

Analysis of Accelerating Ideation Process with Large Language Model

Yuhan Wei

Applied Maths / Computer Science Track, Duke Kunshan University, Suzhou, China

yw641@duke.edu

Abstract. Large Language Models (LLMs) have rapidly advanced, demonstrating near-human intelligence in language comprehension, problem-solving, and autonomous decision-making, which can be leveraged in scientific discovery. Previous research of LLM for science often focuses on experiment execution part and the first stage of research, i.e., idea generation with large language models is lack of research. This paper fills this gap by surveying the use of LLM-based agent systems in research idea generation. This research first proposes three stages in research idea generation pipeline, pre-ideation (knowledge preparation), ideation (generation and iteration), post-ideation (evaluation), and provide detailed summary in different methods in each stage, then it summarizes different metrics and angles of evaluating generated research ideas, finally it discusses limitations and ethical concerns of existing works and suggest potential solutions and future directions. These results aim to provide a solid foundation for future research in improving LLM-based ideation systems and fostering responsible AI usage in scientific discovery.

Keywords: LLM-based agent, research idea generation, scientific discovery, multi-agent collaboration.

1. Introduction

Large language models (LLMs) have rapidly advanced, demonstrating capabilities that approach human-level intelligence. Trained on massive datasets, they have been applied across fields such as language comprehension, problem-solving, and autonomous decision-making. These developments have led to the rise of autonomous agents—systems powered by LLMs that can perform complex tasks like planning and interacting with their environment without human intervention [1, 2]. An example is the generative AI town simulation, where multiple agents simulate human behaviors and decision-making processes. This highlights the potential of LLMs for broader applications, particularly in scientific discovery, where their ability to process vast amounts of data and generate insights is promising.

Existing research on LLMs in scientific fields has largely focused on accelerating experimental processes [3, 4]. In chemistry and drug discovery, LLMs have been used to design experiments, automate data analysis, and simulate results, significantly speeding up research. In biology and physics, these models assist in generating predictions and models, transforming how experiments are conducted and allowing for faster research cycles. A major contribution in this domain is the integration of knowledge graphs with LLMs to link scientific data and identify relationships that are not immediately

obvious. Literature-based discovery (LBD) techniques have been used to uncover connections between disparate knowledge, making LLMs valuable tools in automating literature reviews, hypothesis generation, and experiment validation. Some research details how these models can process large volumes of scientific literature, summarize key findings, and suggest new research directions, improving researchers' productivity [3]. Despite these advances, the use of LLMs in the ideation process—the generation of novel research ideas—remains underexplored. Most current works focus on experiment design and hypothesis validation rather than the early stages of scientific inquiry. This presents a significant opportunity to expand the role of LLMs in ideation, helping researchers formulate innovative research questions.

Generating new ideas is critical to scientific discovery, requiring not only comprehensive literature review but also the ability to synthesize information into novel hypotheses. LLMs, with their capacity to process large datasets and identify patterns across disciplines, offer a promising solution. They can help researchers uncover new research directions and identify gaps in existing knowledge that might otherwise be overlooked. LLMs have the potential to assist significantly in this ideation process. By handling large amounts of information, they can identify research opportunities and suggest areas for further exploration that would be difficult for individual researchers to pinpoint. The ability of LLMs to integrate external knowledge from various disciplines also enhances their capacity for generating interdisciplinary research ideas, an increasingly important aspect of modern scientific breakthroughs.

This paper will talk about the advancements in LLM-based autonomous systems for research idea generation. It will explore how these systems leverage key components like profiling, memory, planning, and action modules to generate novel ideas, and how methods such as verbal self-reflection, multi-agent debates, and role-playing can enhance both creativity and factual accuracy in ideation processes. The paper will further examine the stages of the research ideation pipeline, from knowledge preparation to post-ideation evaluation, and discuss the role of human-AI collaboration in improving the quality of research ideas. Finally, it will address the current limitations, including repetitive idea generation and ethical challenges, offering insights into future directions for refining LLM-based research systems.

2. Autonomous agent systems in research idea generation

In recent years, achievements in LLMs have shown great promise in reaching human level intelligence. Existing studies have also leveraged this intelligence to build LLM-based agents, which are expected to perform diverse tasks [1]. A typical autonomous LLM agent consists of several essential components: a profiling module, a memory module, a planning module, and an action module. The profiling module allows the agent to understand the task context and user preferences. The memory module stores and retrieves information from past interactions, enabling the agent to learn and improve over time. The planning module is responsible for decision-making, helping the agent figure out how to approach and complete tasks. Finally, the action module enables the agent to execute tasks, such as generating responses, performing web search or interacting with external environments [2].

