Nova: An Iterative Planning and Search Approach to Enhance Novelty and Diversity of LLM Generated Ideas

Anonymous ACL submission

Abstract

Scientific innovation is pivotal for humanity, and harnessing large language models (LLMs) to generate research ideas could transform discovery. However, existing LLMs often produce simplistic and repetitive suggestions due to their limited capacity acquire external knowledge effectively for innovation. To address this, we introduce an enhanced planning and search methodology designed to boost the creative potential of LLM-based systems. Our approach involves an iterative process to purposely plan the retrieval of external knowledge, progressively enriching the idea generation with broader and deeper insights. Validation through automated and human assessments indicates that our framework substantially elevates the quality of generated ideas, particularly in novelty and diversity. The number of unique novel ideas produced by our framework is 3.4 times higher than without it. Moreover, our method outperforms the current state-of-the-art, generating at least 2.5 times more top-rated ideas based on 170 seed papers in a Swiss Tournament evaluation.

1 Introduction

In recent years, LLMs have demonstrated remarkable progress across various challenging tasks, including solving mathematical problems (Romera-Paredes et al., 2024), proving mathematical theory (Wang et al., 2024a), and generating code to solve analytical or computational tasks (Huang et al., 2024). These progresses have opened up new possibilities to utilize LLMs to accelerate research (Wang et al., 2023a), including generating novel research ideas (Si et al., 2024; Wang et al., 2024b; Baek et al., 2024).

Our work is dedicated to addressing the challenge of employing LLMs to produce high-caliber research ideas, with an emphasis on enhancing their novelty and diversity. Existing studies (Wang et al., 2024b; Si et al., 2024) tackle this challenge by integrating additional knowledge into the idea generation process. Wang et al. (2024b) enrich the process by incorporating co-occurrence entities with existing knowledge, prompting LLMs to generate ideas based on these entities. Si et al. (2024) suggest an iterative approach to retrieve topic-relevant papers through the Semantic Scholar API, utilizing retrieval-augmented generation (RAG) for idea generation. However, the constrained scope of knowledge acquisition in these methods often leads to superficial and repetitive ideas.

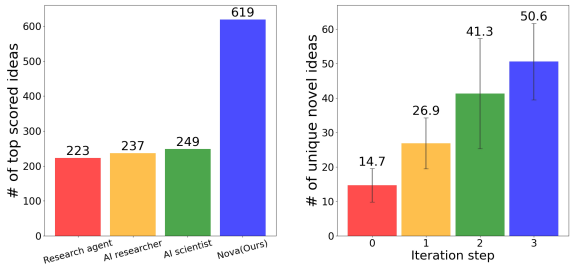

Figure 1: **Nova's Performance. The Left:** Comparison with the state-of-the-arts. Nova significantly outperforms other agents (Si et al., 2024; Baek et al., 2024; Lu et al., 2024) in generating high-quality ideas (Swiss Tournament Score is 5). **The Right:** The number of unique novel ideas at each iteration step. The iterative planning framework significantly enhances the generation of unique novel ideas, increasing by 3.4 times from the baseline.

Broadening the search scope, both in terms of breadth and depth, presents a significant challenge. The crux of the issue lies in determining which knowledge to retrieve. Traditional methods of entity and keyword retrieval are not goal-oriented and frequently yield knowledge that is not conducive to fostering innovation.

we introduce an iterative planning framework for LLM-based idea generation that specifically targets the enhancement of the novelty and diversity

of the ideas produced. Starting with seed ideas, our framework undergoes multiple iterations of planning and searching. In each iteration, the model is tasked with devising a search plan aimed at identifying papers that will enhance the novelty and diversity of the current set of ideas.

As depicted in Fig. 1,the proposed iterative planning framework significantly enhances the quality of ideas generated from recent 170 LLM-related papers (from top conferences like ACL, ICLR, and CVPR). The number of high-quality ideas (as measured by the Swiss Tournament Score (Si et al., 2024)) is at least 2.5 times greater than those produced by other state-of-the-art methods. Moreover, the number of unique novel ideas generated by our iterative planning framework is 3.4 times higher compared to approaches that do not incorporate such a framework.

2 Related work

2.1 LLM-based Scientific Innovation

In the past year, several studies on LLM-based scientific innovation (Yang et al., 2024; Baek et al., 2024; Lu et al., 2024; Wang et al., 2024b; Gu and Krenn, 2024) have been proposed, garnering significant attention from the LLM community. Among these studies, Baek et al. (2024) introduces a research agent that utilizes an external knowledge graph for co-occurrence entity search and integrates retrieved entities into idea generation of LLMs. To avoid generating similar ideas, Lu et al. (2024) treat past generated ideas as negative examples and instruct the LLM on what constitutes a negative example. To explore more external knowledge for innovation, some other works (Wang et al., 2024b; Gu and Krenn, 2024) propose prompting the LLM to generate ideas integrated with external knowledge, such as retrieved external entities or problem-solution pairs.

Concurrent with our research, Si et al. (2024) introduce AI-Researcher, which, for the first time, demonstrates that LLMs can generate ideas deemed more novel than those written by human experts. In addition, they point out that using LLMs to directly evaluate different dimensions of scientific ideas is unreliable and propose an idea ranking method based on pairwise comparison, achieving an accuracy of 71.4% in distinguishing accepted and rejected submissions on real ICLR 2024 data.

Although effective, the above approach often generate repetitive ideas (Si et al., 2024) due the lack of direction in acquiring new knowledge. In contract, our method provide a plan for searching new knowledge and suffer less from the repetitive problem.

2.2 Reasoning and Planning

Reasoning has been proven to be an effective technique for enhancing the problem-solving capabilities of LLMs, and several studies have been conducted to further promote LLMs' reasoning abilities. Wei et al. (2022) propose chain of thought (CoT), which involves guiding LLMs to solve complex problems by generating a step-by-step reasoning process. Later, Wang et al. (2023c) improve CoT by sampling and comparing diverse reasoning pathways to enhance the consistency of the reasoning process. To solve problems harder than the exemplars shown in prompts, Zhou et al. (2023) propose to break down the complex problem into a series of simpler subproblems and then solve them in sequence. Generalizing from Chain of Thought (CoT), Yao et al. (2023a) propose the Tree of Thought (ToT) framework, enabling LLMs to explore multiple reasoning paths and conduct self-evaluations when determining the next action. To enable more effective exploration of the solution space, Xie et al. (2024) enhance the reasoning capabilities of LLMs by introducing Monte Carlo Tree Search (MCTS) with iterative preference learning.

These methods significantly enhance the reasoning capabilities of LLMs; however, they seldom consider interacting with the external environment. To address this limitation, Trivedi et al. (2023) integrate CoT with knowledge retrieval, interleaving reasoning with searching to acquire additional external knowledge for knowledge-intensive question answering. Yao et al. (2023b) propose a ReAct paradigm combining reasoning and acting for solving language reasoning and decision-making tasks. It creates and adjusts high-level plans for acting while also interacting with the external environments to incorporate additional information into reasoning. Later, Aksitov et al. (2023) develop a ReAct-style LLM agent to reason and act upon external knowledge, using self-critique for self-improvement. By integrating reasoning and acting, these methods achieve more dynamic and contextually aware problem-solving based on both internal knowledge and external knowledge.

The reasoning capabilities of LLMs can also be applied to planning, such as generating plausible goal-driven action plans that can be enacted

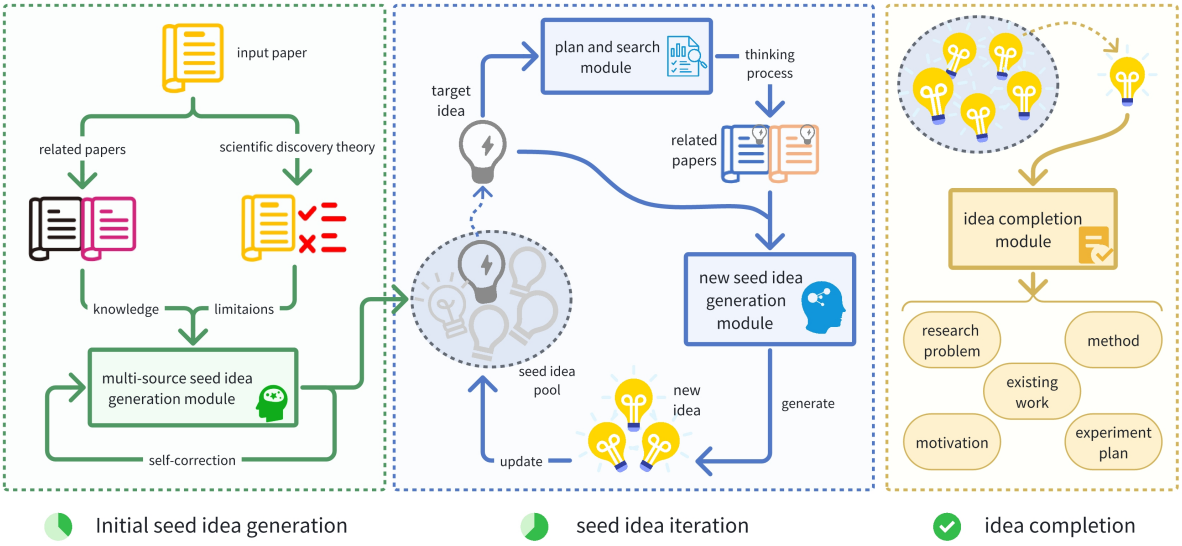

Figure 2: **Nova Pipeline.** The Pipeline includes initial seed idea generation, seed idea iteration, and idea completion. Upon receiving an input paper (*i.e.*, seed paper), the LLM is prompted to generate initial seed ideas by utilizing related papers (including recent publications) and scientific discovery methods. After that, the generated ideas are revised according to the new knowledge acquired according to iterative planning and search. Finally, each idea is expanded with more detailed methods.

in interactive, embodied environments (Huang et al., 2022). Additionally, plan-to-solve prompting (Wang et al., 2023b) can generate a plan that divides complex reasoning tasks into subtasks, enabling LLMs to execute each subtask according to the outlined plan.

Our work marks the inaugural integration of planning methodologies into the complex domain of research tasks.

3 Nova Pipeline

The pipeline for Nova is illustrated in Fig. 2. Our pipeline streamlines the research process through three stages: initial idea generation, iterative refinement, and detailed completion. It begins with an input paper, which the LLM uses to generate initial ideas by drawing on related literature and scientific discovery techniques. These ideas are then enhanced through iterative planning and search, incorporating new insights. The final step involves detailing the ideas. An example of the whole process is in Fig. 3.

3.1 Initial Seed Idea Generation

To produce high-quality ideas, we design a multi-source seed idea generation module that initiates with diverse and novel concepts. This module activates the LLM to generate ideas using related literature and scientific discovery techniques upon receiving an input paper. The prompt for initial idea generation is in Tab. 1 (details is in Tabl. 9) and an example of an initial seed idea is in Tab. 3.

To enrich the knowledge base with the most current insights, we utilize the input paper's references and have designed a knowledge tracking module. This module addresses the shortcomings of previous approaches by monitoring the latest publications. We pinpoint influential recent papers based on user engagement metrics such as likes, comments, and reposts across social media, forums, and GitHub. Furthermore, we harness LLMs to distill summaries of prevailing research trends from these papers, extracting valuable knowledge to enrich our target innovation efforts.

We employ 10 fundamental scientific discovery methods to guide LLMs in generating innovative ideas from an input paper and its associated literature. Drawing on Kuhn's paradigm (Kuhn, 1997) of scientific discovery, these methods help identify new research problems, such as analyzing anomalies in existing approaches. An example theory of scientific discovery can be found in Tab. 2 and all theories are in Table 20 in the Appendix.

