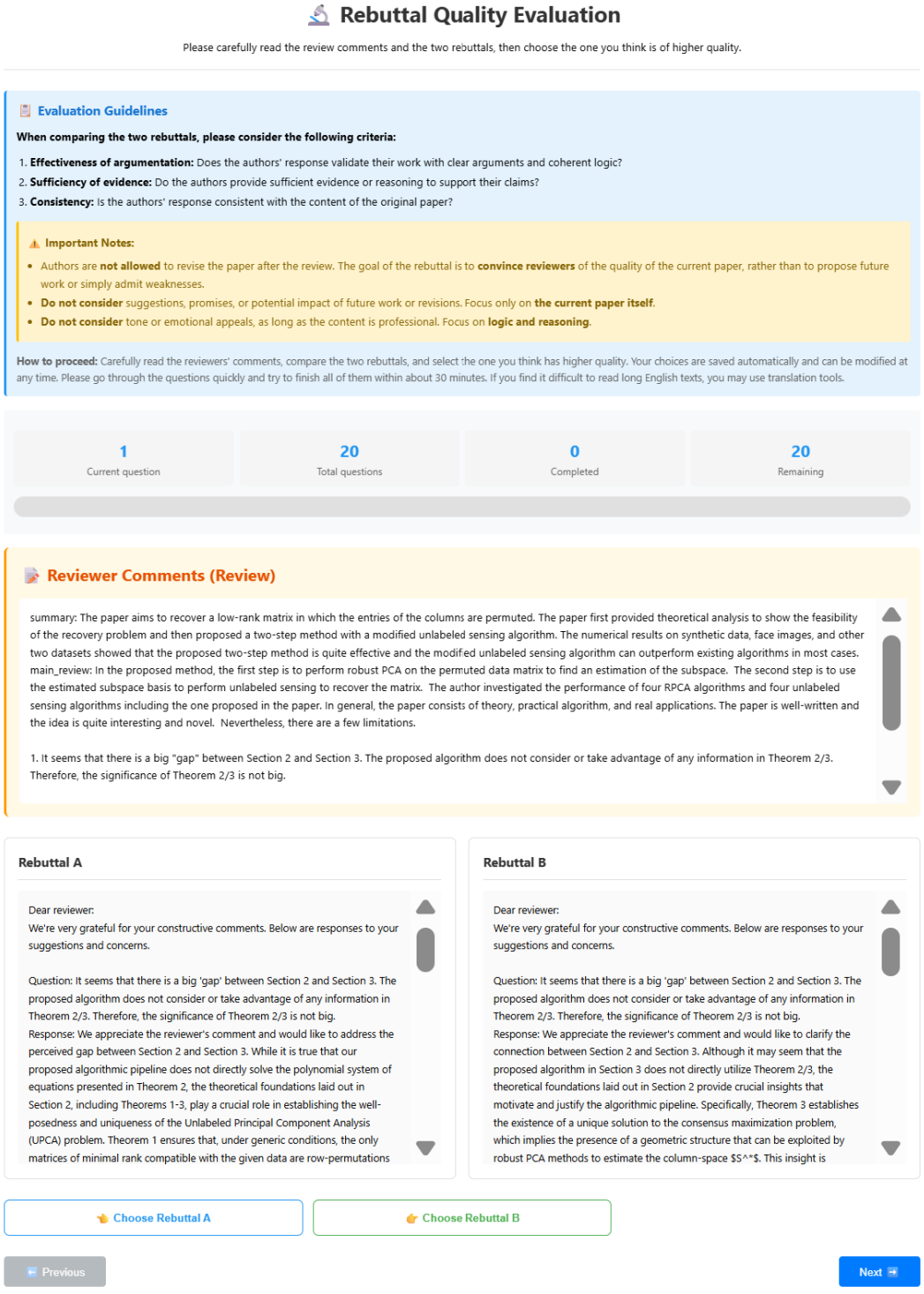

Figure 13: Illustration of the webpage used for human annotation.

Table 8: Details for human study. Rebuttal 1 refers to the rebuttal generated by the DRG pipeline, and Rebuttal 2 refers to the rebuttal generated by DRPG.