With ability to understand and generate natural language, as well as make autonomous decisions, LLM-based agents can not only process complex information but also independently complete tasks through built-in memory modules and planning capabilities. These agents are able to perform reasoning and generate new ideas by analyzing vast datasets and utilizing their own integrated knowledge, demonstrating human-like abilities in problem-solving and creative idea generation. In addition, when LLMs are leveraged in autonomous system to complete tasks, methods to enhance the factuality and creativity are explored [5-7], empowering their use in idea generation. Reflexion introduces verbal self-reflection, allowing agents to learn from their previous actions through natural language feedback stored in memory and continuously improve their performance [5]. LLM Debate leverages a multi-agent debate framework where each agent presents its arguments in favor or against a solution, while a judge evaluates the strength of their claims [6]. This improves the factuality of LLM generation by encouraging divergent thinking and mitigating the Degeneration of Thought (DoT) problem, which can occur when a single model sticks to an incorrect answer due to overconfidence. LLM Discussion builds a multi-agent discussion framework which involves initiation, discussion, and convergence stages to

stimulate both divergent and convergent thinking [7]. The role-playing technique further boosts creativity of the system by assigning diverse roles to each LLM, ensuring that the agents bring different perspectives and approaches to the discussion.

These advancements in LLM-based autonomous systems—through verbal self-reflection, debate, and multi-agent discussion frameworks—significantly improve both the factuality and creativity of research idea generation. By leveraging these methods, LLMs show promise to autonomously generate novel and accurate research ideas. Building on this foundation, recent scientific ideation studies take these innovations a step further, diving into agent systems to generate scientific hypotheses, explore new research directions, evaluate and iteratively refine these research ideas.

3. Scientific ideation pipeline

Similar to the process of human researchers, ideation pipeline for LLM-based agents also consists of three stages as shown in Fig.1. Pre-ideation stage starts from taking in knowledge and reading large amounts of papers and formalize this knowledge into knowledge base. Ideation stage requires researchers to find unexplored potential relationship between knowledge or get inspiration by some of the papers and initiate query to form new research ideas. During the ideation process, review and reflect on the proposed ideas, search the previous work to ensure its originality and iterate them to make them better are also important. Post-ideation includes more evaluations like further validation on novelty, assessment on feasibility and check on human-alignment. This section provides a concise summary of previous works in each stage.

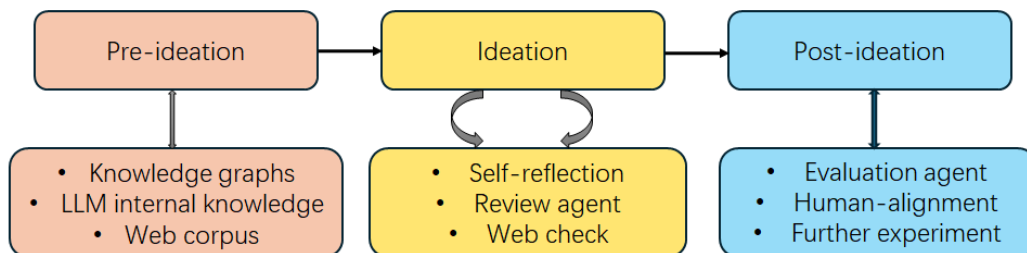


Figure 1. The Research Ideation Pipeline Using LLM-based Agents (Photo/Picture credit: Original).

When doing scientific research, abundant knowledge input and comprehensive literature review process are usually crucial inductive for an innovative and compelling idea. While there are several works trying to propose new research ideas and do scientific discovery simply based on text mining on large datasets of scientific literature [8, 9] or link prediction in knowledge graph [10-12], these ideas have less flexibility and are limited by the relationship between entities [13, 14]. Several studies start to leverage the internal scientific knowledge in large language models and combine few shot examples or a set of relevance papers in the prompt serving for the literature review purpose [15-18]. Another study utilizes pure web corpus instead of manually selected paper sets, targeting at a more open-domain setting [15].