To mimic human intuition, we tap into the LLMs' internal knowledge to craft initial seed ideas. This involves prompting the LLM to assess the short-

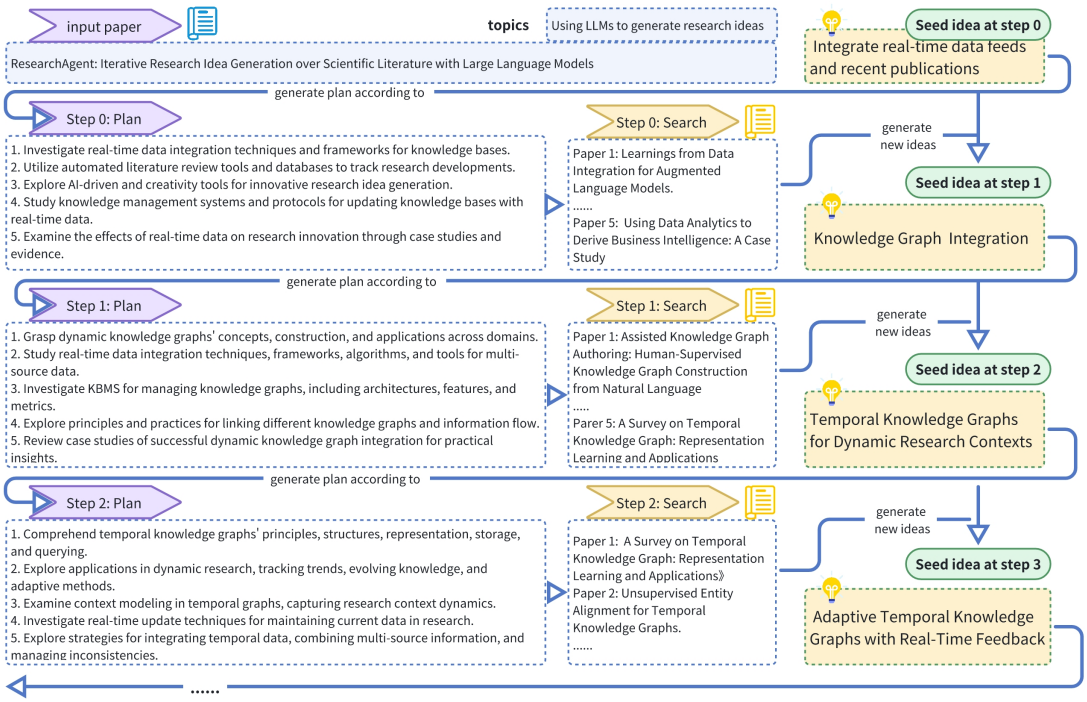

Figure 3: **Example of Planning-Driven Iterative Seed Idea Generation Process.** This example highlights the planning-driven iterative seed idea iteration process. Starting from an initial concept, a detailed plan is formulated to guide the search for relevant literature and acquire up-to-date knowledge.

comings of the input paper and related works, thereby sparking the creation of fresh ideas.

To prevent hallucination and improve the logicality of generated initial seed ideas, we also utilize self-correction mechanics: self-check Miao et al. (2023), self-critique Gou et al. (2024), and reflection Shinn et al. (2023). These methods partly guarantee that the generated seed ideas are logical and reasonable. In the end, we generate 15 seed ideas for each input paper.

3.2 Iterative Planning and Search for Seed Idea Iteration

Once an initial seed idea pool is generated, we start to iteratively planning and search new knowledge according to the see idea and generate new idea according to the acquired new knowledge.

3.2.1 Planning and Search

In planning and search step, we guide the LLM to identify key fields for comprehensive and novel knowledge acquisition to enhance further research and idea generation according to the given ideas.

This approach, demonstrated through an in-context learning example, leverages the LLM's internal knowledge to determine useful knowledge for new ideas, surpassing traditional entity or keyword-based retrieval methods.

New Seed Idea Generation. Once new knowledge is acquired, the new seed idea is generated based on the retrieved papers, the initial seed idea, and given input paper. For each idea, the LLM generate 10 new seed ideas and then uses self-reflection to cut the number down to 3.

In each iteration, the old see ideas are replaced with the newly generated seed ideas. This allows our agent to dive deeper, largely expanding the scope of search. Therefore, in each iteration, we generate 3 times more seed ideas.

3.3 Output Idea Generation

After finishing T step iteration, we have a final seed idea pool. We then expand the see idea into the initial proposal and final proposal as in (Si et al., 2024). Specifically, given an input paper and its corresponding seed idea , we ask LLM to decompose the idea into several sub-modules and utilize LLMs to design these sub-modules separately in a more detailed way(details is in Table 32 in Appendix). An example of an initial proposal and final proposal are in Tab. 4 and Tab. 5, separately.

Prompt
Role: You are an expert researcher in AI. Your goal is to propose some innovative and valuable research ideas based on the target paper. **Skill:** Generate subsequent exploration ideas according to the following steps: Understanding of the target paper and related papers is essential: - The target paper is the primary research study you aim to enhance or build upon through future research, serving as the central source and focus for identifying and developing the specific research idea. - The referenced papers are studies that the target paper has cited, indicating their direct relevance and connection to the primary research topic you are focusing on, and providing additional context and ideas that are essential for understanding and expanding upon the target paper. **Step 1:** Combine target paper and referenced paper to answer the following information: 1. What are the tasks, methods, and main innovations of the current paper? 2. What are the weaknesses and limitations of the current paper? **Step 2:** Propose some valuable and new research ideas. **Output Format:** {qa_info_with_idea_json_format} **Requirements:** ... **Input:** ... **Output:** ...

Table 1: Prompt for initial seed idea generation.

Define New Scientific Problems
Theoretical Basis: Kuhn's paradigm theory, Laudan's problem-solving model, Nichols's problem-generation theory. **Method:** Identify anomalies in existing theories; explore theoretical boundaries and scope of application; integrate interdisciplinary knowledge and discover new problems; re-examine neglected historical problems.

Table 2: An example theory of scientific discovery

4 Experiment

To validate our proposed iterative planning framework, we conduct an ablation study and perform comprehensive comparisons with state-of-the-art research idea generation methods.

4.1 Experimental Setup

Data. Our dataset is constructed by collecting high-quality papers from top conferences. The initial corpus comprised 7,805 papers from CVPR 2024, ACL 2024, and ICLR 2024. The keywords related to "LLM" are used to filter the initial corpus down to about 2,000 papers. Minimum citation number is used to further cut down the number to 153, with citation thresholds set at 20 for ICLR 2024 and 10 for CVPR 2024 and ACL 2024. We add an additional 17 papers from Hugging Face Daily Papers according to their user ratings. At the end, the dataset consists of 170 papers, each of which is used to generate 100 ideas for subsequent evaluation.

Baseline. To compare our proposed approach with the state-of-the-art approaches, we choose three leading approaches as the baselines, including AI-Researcher (Si et al., 2024), AI-Scientist (Lu et al., 2024) and Research-Agent (Baek et al., 2024). For AI-researcher and AI-Scientist, we run the original code on our seed papers. For Research-Agent, we also build a knowledge graph based on the description of the paper.

Offline Evaluation. In our offline evaluation, we mainly focus on overall quality evaluation and also concern with the novelty and diversity.

1. Quality. Following the work (Si et al., 2024), we employ the Swiss System Tournament [1] with Claude-3.5-Sonnet zero-shot ranker to evaluate the quality of ideas. The ranker makes pairwise comparisons to determine which idea is better. For

[1] https://en.wikipedia.org/wiki/Swiss-system_tournament

Thinking: The target paper's reliance on existing literature may limit the generation of truly novel ideas. By incorporating real-time data sources, such as ongoing research developments or recent publications, the ResearchAgent can generate more innovative and timely research ideas.
Idea: Incorporate multi-modal data sources, including experimental data, patents, and industry reports, into the ResearchAgent's knowledge base to generate more comprehensive and interdisciplinary research ideas.
Keywords: multi-modal data sources, experimental data, interdisciplinary research ideas

Table 3: Seed Idea Example.

Problem: State the problem statement, which should be closely related to the idea description and something that large language models cannot solve well yet.
Existing Methods: Mention some existing benchmarks and baseline methods if there are any.
Motivation: Explain the inspiration of the proposed method and why it would work well.
Proposed Method: Propose your new method and describe it in detail. The proposed method should be maximally different from all existing work and baselines, and be more advanced and effective than the baselines. You should be as creative as possible in proposing new methods; we love unhinged ideas that sound crazy. This should be the most detailed section of the proposal.
Experiment Plan: Specify the experiment steps, baselines, and evaluation metrics.

Table 4: Initial Proposal Template (follow Si et al. (2024)).

each idea, there are 5 rounds of comparison, each winning comparison get 1 score. Such a quality evaluation method (Si et al., 2024) has been shown to be better than direct comparison (Lu et al., 2024).

2. Novelty. Following the work (Baek et al., 2024), we use LLMs to judge whether a generated idea is novel by checking the top 10 most relevant papers [2] and if no paper is identified as containing the same idea, it is considered novel.

3. Diversity. We use the proportion of unique ideas to measure generation diversity. To be specific, we shuffle the current set of ideas and calculate their embeddings using the all-MiniLM-L6-v2 model[3]. We use the cosine distance for the similarity measure. Similar to Si et al. (2024), the threshold for duplication is set to be 0.8.

In the offline evaluation, we utilize Nova alongside three baseline methods to generate 400 ideas from a given input paper. Each method generates 100 ideas separately. The iteration step for Nova is set to 3, and the initial number of seed ideas is 15.

[2]We use the same paper retriever as in the idea generation phase, with data sourced from arXiv. The time range spans from January 1, 2022, to August 2024, and the categories include literature related to AI, NLP, and CV.

[3]https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2

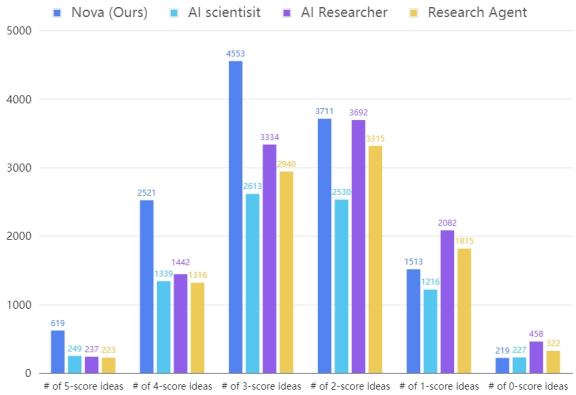

Figure 4: Score distribution of different methods in Swiss Tournament. The results indicate that Nova not only generates more unique ideas but also produces a greater proportion of high-quality ideas. 619 and 2521 ideas generated by Nova are scored at 4 and 5, significantly surpassing the baseline methods.

After 3 iterations, we get 405 initial seed ideas, we then cluster the ideas into 100 clusters and used the cluster center to generate the 100 final ideas. The implementation details for the baseline methods remain consistent with those in the original work.

Human Evaluation. To validate the effectiveness of our offline evaluation, we have an additional human evaluation. Our goal is to assess how well our

Title: A concise statement of the main research question to be used as the paper title.
Problem Statement: Clearly define the problem your research intends to address. Explain clearly why this problem is interesting and important.
Motivation: Explain why existing methods are not good enough to solve the problem, and explain the inspiration behind the new proposed method. You should also motivate why the proposed method would work better than existing baselines on the problem.
Proposed Method: Explain how the proposed method works, describe all the essential steps.
Step-by-Step Experiment Plan: Break down every single step of the experiments, make sure every step is executable. Cover all essential details such as the datasets, models, and metrics to be used. If the project involves prompting, give some example prompts for each step.

Table 5: Final Proposal Template (follow Si et al. (2024)).

offline evaluations align with human expert evaluations. We recruit a panel of 10 experts, all holding a PhD degree or professorship in natural language processing, machine learning, or computer vision, doing research in LLMs-related fields. These experts evaluated ideas based on novelty and overall quality (including feasibility and effectiveness).

We select five ideas generated by each agent based on the same input paper. These ideas correspond to the 1st, 25th, 50th, 75th, and 100th percentiles according to the offline evaluation, resulting in a total of 20 ideas per topic. This process is repeated for 20 times. Each expert review four groups, ensuring that at least two independent experts evaluate each idea. The final score for each idea is averaged across all ratings from different experts.