Number	Paper Source	Human Annotator 1	Human Annotator 2	Human Annotator 3	GPT-4o
1	EMNLP2020	Rebuttal 1	Rebuttal 1	Rebuttal 1	Rebuttal 1
2	ICML2020	Rebuttal 2	Rebuttal 1	Rebuttal 1	Rebuttal 1
3	KDD2023	Rebuttal 1	Rebuttal 1	Rebuttal 1	Rebuttal 1
4	ICML2020	Rebuttal 1	Rebuttal 2	Rebuttal 2	Rebuttal 1
5	CoRL2023	Rebuttal 1	Rebuttal 1	Rebuttal 1	Rebuttal 1
6	AAAI2022	Rebuttal 1	Rebuttal 1	Rebuttal 1	Rebuttal 1
7	ICLR2023	Rebuttal 2	Rebuttal 2	Rebuttal 2	Rebuttal 1
8	ACL2020	Rebuttal 1	Rebuttal 1	Rebuttal 1	Rebuttal 1
9	ICLR2022	Rebuttal 2	Rebuttal 1	Rebuttal 1	Rebuttal 1
10	NeurIPS2022	Rebuttal 1	Rebuttal 1	Rebuttal 1	Rebuttal 1
11	ICLR2022	Rebuttal 2	Rebuttal 2	Rebuttal 2	Rebuttal 2
12	ECCV2022	Rebuttal 1	Rebuttal 1	Rebuttal 1	Rebuttal 2
13	NeurIPS2022	Rebuttal 2	Rebuttal 2	Rebuttal 2	Rebuttal 2
14	ACL2020	Rebuttal 1	Rebuttal 2	Rebuttal 1	Rebuttal 2
15	ICLR2022	Rebuttal 2	Rebuttal 2	Rebuttal 2	Rebuttal 2
16	AAAI2022	Rebuttal 2	Rebuttal 2	Rebuttal 1	Rebuttal 2
17	SIGKDD2022	Rebuttal 2	Rebuttal 2	Rebuttal 2	Rebuttal 2
18	CoRL2023	Rebuttal 2	Rebuttal 2	Rebuttal 2	Rebuttal 2
19	ICLR2023	Rebuttal 1	Rebuttal 2	Rebuttal 1	Rebuttal 2
20	ICML2021	Rebuttal 2	Rebuttal 2	Rebuttal 2	Rebuttal 2

Review Point: Adding factual data in fact may not always lead to clear improvements, while in general most improvement appears to be driven by the heuristic ATE weighting.
Decomp Response: We appreciate the reviewer's comment and would like to clarify that the addition of factual data does indeed contribute to the overall improvement, even if its individual contribution may seem marginal in some cases. The comparison between the purple and green bars in Figure 2 and Figure 3, as well as Table 2, shows that the DT policy trained on both factual and counterfactual simulations (purple bar) outperforms the DT policy trained on counterfactual simulations only (green bar). This suggests that the factual data provides some benefit, even if the majority of the improvement comes from the counterfactual simulations and the ATE weighting scheme. Furthermore, the factual data helps to anchor the DT policy in the source environment, making the transfer to the target environment more effective. Therefore, while the ATE weighting scheme is a key component of our approach, the inclusion of factual data is also an important factor in achieving the best results.
DRPG Response: We appreciate the reviewer's observation that the addition of factual data may not always lead to clear visual improvements, as seen in the comparison between the purple and green bars or in Table 2. However, we would like to clarify that the inclusion of factual data is intended to provide a more comprehensive understanding of the results, rather than solely driving improvements. The ATE weighting scheme is indeed a key factor in enhancing the generalization ability of the DT agents, as demonstrated in Figures 2 and 3. The factual data serves as a foundation, allowing the counterfactual simulations to build upon and improve the overall performance. While the improvements may not always be drastic, the combination of factual and counterfactual data with ATE weighting consistently yields better results, as evident in the overall trend of our experimental findings. This suggests that the addition of factual data, although sometimes subtle in its impact, contributes to a more robust and effective training scheme for DT agents.
Comment: Both responses provide coherent arguments and reasoning, but Response 2 has a slight edge due to its emphasis on the foundational nature of the work and its detailed explanation on the role of factual data and ATE weighting.

Figure 14: A case study comparing Decomp amd DRPG.

Review Point: The dataset contains only 20 images.
DRG Response: We acknowledge the reviewer's concern regarding the size of the dataset used in our experiments. However, we would like to clarify that the dataset size was intentionally kept small to demonstrate the feasibility and effectiveness of the proposed SinGAN approach in hiding images in deep probabilistic models. With only 20 test image pairs, we were able to achieve promising results in terms of extraction accuracy and model security, as shown in Tables 1 and 2. Furthermore, the use of a small dataset allowed us to focus on the quality and diversity of the generated cover images, as well as the weight distribution similarity between the original and stego SinGANs. While a larger dataset may be desirable for more comprehensive evaluation, our current results provide a solid foundation for future work and demonstrate the potential of the proposed framework.
DRPG Response: We appreciate the reviewer's comment regarding the dataset size. Although our experiments involve a dataset of 20 images, we would like to emphasize that this is sufficient for a proof-of-concept demonstration of our probabilistic image hiding framework. The primary goal of our work is to showcase the feasibility and effectiveness of hiding images in deep probabilistic models, particularly using SinGANs. With a smaller dataset, we are able to establish a baseline for the performance of our approach, as evidenced by the extraction accuracy results in Table 1 and the security evaluation in Table 2. Moreover, our method's ability to hide multiple images for different receivers, as shown in Table 3, further demonstrates its potential. While we acknowledge that larger datasets may be necessary for more comprehensive evaluations, our current results provide a solid foundation for future work, which can indeed focus on scaling up to larger datasets and exploring more complex scenarios.
Comment: Both responses argue that the small dataset size is adequate for a proof-of-concept demonstration. However, Response 2 provides a more detailed explanation by referencing specific tables in the paper that demonstrate the method's effectiveness, thus providing more evidence to support their claim.

Figure 15: A case study comparing DRG amd DRPG.