Based on the knowledge graph, there are several ideation methods build models to discover potential research problems without using large language models. For example, the ComplEx model formalized hypothesis prediction task as link prediction and generate new hypotheses by predicting potential links in complex, multi-relational data [10]. In chemistry or drug discovery, a research uses Generative Flow Net (GFlowNet) to generate high-quality hypothesis on molecules satisfying some criteria [16-25]. In the biomedical field, frameworks (e.g., ExEMPLAR [26]) use knowledge graphs to link heterogeneous data, helping researchers to reason over complex datasets and generate novel hypotheses. Simply based on the relationship between entities, the research ideas generated seem to be limited and less capable in exploration of more open-ended research directions. This part will focus on reviewing LLM-based

research ideation framework. Ideation process with large language models can be formalized as input the prompt into large language models and get the output ideas:

$$o = \text{LLM} (T (\{l_0, l_1, \dots, l_n\})) \quad (1)$$

In term of the prompt formats and what is incorporated in the prompt, several works have explored different methods and frameworks to improve the quality of the generated ideas. One popular way of prompt used in both experiments in zero shot [17] and AI scientist [18] is input all the seeds ideas and their experiment details. These seeds ideas are records of past generated ideas which serves as inspiration for the new round of idea generation. With all the seeds ideas, new ideation is kept relevant with the intended topic and can be adaptation and improvement of the previous ones. The work zero shot [17] also compares the quality of ideas generated by few-shot and zero-shot prompt. It finds that increasing uncertainty and model's space through zero-shot prompt enhances model's proficiency in generating hypotheses.

Another classical way of designing prompt is shown in ResearchAgent [14]. Like human researchers often get inspiration from one certain research paper as well as checking its relevant papers, ResearchAgent starts with a single targeted paper and also incorporates relevant papers according to their reference list in the prompt. In order to encourage more interdisciplinary research ideas and not limited by the narrow scope of the set of relevant papers, ResearchAgent also leverages external entity-centric knowledge store and retrieves top-k relevant entities besides the existing literature set. Adding these entities to prompts helps to generate more interdisciplinary ideas and tackle open-ended scenarios in scientific discovery. In term of iteration process which keeps improving ideas, previous studies have also designed different methods or built various multi-agent collaboration frameworks to enhance the proficiency in generating research ideas. Zero shot [17] designs a role-play framework in the hypothesis generation loop, entailing an analyst who interprets and extracts key words from literatures, an engineer who performs search based on the key words and compiles and structures the findings, a scientist who offers fresh interpretation and crafts hypotheses based on the findings, and a critic scrutinizing hypotheses. In addition, the framework also implements tool use methods like ReAct [19] in the engineer part, allowing the engineer agent to think before taking action, do the planning and make observations on the feedback from the environment. In ResearchAgent [14], the multi-agent collaboration occurs primarily through the interaction between research agents and reviewing agents. Three research agents responsible for problem identification, method development and experiment design have a reviewing agent for each process respectively. Based on the feedback of the reviewing agent, each research agent iteratively refines its proposal several times before their proposal is combined into the prompt for the research agent in the next stage. SciMON retrieves inspiration from related prior literature, including semantic neighbors, knowledge graph neighbors, and citation neighbors [13]. These inspirations are combined into the prompt to the large language model to generate an initial hypothesis grounded in previous work. SciMON then compares it against existing research to assess its novelty. If the hypothesis is too similar to existing ideas, the system prompts the model to revise and enhance the hypothesis to be more novel. The iteration process stops when the hypothesis reaches a sufficient level of novelty. The AI scientist builds a complete scientific research pipeline more than just idea generation [18]. In the idea generation stage, the AI scientist uses multiple rounds of chain-of-thought [20] and self-reflection [5] to refine each idea. It also checks the similarity of the idea with existing studies using tools like the Semantic Scholar API and web access [21]. After the whole pipeline of generating a paper, the AI scientist leverages a reviewing agent to review the paper, and the feedback can be given back to the idea generation agent for refinement.

Besides the autonomous agent framework, human-AI collaboration framework is another option for leveraging large language models to accelerate research ideation process. In the CoQuest system [22], the user first provides an initial idea or a broad topic for AI and then AI can generate research ideas through the breadth-first generation approach for the user to explore multiple research directions at the same time and be more creative or through depth-first generation for more innovative ideation. The AI Thoughts panel integrates a Paper Graph Visualizer which allows users to explore relevant academic

papers related to the generated research questions, ensuring that they are grounded in existing research. By retrieving and visualizing related literature, this design also helps users to understand the research landscape and identify potential gaps. The panel also offers explanations of why the AI generated specific research questions, giving transparency to the reasoning behind RQ generation, improving users' trust and encouraging them to refine their questions further. The user can provide feedback on the generated questions and AI will refine and regenerate new research ideas. A research idea can be evaluated from different angles as shown in Table 1. Various evaluation metrics are designed, but they usually contain the following angles: novelty, significance and feasibility. Ideas' relevance to one topic is also assessed [14, 17]. In addition, clarity also serves as one evaluation criteria [14, 18]. Human researchers score interestingness of ideas [23].