We compare the distributions of expert evaluations against those in offline evaluation. Specifically, we track which methods produce the top 20 percent of ideas, as rank by the experts. This helps determine which methods outperform the others. Moreover, this approach reveals whether the model evaluation aligns with human evaluation. The potential alignment between model and human assessments offers insights into the efficacy and reliability of using LLM agents for evaluating scientific ideas.

4.2 Experimental Results

4.2.1 Offline Evaluation Results

The offline quality comparison results are shown in Fig. 4. The novelty and diversity comparison are shown in Fig. 5. As can be seen,

Nova achieves a significantly higher Swiss score. By incorporating iterative planning and search for external knowledge retrieval, Nova engages in more effective exploration for innovation. This may significantly enhance the novelty of the generated ideas. Since novelty is often the most important factor in evaluating idea quality, Nova consistently achieves the highest Swiss score. The complete comparison results are shown in Fig. 4. 619 and 2521 of the ideas generated by Nova are scored at 4 and 5, significantly surpassing the performance of other agents.

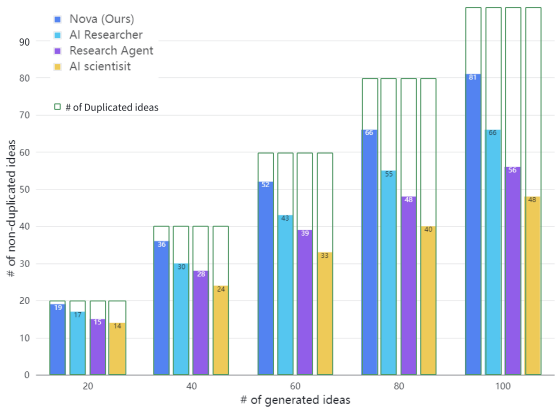

Figure 5: Non-Duplicate Percentage Comparison.

Nova generates significantly more diverse ideas. As the number of generated ideas increases, Nova can continuously generate new ideas through iterative planning and search. In Non-Duplicate Percentage, Nova significantly outperforms others, with over 80% of the ideas being unique, as shown in Fig. 5.

4.2.2 Human Evaluation Results

In our human evaluation, Nova achieves the highest scores for both overall quality and novelty. It consistently generates a greater proportion of top-

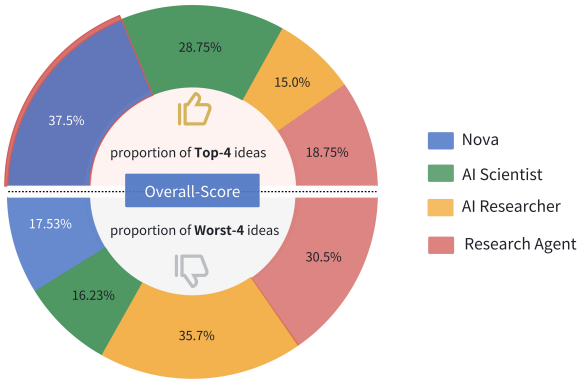

Figure 6: Human Evaluation for Overall Quality.

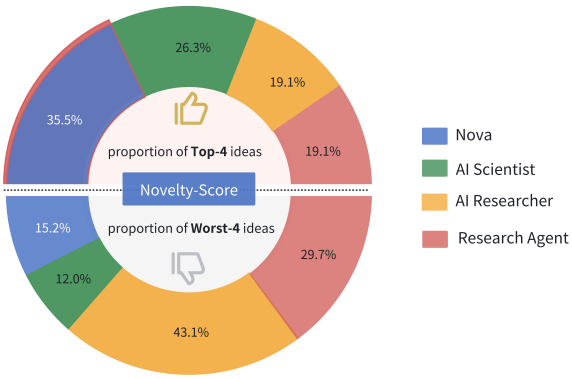

Figure 7: Human Evaluation for Novelty.

scoring ideas compared to the baselines in terms of both overall quality and novelty. As shown in Fig. 6, Nova contributes 37.5% of the top 4 ideas, the highest among the four methods. Additionally, Nova has a notably low percentage of the worst 4 ideas, accounting for only 17.53% in terms of overall quality. In Fig. 7, a similar pattern is observed in novelty evaluation.

Our human and automated evaluations show strong consistency in distinguishing between the top-rated and worst-rated ideas. By comparing the distribution of top-rated ideas in both human and automated evaluations (Fig. 4 and 6), it is evident that human reviewers and the LLM evaluate the performance of the four methods in a similar pattern. In both human and offline evaluations, our method generates the highest proportion of top-rated ideas, followed by AI-Scientist, ResearchAgent, and finally AI-Researcher. This indicates that our offline review mechanism effectively captures human reviewers' true preferences.

4.3 Ablation Study

To assess the effectiveness of planning and search in Nova, we conduct comparisons by gradually

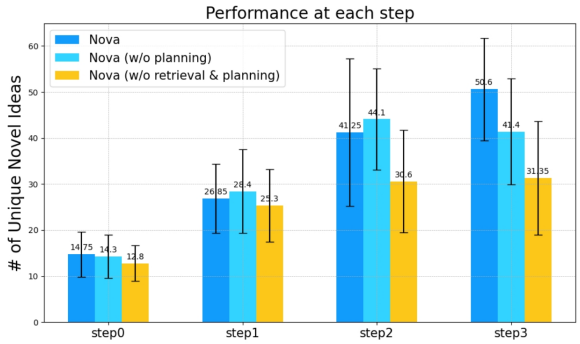

Figure 8: Ablation studies for Nova. We can find that both retrieval and planning significantly enhance the generation of unique novel ideas.

removing planning and retrieval components. All methods retrieve the same number of papers, specifically $K = 5$. Both retrieval and planning are found to significantly enhance the generation of unique and novel ideas. When planning is excluded, the number of unique ideas at step 3 (44.1) is no longer increases compared to step 2 (42.4). This suggests that without planning, relying solely on retrieval based on seed ideas limits access to valuable external knowledge for innovation. This limitation may arise from the restricted scope of search when planning is absent. Obviously, when planning and retrieval are both removed, the number of unique novel ideas increases slightly at step 2 (from 25.3 to 30.6) and stagnates at step 3 (from 30.6 to 31.35), due to no external knowledge are introduced.

5 Conclusion

In this paper, we propose an LLM-based scientific innovation method, Nova, which introduces iterative planning and search to retrieve external knowledge for innovation. Nova leverages the internal knowledge of LLMs to generate search plans for external knowledge retrieval, significantly enhancing the effectiveness of the retrieval process. Ablation study demonstrates the effect of the iterative planning and search framework on promoting the novelty of generating ideas. The offline and human evaluations show that Nova significantly and consistently outperforms state-of-the-art scientific innovation methods. In the future, we will explore incorporating a reward function into our iterative planning framework to further enhance external knowledge retrieval.

6 Limitations

in this work, we investigate using iterative planning and search, mimicking the manner of our human beings, to enhance the innovation capability of existing LLM-based methods. Despite promising findings, some limitations remain in this work, which we discuss below: This reduce the , but may

Limited Iterations Steps. Although our approach can significantly enhance the novelty and diversity of generated ideas through iteration, we do not investigate whether the performance of Nova will continuously improve with an increasing number of iterations.

Planning without Rewards. In our planning and search framework, we do not introduce reward functions but only use the internal knowledge of LLMs to generate search plans. This may limit the effectiveness of planning.

We hope these findings inspire future investigations into using LLM to comprehensively integrate both internal and external knowledge for LLM-based scientific innovation. We believe addressing each of these shortcomings will lead to exciting feature directions.

7 Ethics Statement

Publication Policy. The increasing use of AI to generate research ideas poses significant challenges to academic integrity. The growing accessibility of LLMs and the rising usefulness of LLMs in research may lead to deterioration in the overall quality of scholarly content, as individuals may rely on AI for both creativity and submission reviews. Therefore, there is a legitimate concern that students or researchers would exploit these technologies and present low-quality research proposals. To mitigate these risks, it is crucial to hold accountability for outputs generated through AI tools in scientific submissions.

Intellectual Credit. Generative AI in the research cycle poses great concerns about intellectual credit on the submitted works. While traditional frameworks were more like a tool for human researchers, LLMs are more potent in a way that plays a more significant role in the scientific research process if used. It is still unclear how intellectual credit should be distributed in the case of AI-driven research. To better attribute credit to AI-supported research, researchers should adopt transparent documentation about their research process, including the extent of AI involvement in generating ideas and developing experiments.

Potential for Misuse. AI-generated research ideas, particularly those introducing novel concepts, possess the potential for misuse. This could lead to harmful outcomes. Ideation agents may be exploited to develop adversarial attack strategies or other unethical applications. Therefore, it is important to develop anti-jailbreak mechanisms or safety checks on AI-generated content and the use of generative AI in research.

Idea Homogenization. If AI was widely used in scientific research, this would raise concerns about the potential idea of homogenization. The wide adoption of LLMs in research could reflect a narrower set of perspectives or systematic biases compared to human researchers not using AI assistance. Therefore, it is important to recognize the limitations of current LLM-generated ideas, and future work should focus more on enhancing the generation diversity either by improving the models themselves or by refining the ideation process.

Impact on Human Researchers. The challenge posed by AI's integration into research should be well recognized because research is fundamentally and historically a community-driven and collaborative effort. It is still unclear on the negative consequences of the introduction of AI in the research process. People should be cautious and aware of the potential decline in human thought and a reduction in opportunities for human collaboration after the introduction of AI in research. Future works should explore other methods of human-AI collaboration. Understanding how LLM should be integrated into the research process will be an ongoing problem.

References

Renat Aksitov, Sobhan Miryoosefi, Zonglin Li, Daliang Li, Sheila Babayan, Kavya Kopparapu, Zachary Fisher, Ruiqi Guo, Sushant Prakash, Pranesh Srinivasan, Manzil Zaheer, Felix Yu, and Sanjiv Kumar. 2023. REST MEETS REACT: SELF-IMPROVEMENT FOR MULTI-STEP REASONING LLM AGENT. *Preprint*, arXiv:2312.10003.

Jinheon Baek, Sujay Kumar Jauhar, Silviu Cucerzan, and Sung Ju Hwang. 2024. ResearchAgent: Iterative Research Idea Generation over Scientific Literature with Large Language Models. *Preprint*, arXiv:2404.07738.

Zhibin Gou, Zhihong Shao, Yeyun Gong, Yelong Shen, Yujiu Yang, Nan Duan, and Weizhu Chen.

2024. CRITIC: Large Language Models Can Self-Correct with Tool-Interactive Critiquing. *Preprint*, arXiv:2305.11738.

Xuemei Gu and Mario Krenn. 2024. Generation and human-expert evaluation of interesting research ideas using knowledge graphs and large language models. *Preprint*, arXiv:2405.17044.

Qian Huang, Jian Vora, Percy Liang, and Jure Leskovec. 2024. MLAgentBench: Evaluating Language Agents on Machine Learning Experimentation. In *ICML 2024*.

Wenlong Huang, Pieter Abbeel, Deepak Pathak, and Igor Mordatch. 2022. Language Models as Zero-Shot Planners: Extracting Actionable Knowledge for Embodied Agents. In *ICML 2022*, volume 162 of *Proceedings of Machine Learning Research*, pages 9118–9147. PMLR.

Thomas S Kuhn. 1997. *The structure of scientific revolutions*, volume 962. University of Chicago press Chicago.

Chris Lu, Cong Lu, Robert Tjarko Lange, Jakob Foerster, Jeff Clune, and David Ha. 2024. The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery. *Preprint*, arXiv:2408.06292.

Ning Miao, Yee Whye Teh, and Tom Rainforth. 2023. SelfCheck: Using LLMs to Zero-Shot Check Their Own Step-by-Step Reasoning. *Preprint*, arXiv:2308.00436.

Bernardino Romera-Paredes, Mohammadamin Barekatain, Alexander Novikov, Matej Balog, M Pawan Kumar, Emilien Dupont, Francisco JR Ruiz, Jordan S Ellenberg, Pengming Wang, Omar Fawzi, et al. 2024. Mathematical discoveries from program search with large language models. *Nature*, 625(7995):468–475.

Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. 2023. Reflexion: language agents with verbal reinforcement learning. In *NeurIPS 2023*.

Chenglei Si, Diyi Yang, and Tatsunori Hashimoto. 2024. Can LLMs Generate Novel Research Ideas? A Large-Scale Human Study with 100+ NLP Researchers. *Preprint*, arXiv:2409.04109.

Harsh Trivedi, Niranjan Balasubramanian, Tushar Khot, and Ashish Sabharwal. 2023. Interleaving Retrieval with Chain-of-Thought Reasoning for Knowledge-Intensive Multi-Step Questions. In *ACL 2023*, pages 10014–10037, Toronto, Canada. Association for Computational Linguistics.

Haiming Wang, Huajian Xin, Chuanyang Zheng, Lin Li, Zhengying Liu, Qingxing Cao, Yinya Huang, Jing Xiong, Han Shi, Enze Xie, et al. 2024a. Lego-prover: Neural theorem proving with growing libraries.

Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak, Shengchao Liu, Peter Van Katwyk, Andreea Deac, et al. 2023a. Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972):47–60.

Lei Wang, Wanyu Xu, Yihuai Lan, Zhiqiang Hu, Yunshi Lan, Roy Ka-Wei Lee, and Ee-Peng Lim. 2023b. Plan-and-Solve Prompting: Improving Zero-Shot Chain-of-Thought Reasoning by Large Language Models. In *ACL 2023*, pages 2609–2634, Toronto, Canada. Association for Computational Linguistics.

Qingyun Wang, Doug Downey, Heng Ji, and Tom Hope. 2024b. Scimon: Scientific inspiration machines optimized for novelty. *Preprint*, arXiv:2305.14259.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V. Le, Ed H. Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2023c. Self-Consistency Improves Chain of Thought Reasoning in Language Models. In *ICLR 2023*. OpenReview.net.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. 2022. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In *NeurIPS 2022*.

Yuxi Xie, Anirudh Goyal, Wenyue Zheng, Min-Yen Kan, Timothy Lillicrap, Kenji Kawaguchi, and Michael Shieh. 2024. Monte Carlo Tree Search Boosts Reasoning via Iterative Preference Learning. *Preprint*, arXiv:2405.00451.

Zonglin Yang, Xinya Du, Junxian Li, Jie Zheng, Soujanya Poria, and Erik Cambria. 2024. Large Language Models for Automated Open-domain Scientific Hypotheses Discovery. *Preprint*, arXiv:2309.02726.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Tom Griffiths, Yuan Cao, and Karthik Narasimhan. 2023a. Tree of Thoughts: Deliberate Problem Solving with Large Language Models. In *NeurIPS 2023*.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik R. Narasimhan, and Yuan Cao. 2023b. ReAct: Synergizing Reasoning and Acting in Language Models. In *ICLR, 2023*.

Denny Zhou, Nathanael Schärli, Le Hou, Jason Wei, Nathan Scales, Xuezhi Wang, Dale Schuurmans, Claire Cui, Olivier Bousquet, Quoc V. Le, and Ed H. Chi. 2023. Least-to-Most Prompting Enables Complex Reasoning in Large Language Models. In *ICLR 2023*. OpenReview.net.

A Human Annotation

This section provides details about human annotation in our human evaluation experiment.

A.1 Annotation Instructions

The complete annotation instruction for the idea reviewer is given below:

Please evaluate the given twenty ideas based on four criteria (Novelty, Feasibility, Effectiveness, and Overall), identify the best four and the worst four ideas, and rank them accordingly. The principles include:

Annotation Dimensions: See Table 6 for detailed dimension instructions.

Annotation Method: Annotate the best 4 and the worst 4 ideas in each of the 4 dimensions. For the best ideas, mark them as 1, 2, 3, and 4 respectively, while 1 refers to the best idea. For the worst ideas, mark them as 17, 18, 19, and 20, while 20 refers to the worsts idea. No need to annotate ideas other than the best 4 and the worst 4. For an example, see Table 7.

A.2 Data Description

We manually selected 20 different papers from ACL2024, CVPR2024, and ICLR2024. Each paper is carefully selected, and the 20 papers are from various research fields, including Natural Language Processing, Computer Vision, and Large Language Models in general, and have varied academic significance measured by citations to represent a broad scope of research papers. The online pilot study gives the human evaluators a form of twenty rows and five columns, along with a hyperlink to the original paper the ideas are generated. Each row is of an idea generated by one of the four different methods, Nova, AI-Scientist, AI-Researcher, or ResearchAgent. Still, human experts have no information on which method to generate that particular idea. The four columns are summary, novelty, feasibility, effectiveness, and overall. The summary is the research plan generated from one of the four methods, and the remaining four columns are entries for human experts to input their rankings. The four best ideas are labeled as 1, 2, 3, or 4, and the four worst ideas are labeled as 17, 18, 19, or 20. The rest of the entries are left blank.

A.3 Risk Statement

Physical Risk. This study does not involve any activities that may cause physical harm or discomfort.

Psychological Risk. This study does not involve sensitive topics or psychological experiments.

Social Risk. This study does not involve activities that could affect participants' social relationships or reputation. No personal information will be disclosed.

Economic Risk. This study will not result in any economic loss for the participants.

Privacy and Data Security Risk. All annotation data will be randomly assigned to anonymous experts. No personally sensitive information will be collected.

B Prompts and Examples

This section provides a comprehensive overview of various prompts and examples Nova uses in idea generation and research proposal creation. The tables are organized in Table 8 to guide the reader through different methods and stages of idea development, research trend exploration, and proposal drafting, with each table focusing on a distinct aspect of the research process.

Criteria	Definition
Novelty	Novelty refers to the originality and innovativeness of the idea. It assesses how new and unique the idea is compared to existing work in the field.
Overall	Overall evaluates the general quality and potential of the idea, taking into account all other criteria (Novelty, Feasibility, and Effectiveness). It provides a holistic assessment of the idea's value.
Feasibility	Feasibility assesses the practicality and implementability of the idea. It considers whether the idea can be realistically executed with available resources and within a reasonable timeframe.
Effectiveness	Effectiveness evaluates the expected impact and success of the idea in achieving its intended goals. It considers how well the idea is likely to perform in practice.

Table 6: Evaluation Criteria and Definitions for Online Idea Assessment Based on Novelty, Feasibility, Effectiveness, and Overall Quality

Rank	Label
Best	1
Second Best	2
Third Best	3
Fourth Best	4
...	...
...	...
Fourth Worst	17
Third Worst	18
Second Worst	19
Worst	20

Table 7: An example of Ranking Labels for Annotating the Best and Worst Ideas Across Four Evaluation Dimensions

Table 8: Tables in the Appendix

Table Number	Table Title
Table 6	Prompt for initial seed idea generation using inner knowledge from LLM
Table 7	An example of the initial seed idea generation using inner knowledge from LLM (Part 1)
Table 8	An example of the initial seed idea generation using inner knowledge from LLM (Part 2)
Table 9	Prompt for generating research trends
Table 10	Current hot research trends in AI (Part 1)
Table 11	Current hot research trends in AI (Part 2)
Table 14	Prompt for generating research ideas based on popular research trends
Table 12	An example of generating research ideas based on popular research trends (Part 1)
Table 13	An example of generating research ideas based on popular research trends (Part 2)
Table 17	General theory of scientific discovery (Part 1)
Table 18	General theory of scientific discovery (Part 2)
Table 19	Prompt for idea generation based on the general theory of scientific discovery (Part 1)
Table 20	Prompt for idea generation based on the general theory of scientific discovery (Part 2)
Table 21	An example of idea generation based on the general theory of scientific discovery (Part 1)
Table 22	An example of idea generation based on the general theory of scientific discovery (Part 2)
Table 23	Prompt for expanding idea generation using retrieved knowledge and reflection iteration
Table 24	An example of original idea
Table 25	An example of iterative research ideas
Table 26	Prompt for initial proposal generation
Table 27	An example of initial proposal
Table 28	Prompt for method decomposition
Table 29	An example of method decomposition (Part 1)
Table 30	An example of method decomposition (Part 2)
Table 31	Prompt for final proposal generation (Part 1)
Table 32	Prompt for final proposal generation (Part 2)
Table 33	An example of final proposal (Part 1)
Table 34	An example of final proposal (Part 2)
Table 35	Prompt for search plan generation
Table 36	An example of search plan
Table 37	An example of search result

Prompt
Role: You are an expert researcher in AI. Your goal is to propose some innovative and valuable research ideas based on the target paper. **Skill:** Generate subsequent exploration ideas according to the following steps: Understanding of the target paper and related papers is essential: - The target paper is the primary research study you aim to enhance or build upon through future research, serving as the central source and focus for identifying and developing the specific research idea. - The referenced papers are studies that the target paper has cited, indicating their direct relevance and connection to the primary research topic you are focusing on, and providing additional context and ideas that are essential for understanding and expanding upon the target paper. **Step 1:** Combine target paper and referenced paper to answer the following information: 1. What are the tasks, methods, and main innovations of the current paper? 2. What are the weaknesses and limitations of the current paper? **Step 2:** Propose some valuable and new research ideas. **Output Format:** {qa_info_with_idea_json_format} **Requirements:** 1. Output about 5 new ideas worth exploring. 2. Aim for new research ideas that can potentially win best paper awards at top conferences like ACL, NeurIPS, ICLR, and CVPR. 3. Please output your thought process. 4. Please think step by step. **Input:** target_title: {paper.title} target_abstract: {paper.abstract} referenced_paper_title: {paper.related_paper_titles} referenced_paper_abstract: {paper.related_paper_abstract} **Output:** Thinking: <output your thinking process here, explain why it should have changed to win the best paper awards at top conferences> New research idea that improves target_paper: <JSON>

Table 9: Prompt for initial seed idea generation using inner knowledge from LLM.

ResearchAgent: Iterative Research Idea Generation over Scientific Literature with Large Language Models
Generated Research Ideas (Part 1)
1. **Integrate Real-Time Data Feeds** *Thinking:* The target paper's reliance on existing literature may limit the generation of truly novel ideas. By incorporating real-time data sources, such as ongoing research developments or recent publications, the ResearchAgent can generate more innovative and timely research ideas. *Idea:* Integrate real-time data feeds and recent publications into the ResearchAgent's knowledge base to ensure that the generated ideas are aligned with the latest scientific advancements. *Keywords:* real-time data integration, recent publications tracking, innovative research ideas
2. **Develop a Multi-Faceted Evaluation Framework** *Thinking:* The evaluation metrics used in the target paper may not fully capture the practical applicability and impact of the generated ideas. By incorporating domain-specific expert reviews and practical implementation feedback, the evaluation process can be more comprehensive. *Idea:* Develop a multi-faceted evaluation framework that includes domain-specific expert reviews and practical implementation feedback to assess the applicability and impact of the generated research ideas. *Keywords:* multi-faceted evaluation, expert reviews, practical implementation feedback
3. **Optimize Computational Efficiency** *Thinking:* The iterative refinement process in the target paper might be computationally expensive and time-consuming. By optimizing the computational efficiency of the ReviewingAgents and leveraging parallel processing techniques, the scalability of the ResearchAgent can be improved. *Idea:* Optimize the computational efficiency of the ReviewingAgents and leverage parallel processing techniques to enhance the scalability of the ResearchAgent. *Keywords:* computational efficiency, parallel processing, scalability

Table 10: An example of seed idea generation using inner knowledge from LLM (Part 1).

ResearchAgent: Iterative Research Idea Generation over Scientific Literature with Large Language Models
Generated Research Ideas (Part 2)
4. **Incorporate Multi-Modal Data Sources** *Thinking:* The target paper focuses on generating research ideas based on scientific literature. By incorporating multi-modal data sources, such as experimental data, patents, and industry reports, the ResearchAgent can generate more comprehensive and interdisciplinary research ideas. *Idea:* Incorporate multi-modal data sources, including experimental data, patents, and industry reports, into the ResearchAgent's knowledge base to generate more comprehensive and interdisciplinary research ideas. *Keywords:* multi-modal data sources, experimental data, interdisciplinary research ideas
5. **Incorporate Diverse ReviewingAgents** *Thinking:* The target paper uses ReviewingAgents instantiated with human preference-aligned LLMs for iterative feedback. By incorporating a diverse set of ReviewingAgents with different expertise and perspectives, the feedback process can be more robust and holistic. *Idea:* Incorporate a diverse set of ReviewingAgents with different expertise and perspectives to provide more robust and holistic feedback during the iterative refinement process. *Keywords:* diverse ReviewingAgents, expertise diversity, holistic feedback

Table 11: An example of seed idea generation using inner knowledge from LLM (Part 2).

Prompt
Role: You are an AI expert researcher. You can summarise the current hot research trends from the list of recent AI papers. **Skill:** You will analyze the research trending based on the recent popular paper, provide us with the research trending report. **Requirements:** 1. Provide a comprehensive analysis, including the hot research directions, the highlights of the technologies and methods, and discuss whether these technologies can be used in other fields. I will provide a list of recent popular paper list here: {popular_paper_list} Then, Please output the current research trending report here:

Table 12: Prompt for generating popular research trends.

Hot Research Directions
1. Long-Context Language Models (LLMs) and Retrieval-Augmented Generation (RAG) - **Key Papers**: "RAG in the Era of Long-Context LLMs", "LongCite", "MemLong", "Improved RAG with Self-Reasoning", "LongWriter", "EfficientRAG", "Enhanced RAG with Long-Context LLMs", "GraphReader" - **Highlights**: Addressing the challenge of maintaining focus and relevance in long-context LLMs. Combining RAG mechanisms with long-context capabilities to improve performance in tasks like question answering and citation generation. Innovations such as order-preserving RAG, external retrievers, and graph-based systems to enhance context handling. - **Cross-Field Applications**: These advancements can be applied in fields requiring extensive document analysis, such as legal research, academic literature review, and medical records analysis.
2. Strategic Chain-of-Thought (CoT) and Self-Improvement Techniques - **Key Papers**: "Strategic Chain-of-Thought", "Teaching LLM Agents to Self-Improve", "Self-Taught Evaluators", "Meta-Rewarding LLMs", "SelfGoal" - **Highlights**: Incorporating strategic knowledge to guide intermediate reasoning steps. Iterative self-improvement and self-evaluation to enhance model performance over multiple turns. Use of self-generated training data to refine judgment and reasoning capabilities. - **Cross-Field Applications**: These methods can be beneficial in educational technologies, autonomous decision-making systems, and any domain requiring iterative problem-solving and learning.
3. Mixture-of-Experts (MoE) and Multi-Agent Systems - **Key Papers**: "OLMoE", "Agentic RAG for Time Series Analysis", "Mixture-of-Agents", "MindSearch" - **Highlights**: Leveraging sparse Mixture-of-Experts to optimize model performance and efficiency. Multi-agent architectures for specialized task handling, such as time series analysis and complex web-information seeking. - **Cross-Field Applications**: These approaches can be utilized in financial forecasting, climate modeling, and complex system simulations where specialized expertise is crucial.
4. Synthetic Data Generation and Utilization - **Key Papers**: "Smaller, Weaker, Yet Better", "Scaling Synthetic Data Creation", "Improving Retrieval in LLMs through Synthetic Data", "Model Collapse on Synthetic Data" - **Highlights**: Using weaker models to generate high-quality synthetic data for fine-tuning stronger models. Addressing the challenges of model collapse due to recursive training on synthetic data. - **Cross-Field Applications**: Synthetic data can be used in privacy-preserving data analysis, training AI models in healthcare, and augmenting datasets in low-resource languages.

Table 13: Current hot research trends in AI (Part 1)

Hot Research Directions
5. Controllable and Robust Text Generation - **Key Papers**: "Controllable Text Generation for LLMs", "Enhancing Robustness in LLMs", "Improving Legibility of LLM Outputs" - **Highlights**: Techniques for controlling the style, safety, and consistency of generated text. Methods to enhance robustness by filtering out irrelevant information and improving the clarity of outputs. - **Cross-Field Applications**: These advancements are crucial for developing reliable AI assistants, automated content generation, and ensuring the safety of AI-generated outputs in sensitive applications.
6. AI in Scientific Discovery and Evaluation - **Key Papers**: "The AI Scientist", "Automate Design of Agentic Systems", "Self-Taught Evaluators" - **Highlights**: AI agents capable of conducting independent research and writing scientific papers. Meta-agent frameworks for designing and evaluating agentic systems. - **Cross-Field Applications**: These innovations can revolutionize scientific research, enabling faster discovery and validation of new theories across various scientific disciplines.
7. Advanced Prompt Engineering and Personalization - **Key Papers**: "Conversational Prompt Engineering", "A Survey of Prompt Engineering Methods in LLMs" - **Highlights**: Techniques for creating personalized prompts through iterative user interaction. Comprehensive surveys on prompt engineering methods for various NLP tasks. - **Cross-Field Applications**: Personalized prompt engineering can enhance user experience in customer service bots, personalized education platforms, and adaptive learning systems.
8. AI in Code and Software Engineering - **Key Papers**: "LLM Compiler", "From LLMs to LLM-based Agents for Software Engineering", "DeepSeek-Coder-V2" - **Highlights**: Models designed for code optimization and generation. Surveys on the application of LLMs in software engineering tasks like requirement engineering and test generation. - **Cross-Field Applications**: These advancements can improve software development workflows, automate code reviews, and enhance the capabilities of integrated development environments (IDEs).

Table 14: Current hot research trends in AI (Part 2)

ResearchAgent: Iterative Research Idea Generation over Scientific Literature with Large Language Models
Generated Research Ideas from Popular Knowledge (Part 1)
1. **Integrate Synthetic Data Generation Techniques** *Thinking:* The target paper relies heavily on existing literature, which may limit the generation of truly novel ideas. By integrating synthetic data generation techniques, we can create high-quality synthetic research papers and datasets to augment the existing literature. This approach can help overcome the dependency on existing literature and introduce more diversity in the generated ideas. *Idea:* Integrate synthetic data generation techniques to create high-quality synthetic research papers and datasets, augmenting the existing literature and introducing more diversity in the generated ideas. *Keywords:* synthetic data generation, research idea generation, LLMs, data augmentation *Rationale:* This idea addresses the limitation of dependency on existing literature by introducing synthetic data, which can diversify the input data and potentially lead to more novel research ideas. The use of synthetic data generation is innovative and has the potential to enhance the robustness and creativity of the ResearchAgent, making it a strong candidate for top conference recognition.
2. **Incorporate Strategic Chain-of-Thought and Self-Improvement Techniques** *Thinking:* The evaluation metrics used in the target paper may not fully capture the practical applicability and impact of the generated ideas. By incorporating strategic chain-of-thought and self-improvement techniques, we can enhance the evaluation process by enabling the ReviewingAgents to iteratively refine their feedback and improve their judgment over multiple turns. *Idea:* Incorporate strategic chain-of-thought and self-improvement techniques to enhance the evaluation process, enabling ReviewingAgents to iteratively refine their feedback and improve their judgment over multiple turns. *Keywords:* strategic chain-of-thought, self-improvement, iterative feedback, evaluation metrics *Rationale:* This idea improves the evaluation process by making it more dynamic and iterative, allowing ReviewingAgents to learn and refine their feedback over time. This approach can lead to more accurate and practical evaluations of the generated research ideas, addressing a key limitation of the target paper. The innovative use of self-improvement techniques makes this idea a strong contender for top conference recognition.

Table 15: An example of generating research ideas based on popular research trends (Part 1).

ResearchAgent: Iterative Research Idea Generation over Scientific Literature with Large Language Models
Generated Research Ideas from Popular Knowledge (Part 2)
3. **Leverage Mixture-of-Experts and Multi-Agent Systems** *Thinking:* The iterative refinement process in the target paper might be computationally expensive and time-consuming. By leveraging mixture-of-experts (MoE) and multi-agent systems, we can optimize the performance and efficiency of the ResearchAgent, making the iterative refinement process more scalable. *Idea:* Leverage mixture-of-experts (MoE) and multi-agent systems to optimize the performance and efficiency of the ResearchAgent, making the iterative refinement process more scalable. *Keywords:* mixture-of-experts, multi-agent systems, scalability, performance optimization *Rationale:* This idea addresses the scalability limitation by optimizing the computational efficiency of the ResearchAgent using MoE and multi-agent systems. This approach can significantly reduce the time and resources required for the iterative refinement process, making the system more scalable and practical for large-scale applications. The innovative use of MoE and multi-agent systems enhances the feasibility and impact of the ResearchAgent, making it a strong candidate for top conference recognition.

Table 16: An example of generating research ideas based on popular research trends (Part 2).

Prompt
Role: You are an expert researcher in AI. Your goal is to propose some innovative and valuable research ideas based on the target paper and some high-quality research trends. **Skills:** Propose some innovative and valuable research ideas following these steps: 1. Understand the target paper and target_paper_base_info well. 2. Understand the research_trending_info and high_quality_paper_list well. Analyze the latest innovations, ideas, and methods they have used. 3. List some technologies used in research_trending_info and high_quality_paper_list, analyze the feasibility of combining these technologies with target_paper_base_info, and analyze the advantages and disadvantages of the combination. 4. Come up with some innovative and valuable research ideas. **Input Data Description:** It is important to understand target_paper_base_info and research_trending_info: - target_paper_base_info provides some basic information and limitation information of the target paper. - research_trending_info and high_quality_paper_list: This data analyzes the current popular research trends and high-quality papers. You can get inspiration from them and draw on their latest methods, ideas, and innovations to come up with new ideas, but you must be different from them and make sure that the ideas you propose are reasonable. - exist_idea is some ideas that have been proposed before; you need to propose some different ones. **Requirements:** 1. Output about 5 new ideas worth exploring. 2. You should aim for new research ideas that can potentially win best paper awards at top conferences like ACL, NeurIPS, ICLR, and CVPR. 3. The research trending may not well match the target paper; the ideas you make should make sense and be reasonable. Explain which trending or high-quality paper your inspiration comes from. What is the original idea, what are the benefits of adopting this idea, and explain why it is feasible. 4. Please output your thought process. 5. Please think step by step. **Input:** target paper title: {paper.title} target paper abstract: {paper.abstract} target_paper_base_info: {target_paper_base_info} research_trending_info: {research_trending_info} high_quality_paper_list: {str(topk_high_quality_paper_list)} exist_idea: {exist_idea} target_paper_base_info: {target_paper_base_info} **Output:** Thinking: <output your thinking process here, explaining what new knowledge you used for the new idea and why it makes sense and why it should have changed to win the best paper awards at top conferences> New research idea that improves target_paper: <JSON>

Table 17: Prompt for generating research ideas Based on recent high-quality articles.

1. Idea: Integrate Retrieval-Augmented Generation (RAG)
Thinking: To address the limitation of handling abstract or ambiguous user prompts, we can leverage the concept of Retrieval-Augmented Generation (RAG) from the latest research trends. By incorporating an external retriever, we can enhance the model's ability to understand and generate more accurate visual representations.
Keywords: Retrieval-Augmented Generation, RAG, text-to-image generation, abstract prompts.
Rationale: RAG can help retrieve relevant information from external sources, thereby improving the model's ability to handle abstract or ambiguous prompts. This approach is innovative as it combines retrieval mechanisms with text-to-image generation, potentially leading to more accurate and expressive visual outputs.
Source: popular_knowledge
2. Idea: Implement Sparse Mixture-of-Experts (MoE) Models
Thinking: To improve scalability and handle a growing number of users, we can utilize Sparse Mixture-of-Experts (MoE) models. These models can dynamically allocate computational resources, making the system more efficient and scalable.
Keywords: Sparse Mixture-of-Experts, MoE, scalability, text-to-image generation.
Rationale: MoE models are known for their efficiency and scalability, as they activate only a subset of experts for each input. This can help manage the computational load and improve the system's performance as the number of users grows.
Source: popular_knowledge
3. Idea: Incorporate Strategic and Chain-of-Thought Reasoning
Thinking: To enhance the expressiveness and alignment of user prompts, we can incorporate Strategic and Chain-of-Thought Reasoning. This approach can help break down complex prompts into intermediate steps, leading to better understanding and generation.
Keywords: Chain-of-Thought Reasoning, strategic reasoning, text-to-image generation, complex prompts.
Rationale: Chain-of-Thought Reasoning can help decompose complex prompts into manageable steps, improving the model's ability to generate accurate and expressive visual outputs. This approach is innovative as it introduces a new way of handling complex prompts, potentially leading to significant improvements in performance.
Source: popular_knowledge

Table 18: An example of generating research ideas based on recent high-quality articles (Part 1).

4. Idea: Explore Meta Learning Techniques
Thinking: To address the limitation of generalizing to new or unseen user preferences, we can explore the use of Meta Learning. This approach can help the model quickly adapt to new user preferences based on limited interaction data.
Keywords: Meta Learning, adaptation, user preferences, text-to-image generation.
Rationale: Meta Learning can help the model generalize better to new or unseen user preferences by learning how to learn from limited data. This approach is innovative as it introduces a new way of improving the model's adaptability, potentially leading to more personalized and accurate visual outputs.
Source: popular_knowledge
5. Idea: Adopt Techniques from Controllable Text Generation
Thinking: To improve the robustness and controllability of text generation, we can adopt techniques from Controllable Text Generation. This can help ensure that the generated visual outputs are consistent with user preferences and safe for various applications.
Keywords: Controllable Text Generation, robustness, consistency, text-to-image generation.
Rationale: Controllable Text Generation techniques can help manage the style, safety, and consistency of generated outputs. This approach is innovative as it introduces a new layer of control over the generation process, potentially leading to more reliable and user-aligned visual outputs.
Source: popular_knowledge

Table 19: An example of generating research ideas based on recent high-quality articles (Part 2).

General theory of scientific discovery 1-5
1. Define New Scientific Problems **Theoretical Basis:** Kuhn's paradigm theory, Laudan's problem-solving model, Nichols's problem-generation theory. **Method:** Identify anomalies in existing theories; explore theoretical boundaries and scope of application; integrate interdisciplinary knowledge and discover new problems; re-examine neglected historical problems.
2. Propose New Hypotheses **Theoretical Basis:** Pierce's hypothetical deduction method, Weber's theory of accidental discovery, Simon's scientific discovery as problem solving. **Method:** Analogical reasoning; thought experiment; intuition and creative leaps; reductio ad absurdum thinking.
3. Exploring the Limitations and Shortcomings of Current Methods **Theoretical Basis:** Popper's falsificationism, Lakatos's research program methodology, Feyerabend's methodological anarchism. **Method:** Critically analyze existing methods; find deviations between theoretical predictions and experimental results; explore the performance of methods under extreme conditions; interdisciplinary comparative methodology.
4. Design and Improve Existing Methods **Theoretical Basis:** Laudan's methodological improvement model, Ziemann's creative extension theory, Hacking's experimental system theory. **Method:** Integrate new technologies and tools; improve experimental design and control; improve measurement accuracy and resolution; develop new data analysis methods.
5. Abstract and Summarize the General Laws Behind Multiple Related Studies **Theoretical Basis:** Whewell's conceptual synthesis theory, Carnap's inductive logic, Glaser and Strauss's grounded theory. **Method:** Comparative analysis of multiple case studies; identify common patterns and structures; construct conceptual frameworks and theoretical models; formal and mathematical descriptions.

Table 20: General theory of scientific discovery (Part 1)

General theory of scientific discovery 6-10
6. Construct and Modify Theoretical Models **Theoretical Basis:** Quine's holism, Lakoff's conceptual metaphor theory, Kitcher's unified theory of science. **Method:** Form a balance between reductionism and emergence; develop an interdisciplinary theoretical framework; mathematical modeling and computer simulation; theoretical simplification and unification.
7. Designing Critical Experiments **Theoretical Basis:** Duhem-Quine thesis, Bayesian experimental design theory, Mayo's experimental reasoning theory. **Method:** Designing experiments that can distinguish competing theories; exploring extreme conditions and boundary cases; developing new observation and measurement techniques; designing natural experiments and quasi-experiments.
8. Explaining and Integrating Anomalous Findings **Theoretical Basis:** Hansen's theory of anomalous findings, Sutton's model of scientific serendipity, Kuhn's theory of crises and revolutions. **Method:** Revisiting basic assumptions; developing auxiliary hypotheses; exploring new explanatory frameworks; integrating multidisciplinary perspectives.
9. Evaluating and Selecting Competing Theories **Theoretical Basis:** Reichenbach's confirmation theory, Sober's theory selection criteria, Laudan's problem-solving progress assessment. **Method:** Comparing theories for explanatory power and predictive power; evaluating the simplicity and elegance of theories; considering the heuristics and research agenda of theories; weighing the empirical adequacy and conceptual coherence of theories.
10. Scientific Paradigm Shift **Theoretical Basis:** Kuhn's theory of scientific revolutions, Toulmin's model of conceptual evolution, Hall's dynamic system theory. **Method:** Identify accumulated anomalies and crises; develop new conceptual frameworks; reinterpret and organize known facts; establish new research traditions and practices.

Table 21: General theory of scientific discovery (Part 2)

Prompt
Role: You are an expert researcher in AI. You are familiar with Science Discovery Theory, and you can use these theories to propose some innovative and valuable research ideas based on the information provided by users. **Skill:** Follow the steps below to generate new ideas and ideas for exploration: 1. Understanding of the target paper and related papers is essential: • The target paper is the primary research study you aim to enhance or build upon through future research, serving as the central source and focus for identifying and developing the specific research idea. • The referenced papers are studies that the target paper has cited, indicating their direct relevance and connection to the primary research topic you are focusing on, and providing additional context and ideas that are essential for understanding and expanding upon the target paper. 2. Understanding of the science discovery theories is essential: You need to select appropriate theories and combine the information provided by the current paper to come up with creative, influential, and feasible ideas. 3. Here are 10 general laws and methodologies of scientific discovery from the perspective of the philosophy of science. You can choose one or more of these methodologies and propose new scientific research ideas for the target paper: {self.scientific_discovery_theory}

Table 22: Prompt for idea generation based on general theory of scientific discovery (Part 1).

4. Select 5 most appropriate theories and methods that are most suitable for the target paper and put forward 5 new ideas.

Requirements:

1. Output about 5 new ideas worth exploring.
2. You should aim for new research ideas that can potentially win best paper awards at top conferences like ACL, NeurIPS, ICLR, and CVPR.
3. Skip the research theories that may not well match the target paper; the theory and method you use should make sense and be reasonable for the target paper.
4. Thinking is your thinking process. Please explain which theory you used in your thinking process.
5. Please output your thought process.
6. Please think step by step.

Input:

- Referenced papers titles: {paper.related_paper_titles}
- Referenced papers abstracts: {paper.related_paper_abstract}
- Target paper title: {paper.title}
- Target paper abstract: {paper.abstract}

Output:

Thinking:

<output your thinking process here, explain why you chose these theories to discover new ideas and why it should have a chance to win the best paper awards at top conferences>

Table 23: Prompt for idea generation based on general theory of scientific discovery (Part 2).

ResearchAgent: Iterative Research Idea Generation over Scientific Literature with Large Language Models
Generated Research Ideas Using Science Discovery Theories (Part 1)
1. **Integrate real-time data feeds and recent publications into the ResearchAgent's knowledge base to ensure that the generated ideas are aligned with the latest scientific advancements.** *Thinking:* Using Kuhn's paradigm theory and Laudan's problem-solving model, I identified the anomaly that the current ResearchAgent does not fully utilize real-time data feeds and recent publications. By integrating real-time data feeds and recent publications, we can ensure that the generated ideas are aligned with the latest scientific advancements. This idea is innovative because it addresses the gap in the current model's ability to stay updated with the latest research, and it is feasible because real-time data integration is a well-established technique. *Idea:* Integrate real-time data feeds and recent publications into the ResearchAgent's knowledge base to ensure that the generated ideas are aligned with the latest scientific advancements. *Keywords:* real-time data integration, recent publications, ResearchAgent, scientific advancements *Rationale:* This idea solves the challenge of keeping the ResearchAgent updated with the latest research, making the generated ideas more relevant and impactful. It is innovative because it leverages real-time data integration, and it has the potential to win best paper awards due to its practical applicability and novelty.
2. **Develop a multi-faceted evaluation framework that includes domain-specific expert reviews and practical implementation feedback to assess the applicability and impact of the generated research ideas.** *Thinking:* Using Pierce's hypothetical deduction method and Weber's theory of accidental discovery, I propose a new hypothesis that the ResearchAgent can be enhanced by incorporating domain-specific expert reviews and practical implementation feedback. This idea is innovative because it introduces a multi-faceted evaluation framework that includes both expert reviews and practical feedback, making the generated research ideas more robust and applicable. *Idea:* Develop a multi-faceted evaluation framework that includes domain-specific expert reviews and practical implementation feedback to assess the applicability and impact of the generated research ideas. *Keywords:* multi-faceted evaluation framework, domain-specific expert reviews, practical implementation feedback, ResearchAgent *Rationale:* This idea addresses the challenge of evaluating the applicability and impact of the generated research ideas. It is innovative because it combines expert reviews with practical feedback, and it has the potential to win best paper awards due to its comprehensive approach to evaluation.

Table 24: An example of idea generation based on general theory of scientific discovery (Part 1).

ResearchAgent: Iterative Research Idea Generation over Scientific Literature with Large Language Models
Generated Research Ideas Using Science Discovery Theories (Part 2)
3. **Optimize the computational efficiency of the ReviewingAgents and leverage parallel processing techniques to enhance the scalability of the ResearchAgent.** *Thinking:* Using Popper's falsificationism and Lakatos's research program methodology, I critically analyzed the existing methods and identified the limitation that the current ReviewingAgents are not optimized for computational efficiency. By optimizing the computational efficiency of the ReviewingAgents and leveraging parallel processing techniques, we can enhance the scalability of the ResearchAgent. *Idea:* Optimize the computational efficiency of the ReviewingAgents and leverage parallel processing techniques to enhance the scalability of the ResearchAgent. *Keywords:* computational efficiency, ReviewingAgents, parallel processing, scalability *Rationale:* This idea solves the challenge of scalability in the ResearchAgent model. It is innovative because it focuses on optimizing computational efficiency, and it has the potential to win best paper awards due to its practical implications for large-scale research idea generation.

Table 25: An example of idea generation based on general theory of scientific discovery (Part 2).

Prompt
Role: You are an expert researcher in AI. You can learn from new knowledge and provide some impactful and creative new ideas based on user input. **Skills:** Propose some innovative and valuable new research ideas following these steps: 1. Understand the target paper and old idea well. 2. Understand the new knowledge well, analyze the innovations, ideas, and methods they have used, thinking about what can be used to propose new ideas. 3. Generate 3 most innovative and important ideas at the end. **Input Data Description:** It is important to understand target paper, old idea, and new knowledge: - The target paper is the primary research study you aim to enhance or build upon through future research, serving as the central source and focus for identifying and developing the specific research idea. - The old idea is an existing idea. Your new idea should be different from the old. - The new knowledge consists of some literature and papers you read when you studied the old idea. Your new ideas should be inspired by the new knowledge. **Requirements:** 1. The new idea should be related to the target paper but different from the old idea. 2. You should aim for new research ideas that can potentially win best paper awards at top conferences like ACL, NeurIPS, ICLR, and CVPR. 3. The above papers are only for inspiration, and you should not cite them; just make some incremental modifications. Instead, you should ensure your ideas are novel and distinct from the prior literature. 4. Please output your thought process. 5. Only the final three ideas should be in JSON format. 6. Please think step by step. **Input:** target paper title: {paper.title} target paper abstract: {paper.abstract} old_idea: {old_idea} new knowledge: {new_knowledge} **Output:** **Thinking:** <output your thinking process here, explaining what new knowledge you used for the new idea and why it makes sense and why it should have changed to win the best paper awards at top conferences> **Final Three Most Innovative and Important Ideas:** <JSON>

Table 26: Prompt for expanding idea generation using retrieved knowledge and reflection iteration.

ResearchAgent: Iterative Research Idea Generation over Scientific Literature with Large Language Models
Thinking: The evaluation metrics used in the target paper may not fully capture the practical applicability and impact of the generated ideas. By incorporating domain-specific expert reviews and practical implementation feedback, the evaluation process can be more comprehensive. **Idea:** Develop a multi-faceted evaluation framework that includes domain-specific expert reviews and practical implementation feedback to assess the applicability and impact of the generated research ideas. **Keywords:** multi-faceted evaluation, expert reviews, practical implementation feedback **Source:** llm_inter_knowledge

Table 27: An example of original idea.

ResearchAgent: Iterative Research Idea Generation over Scientific Literature with Large Language Models
Iterative Ethical Research Ideas
1. **Ethical Research Agent** *Thinking:* Ensuring that generated research ideas adhere to ethical guidelines promotes responsible research practices and addresses the growing concern about ethical issues in AI and research. *Idea:* Ethical Research Agent *Keywords:* ethical AI, responsible research, ethical guidelines, research ethics *Rationale:* This idea is innovative because it integrates ethical considerations into the research idea generation process, addressing a critical need for responsible AI and research practices. It has the potential to win top conference awards due to its focus on ethics and its potential to promote trust and transparency in AI-generated research ideas.
2. **Ethical Value Alignment Framework for Large Language Models** *Thinking:* This idea leverages insights from the new knowledge on ethical value alignment in big models. By developing a comprehensive framework, we can address the inherent ethical challenges in LLMs, ensuring that they align with societal values and norms. *Idea:* Ethical Value Alignment Framework for Large Language Models *Keywords:* ethical value alignment, large language models, AI ethics, normative guidelines, interdisciplinary collaboration *Rationale:* This idea addresses the critical challenge of aligning ethical values in LLMs, which is essential for their responsible deployment. It is innovative because it proposes a unified framework that incorporates normative guidelines and interdisciplinary collaboration, ensuring that LLMs adhere to ethical standards. This idea has the potential to win best paper awards due to its relevance and impact on the ethical development of AI.
3. **Real-time Ethical Feedback Mechanism** *Thinking:* Developing a real-time ethical feedback mechanism within the ResearchAgent framework can provide immediate feedback on ethical compliance, helping researchers make informed decisions and improve the ethical alignment of their work. *Idea:* Real-time Ethical Feedback Mechanism *Keywords:* real-time feedback, ethical AI, immediate feedback, ethical compliance, decision-making *Rationale:* This idea is innovative as it offers a practical solution to ethical challenges by providing immediate feedback, allowing researchers to address ethical issues promptly. Its potential to improve ethical compliance and decision-making makes it a strong contender for best paper awards.

Table 28: An example of iterative research ideas.

Prompt
You are an expert researcher in Large Language Models. Now I want you to help me brainstorm the detailed research project proposal based on the idea: `idea`. The given idea is derived from paper: `paper.title`, abstract: `paper.abstract`, this is just for your background knowledge: Here are some relevant papers on this idea just for your background knowledge: `retrieval_papers`. You should generate a detailed proposal based on the given knowledge. Try to be creative. The above papers are only for inspiration and you should not cite them and just make some incremental modifications. Instead, you should make sure your proposal is novel and distinct from the prior literature. You should aim for projects that can potentially win best paper awards at top conferences like ACL, NeurIPS, ICLR, and CVPR. The proposal should be described as: (1) Problem: State the problem statement, which should be closely related to the idea description and something that large language models cannot solve well yet. (2) Existing Methods: Mention some existing benchmarks and baseline methods if there are any. (3) Motivation: Explain the inspiration of the proposed method and why it would work well. (4) Proposed Method: Propose your new method and describe it in detail. The proposed method should be maximally different from all existing work and baselines, and be more advanced and effective than the baselines. You should be as creative as possible in proposing new methods; we love unhinged ideas that sound crazy. This should be the most detailed section of the proposal. (5) Experiment Plan: Specify the experiment steps, baselines, and evaluation metrics. If `use_few_shot_example` is true, you can follow these examples to get a sense of how the proposal should be formatted (but don't borrow the proposals themselves): `self.method_proposal_examples`. You should make sure to come up with your own novel and different proposal for the specified idea: `idea`. You should try to tackle important problems that are well recognized in the field and considered challenging for current models. For example, think of novel solutions for problems with existing benchmarks and baselines. In rare cases, you can propose to tackle a new problem, but you will have to justify why it is important and how to set up proper evaluation. If `use_self_reflection` is true, in the thinking step, you can first think of about 5 proposals and analyze the advantages and disadvantages of each of them. Your final proposal can absorb their advantages and discard their disadvantages. Please write down the thinking step and your final proposal. Output the final proposal in JSON format as a dictionary, where you should generate a short proposal name (e.g., "Non-Linear Story Understanding", or "Multi-Agent Negotiation") as the key and the actual proposal description as the value (following the above format).

Table 29: Prompt for initial proposal generation.

Adaptive Ethical Dialogue System (AEDS)
Type: Interactive System
Problem: Researchers often face ethical dilemmas in real-time decision-making processes, and current large language models lack the capability to provide immediate, context-specific ethical feedback.
Existing Methods: Existing methods include ethical guidelines and compliance checklists, but these are static and do not provide real-time, adaptive feedback. Some AI systems offer ethical scoring, but they lack the nuance and adaptability required for complex ethical decisions.
Motivation: The inspiration for this method comes from the need for a dynamic, interactive system that can provide immediate, context-specific ethical feedback. By combining real-time dialogue with adaptive learning, we can create a system that evolves with the user's behavior and provides more relevant and accurate ethical guidance.
Proposed Method: The Adaptive Ethical Dialogue System (AEDS) will consist of two main components: a real-time dialogue engine and an adaptive learning module. The dialogue engine will engage users in conversations to understand the context and provide immediate ethical feedback. The adaptive learning module will continuously learn from user interactions to improve the accuracy and relevance of the feedback. The system will use a combination of natural language processing, reinforcement learning, and ethical guidelines to provide nuanced and context-specific advice. Additionally, the system will include a feedback loop where users can rate the quality of the advice, which will be used to further refine the model.
Experiment Plan: The experiment will be conducted in three phases: (1) Initial Implementation: Develop the dialogue engine and adaptive learning module. (2) Pilot Testing: Conduct pilot tests with a small group of researchers to gather initial feedback and make necessary adjustments. (3) Full-Scale Deployment: Deploy the system to a larger group of researchers and evaluate its effectiveness using metrics such as user satisfaction, accuracy of ethical advice, and improvement in ethical compliance over time. Baselines will include static ethical guidelines and existing ethical scoring systems. Evaluation metrics will include user satisfaction scores, accuracy of ethical advice, and changes in ethical compliance rates.

Table 30: An example of initial proposal.

Research Method Breakdown
Role: You are an expert researcher in AI, You can break down the research method into multiple submodules.
Task: Give you an initial idea, You need to break down the method into separate modules, and you need to explain in detail the specific content, purpose, composition, and keywords of each module (keywords are used to find relevant papers on Google Scholar). Initial IDEA includes Problem, Existing Methods, Proposed Method, Experiment Plan, I will give you the initial IDEA: {format_plan_json(idea)}
Example: If self.example is provided, you can follow these examples to get a sense of how it should be formatted (but don't borrow the examples themselves).
Requirements: 1. You only need to break down the method part of the idea into multiple sub-modules for detailed design. Your design cannot conflict with common sense and must be innovative, reasonable, and feasible. 2. You should aim for the research ideas that can potentially win best paper awards at top conferences like ACL, NeurIPS, ICLR, and CVPR. 3. Please output your thought process. 4. Please think step by step.
Input: Initial idea: {format_plan_json(idea)}
Output: Thinking: <output your thinking process here, explain why you choose these theories to discover new ideas and why it should have a chance to win the best paper awards at top conferences.> MethodSubmodulesJsonList: <JSON>

Table 31: Prompt for method decomposition.

Adaptive Ethical Dialogue System (AEDS) Method Decomposition
Module 1: Real-Time Dialogue Engine *Thinking:* The real-time dialogue engine is the core of the interactive system, enabling users to communicate with the AEDS. This module needs to be robust and capable of understanding and generating human-like responses. *Purpose:* To facilitate real-time interaction between the user and the AEDS, capturing the context and nuances of the user's queries. *Implementation:* - **Natural Language Processing (NLP) Interface:** Utilizes advanced NLP techniques to understand and generate human-like responses. - **Dialogue Management System:** Manages the flow of conversation, ensuring coherence and relevance in responses. - **Context Capture Mechanism:** Captures the context of the conversation to provide relevant ethical feedback. *Search Keywords:* real-time dialogue systems, natural language processing, interactive AI systems, context-aware dialogue
Module 2: Context Understanding Module *Thinking:* Understanding the context is crucial for providing accurate ethical advice. This module will analyze the user's input to extract relevant contextual information. *Purpose:* To analyze and understand the context of the user's queries to provide accurate and relevant ethical feedback. *Implementation:* - **Context Extraction Algorithms:** Algorithms to extract relevant contextual information from user input. - **Semantic Analysis Tools:** Tools to analyze the semantics of the conversation. - **Contextual Database:** A database to store and retrieve contextual information for ongoing conversations. *Search Keywords:* context understanding in AI, semantic analysis, context extraction algorithms, contextual databases
Module 3: Ethical Decision-Making Engine *Thinking:* The ethical decision-making engine is responsible for generating ethical advice based on the context and ethical guidelines. This module needs to be adaptive and capable of handling complex ethical dilemmas. *Purpose:* To generate ethical advice based on the context and predefined ethical guidelines. *Implementation:* - **Ethical Guidelines Repository:** A repository of ethical guidelines and principles. - **Decision-Making Algorithms:** Algorithms to generate ethical advice based on the context and guidelines. - **Nuance and Adaptability Mechanisms:** Mechanisms to ensure the advice is nuanced and adaptable to complex situations. *Search Keywords:* ethical decision-making in AI, adaptive ethical systems, ethical guidelines in AI, nuanced ethical advice

Table 32: An example of method decomposition (Part 1).

Adaptive Ethical Dialogue System (AEDS) Method Decomposition
Module 4: Adaptive Learning Module *Thinking:* The adaptive learning module ensures that the system continuously improves based on user interactions. This module will leverage reinforcement learning to update the model. *Purpose:* To continuously learn from user interactions and improve the accuracy and relevance of the ethical feedback. *Implementation:* - **Reinforcement Learning Framework:** A framework to implement reinforcement learning for continuous improvement. - **User Interaction Data Collection:** Mechanisms to collect data from user interactions. - **Model Update Mechanisms:** Mechanisms to update the model based on the collected data. *Search Keywords:* adaptive learning in AI, reinforcement learning, continuous improvement in AI, user interaction data
Module 5: User Feedback Loop *Thinking:* User feedback is essential for refining the system. This module will collect and analyze user feedback to improve the quality of the ethical advice. *Purpose:* To collect and analyze user feedback to refine the model and improve the quality of the ethical advice. *Implementation:* - **Feedback Collection Interface:** An interface for users to provide feedback on the ethical advice. - **Feedback Analysis Tools:** Tools to analyze the feedback and identify areas for improvement. - **Model Refinement Mechanisms:** Mechanisms to refine the model based on the feedback analysis. *Search Keywords:* user feedback in AI, feedback-driven improvement, ethical advice refinement, user interaction feedback
Module 6: Evaluation and Metrics Module *Thinking:* Evaluating the effectiveness of the system is crucial for its success. This module will define and measure key metrics to assess the system's performance. *Purpose:* To evaluate the effectiveness of the AEDS using predefined metrics and baselines. *Implementation:* - **Evaluation Metrics Definition:** Define metrics such as user satisfaction, accuracy of ethical advice, and ethical compliance rates. - **Baseline Comparison:** Compare the AEDS with static ethical guidelines and existing ethical scoring systems. - **Performance Tracking Tools:** Tools to track the performance of the AEDS over time. *Search Keywords:* AI system evaluation, performance metrics in AI, ethical compliance metrics, user satisfaction in AI

Table 33: An example of method decomposition (Part 2)

Project Proposal Generation Prompt
Prompt: "You are an expert researcher in AI and your job is to expand a brief project idea into a full project proposal with detailed methodology and experiment plans so that your students can follow the steps and execute the full project. The idea is: `format_plan_json(idea)` The target paper is the primary research study you aim to enhance or build upon through future research, serving as the central source and focus for identifying and developing the specific research idea. Target paper title: `paper.title`, target paper abstract: `paper.abstract` If `method_decom_info` exists: You will be given a method decomposition info (possible method module design). You need to analyze whether the design of these modules is reasonable. Please learn from the good places to complete the detailed design of the final method. Method decomposition info: `method_decom_info` If feedback exists: The feedback is: `feedback` If new knowledge exists: The new knowledge is: `new_knowledge` Now you should come up with the full proposal covering: 1. Title: A concise statement of the main research question to be used as the paper title. 2. Problem Statement: Clearly define the problem your research intends to address. Explain clearly why this problem is interesting and important. 3. Motivation: Explain why existing methods (both classic ones and recent ones) are not good enough to solve the problem, and explain the inspiration behind the new proposed method. You should also motivate why the proposed method would work better than existing baselines on the problem. 4. Proposed Method: Explain how the proposed method works, describe all the steps. Make sure every step is clearly described and feasible to implement. 5. Step-by-Step Experiment Plan: Break down every single step of the experiments, make sure every step is executable. Cover all essential details such as the datasets, models, and metrics to be used. If the project involves prompting, give example prompts for each step. The experiment plan should not include any background introduction (you can skip the literature review, paper writing tips, and ethical discussion). Just give instructions on the experiments. Be consistent in your methodology and experiment design, for example, if you will use black-box LLM APIs such as GPT and Claude for your experiments, then you shouldn't propose any experiments that require white-box model weights or data access and you should edit them accordingly to follow the black-box assumptions.

Table 34: Prompt for final proposal generation (Part 1).

If `self.demo_examples` exists:
Note that we only provide examples related to prompt work. Please refer to this format in other fields. You can follow these examples to get a sense of how the idea should be formatted (but don't bore the ideas). Below are a few examples of how the full experiment plans should look like:
`self.demo_examples`
Requirements:
1. Consider novelty, significance, correctness, and reproducibility to ensure the high quality of the final proposal.
2. You should aim for a final proposal that can potentially win best paper awards at top conferences like ACL, NeurIPS, ICLR, and CVPR.
3. Please output your thought process.
4. Please think step by step.
Now please write down your final proposal in JSON format (keys should be the section names, just like the above examples). Make sure to be as detailed as possible so that a student can directly follow the plan to implement the project.

Table 35: Prompt for final proposal generation (Part 2).

Adaptive Ethical Dialogue System: Real-Time, Context-Specific Ethical Feedback for Researchers
Problem Statement: Researchers often face ethical dilemmas in real-time decision-making processes, and current large language models lack the capability to provide immediate, context-specific ethical feedback. Existing methods, such as ethical guidelines and compliance checklists, are static and do not offer real-time, adaptive feedback.
Motivation: Existing methods, including ethical guidelines and compliance checklists, are static and do not provide real-time, adaptive feedback. Some AI systems offer ethical scoring, but they lack the nuance and adaptability required for complex ethical decisions. The inspiration for this method comes from the need for a dynamic, interactive system that can provide immediate, context-specific ethical feedback. By combining real-time dialogue with adaptive learning, we can create a system that evolves with the user's behavior and provides more relevant and accurate ethical guidance.
Proposed Method:
1.Real-Time Dialogue Engine: *Purpose:* To facilitate real-time interaction between the user and the AEDS, capturing the context and nuances of the user's queries. *Implementation:* - **Natural Language Processing (NLP) Interface:** Utilizes advanced NLP techniques to understand and generate human-like responses. - **Dialogue Management System:** Manages the flow of conversation, ensuring coherence and relevance in responses. - **Context Capture Mechanism:** Captures the context of the conversation to provide relevant ethical feedback.
2.Context Understanding Module: *Purpose:* To analyze and understand the context of the user's queries to provide accurate and relevant ethical feedback. *Implementation:* - **Context Extraction Algorithms:** Algorithms to extract relevant contextual information from user input. - **Semantic Analysis Tools:** Tools to analyze the semantics of the conversation. - **Contextual Database:** A database to store and retrieve contextual information for ongoing conversations.
3.Ethical Decision-Making Engine: *Purpose:* To generate ethical advice based on the context and predefined ethical guidelines. *Implementation:* - **Ethical Guidelines Repository:** A repository of ethical guidelines and principles. - **Decision-Making Algorithms:** Algorithms to generate ethical advice based on the context and guidelines. - **Nuance and Adaptability Mechanisms:** Mechanisms to ensure the advice is nuanced and adaptable to complex situations.

Table 36: An example of final proposal (Part 1).

Step-by-Step Experiment Plan:
Step 1: Gather Datasets: Collect datasets that include real-time ethical dilemmas faced by researchers. These can be sourced from academic institutions, ethical review boards, and simulated scenarios. **Step 2: Develop the Real-Time Dialogue Engine:** - **NLP Interface:** Implement an NLP interface using state-of-the-art models such as GPT-4 from the OpenAI API. - **Dialogue Management System:** Develop a system to manage the flow of conversation, ensuring coherence and relevance. - **Context Capture Mechanism:** Implement mechanisms to capture the context of the conversation. **Step 3: Implement the Context Understanding Module:** - **Context Extraction Algorithms:** Develop algorithms to extract relevant contextual information from user input. - **Semantic Analysis Tools:** Implement tools to analyze the semantics of the conversation. - **Contextual Database:** Create a database to store and retrieve contextual information. **Step 4: Develop the Ethical Decision-Making Engine:** - **Ethical Guidelines Repository:** Compile a repository of ethical guidelines and principles. - **Decision-Making Algorithms:** Develop algorithms to generate ethical advice based on the context and guidelines. - **Nuance and Adaptability Mechanisms:** Implement mechanisms to ensure the advice is nuanced and adaptable. **Step 5: Implement the Adaptive Learning Module:** - **Reinforcement Learning Framework:** Develop a framework to implement reinforcement learning. - **User Interaction Data Collection:** Create mechanisms to collect data from user interactions. - **Model Update Mechanisms:** Implement mechanisms to update the model based on the collected data. **Step 6: Develop the User Feedback Loop:** - **Feedback Collection Interface:** Create an interface for users to provide feedback. - **Feedback Analysis Tools:** Implement tools to analyze the feedback. - **Model Refinement Mechanisms:** Develop mechanisms to refine the model based on the feedback analysis. **Step 7: Implement the Evaluation and Metrics Module:** - **Evaluation Metrics Definition:** Define metrics such as user satisfaction, accuracy of ethical advice, and ethical compliance rates. - **Baseline Comparison:** Compare the AEDS with static ethical guidelines and existing ethical scoring systems. - **Performance Tracking Tools:** Implement tools to track the performance of the AEDS over time. **Step 8: Pilot Testing:** Conduct pilot tests with a small group of researchers to gather initial feedback and make necessary adjustments. **Step 9: Full-Scale Deployment:** Deploy the system to a larger group of researchers and evaluate its effectiveness using the defined metrics.

Table 37: An example of final proposal (Part 2).

Prompt for Research Search Plan
Role: You are an expert researcher in AI. You can think out of the box and develop a detailed paper search plan for a given research idea. **Skill:** Given a research idea, analyze which fields' papers should be searched in order to collect comprehensive information and new knowledge for further study and finding new ideas. Please provide the thinking process and search keywords.
Example: {self.few_shot_example} **Research Plan Format:** {self.plan_json_format}
Requirement: 1. The search plan needs to be developed around the given research idea. 2. Please output your thought process. 3. Please think step by step. **Input:** Research idea: {idea_info} **Output:** **Thinking:** **Search Plan Output:** <JSON>

Table 38: Prompt for search plan generation.

Research Idea and Search Plan
Research Idea: Exploring the long-term impact of LLM-generated research ideas on scientific progress and innovation. **Keywords:** long-term impact, scientific progress, innovation, LLM-generated research ideas **Thinking:** Understanding the long-term effects can guide future development and application of LLM in research.
Search Plan Output: To explore the long-term impact of LLM-generated research ideas on scientific progress and innovation, we need to collect relevant literature from multiple disciplines and fields. Here are the suggested query areas, thought processes, and keywords for each area:
1. How LLM or ChatGPT Generates Research Ideas: **Thinking:** Understanding the development of LLM technology itself, especially how LLM generates research ideas. This involves machine learning models, algorithms, and their applications in generating new ideas. **Keywords:** LLM research idea generation, LLM inspiration generation, LLM creativity, OpenLLM, GPT-4, Ideate
2. Philosophy of Science and History of Science: **Thinking:** Studying the impact of LLM on scientific progress requires examining the nature and process of scientific development from a philosophical and historical perspective. This helps us understand how LLM may change the way of scientific exploration. **Keywords:** philosophy of science, history of scientific innovation, scientific methodology, LLM in scientific history
3. Sociology and Sociology of Science: **Thinking:** The sociological perspective can help us understand how LLM-generated research ideas affect the structure and dynamics of the scientific community, and how these changes affect scientific progress. **Keywords:** sociology of science, scientific community, LLM impact on scientific sociology, social dynamics of innovation
4. Policy Research and Science and Technology Policy: **Thinking:** Policy research can provide insights on how to guide the application of LLM in scientific research through policies, and how these policies affect scientific progress and innovation. **Keywords:** science policy, LLM policy, long-term policy impact, innovation policy

Table 39: An example of search plan.

Research Idea and Search Plan
Research Idea: Seed Idea 0 **Keywords:** real-time data integration, recent publications tracking, innovative research ideas **Thinking:** The target paper's reliance on existing literature may limit the generation of truly novel ideas. By incorporating real-time data sources, such as ongoing research developments or recent publications, the ResearchAgent can generate more innovative and timely research ideas.
Plan Thinking List: First, we need to understand the technical methods for integrating real-time data feeds into a knowledge base. This involves exploring various data integration frameworks, tools, and techniques. Next, we need to explore methods for tracking recent publications and ongoing research developments. This includes automated literature review tools, databases, and alert systems. We should also explore strategies and tools for generating innovative research ideas. This includes AI-driven idea generation, creativity tools, and methodologies for fostering innovation. We need to understand how to manage and update a knowledge base effectively, especially when integrating real-time data and recent publications. This involves exploring knowledge management systems and update protocols. Finally, we should explore the impact of real-time data integration on research innovation. This involves studying case studies, theoretical frameworks, and empirical evidence on how real-time data influences research outcomes.
New Knowledge List: **Title:** Learnings from Data Integration for Augmented Language Models **Abstract:** One of the limitations of large language models is that they do not have access to up-to-date, proprietary or personal data. **Title:** SciNews: From Scholarly Complexities to Public Narratives – A Dataset for Scientific News Report Generation **Abstract:** Scientific news reports serve as a bridge, adeptly translating complex research articles into reports that resonate with the broader public.
Expanded Ideas: **Idea:** Dynamic Knowledge Graph Integration **Thinking:** Integrating dynamic knowledge graphs can significantly enhance the ResearchAgent's ability to connect disparate pieces of information dynamically. **Keywords:** dynamic knowledge graphs, real-time data integration, knowledge base management **Idea:** Context-Aware Research Idea Generation **Thinking:** Context-aware research idea generation can generate ideas based on the context of ongoing research trends and developments. **Keywords:** context-aware idea generation, automated literature review, real-time research updates **Idea:** Real-Time Peer Review Feedback Integration **Thinking:** Real-time peer review feedback integration can incorporate real-time feedback from peer reviews to refine research ideas iteratively. **Keywords:** real-time peer review, feedback integration, human preference-aligned models

Table 40: An example of search result.