Wide-accepted angles human accessing ideas and research works can be referred in the conferences' review guidelines. For example, in [18], the reviewing agents need to provide reviews based on NeurIPS review guidelines. In these autonomous agent system with reviewing agents, the alignment between agent evaluations and those of human experts is frequently assessed by checking the correlation of the two scores [14, 17, 18].

Observations on evaluation scores provide new insights for metrics designing. One observation indicates that citation of one concept can be used to evaluate the creativity of one idea: the more frequently a concept has been cited, the less interesting the research projects are evaluated [13]. Several studies also work on predict the impact of new papers or ideas. A study leverages knowledge graph and concept pairs to assess the impact of one idea [13]. Another study finetunes the large language model on Topic Normalized Citation Success Index for the same period, allowing the model to make impact prediction based only on the content of the paper instead of the external information [24]. Others provide a comprehensive higher-level assessment metrics [25]. This assessment rubric evaluates an idea from five angles: idea quality, idea space, impact of ideas on users, social acceptance, human alignment. Novelty, originality, feasibility, elaboration and completeness all fall into the metrics of idea quality. From the angle of idea space, quantity, diversity, evenness and depth are evaluated, which are slightly overlap with metrics in idea quality angle. From ideas' impact and users' perspective, metrics from computer human interaction is included like surprise, usefulness, motivational, task influence. Social acceptance angle evaluates on acceptance, appropriateness, value, realistic and flexibility. Human Alignment angle evaluates on relevance, elaboration and fluency.

Table 1. Comparison of Autonomous Agent Systems for Ideation Enhancement

Idea Quality	Novelty
	Originality
	Creativity
	Relevance
	Feasibility
Idea Space	Quantity
	Diversity
	Depth
Idea Impact on Users	Surprise
	Importance
	Human-alignment
	Usefulness

4. Limitations and prospects

While the advancements in LLM-based systems for research idea generation demonstrate significant potential, limitations and ethical concerns must be considered. One limitation is that similar ideas are found generated during multiple iterations. This may because large language models heavily rely on existing data and knowledge, and in multiple times of similar prompts, large language models are constraint into a narrow domain and may produce research ideas that lack novelty. Future work can

focus on improving the ability of LLMs to generate more diverse and novel research ideas. This could be achieved by incorporating more dynamic prompts, leveraging broader datasets, and integrating mechanisms that encourage interdisciplinary thinking. Degeneration of Thought problem, when large language models are overconfident about their reply, may also result in this repetitive idea generation problem. Future studies may work on the homogeneity problem of large language models by incorporating multi-agent role-play and conversation, which leverages group intelligence to increase the creativity of the system.

Better incorporating Human-AI Collaboration into agent system can also be a direction of improving the idea quality. Combining the strengths of human intuition and expertise with LLMs can help to overcome some of the limitations in autonomous idea generation. Human-AI collaboration frameworks, such as interactive systems where humans provide feedback to refine and adjust the generated ideas, can enhance both the creativity and feasibility of research outcomes. The ease with which LLMs can generate ideas at scale introduces ethical concerns and the possibility of misuse. For instance, these systems could be exploited to flood academic forums with low-quality or superficially novel ideas, potentially overwhelming peer review systems. Moreover, the lack of a solid ethical framework around the use of LLMs in research could result in the propagation of biased, harmful, or dangerous research ideas. Ensuring ethical compliance into the ideation system is also an important future direction.

5. Conclusion

To sum up, this study provides a comprehensive review of LLM-based research ideation systems, systematically examining the full pipeline from pre-ideation knowledge preparation, through ideation and iteration, to post-ideation evaluation. It categorized different approaches, such as knowledge graph utilization, prompt design techniques, and multi-agent collaboration frameworks, demonstrating how these methods contribute to improving quality of the generated research ideas. Additionally, this research highlighted key mechanisms like verbal self-reflection, multi-agent debate, and role-playing techniques, which have been employed to further enhance idea generation by LLMs. Despite these advancements, challenges remain. Repetitive idea generation and ethical concerns surrounding the use of LLMs in autonomous research present significant limitations. Addressing these issues will require more dynamic prompting strategies, interdisciplinary collaboration, and robust ethical frameworks to ensure responsible AI usage in research settings. These results aim to provide a solid foundation for understanding the current capabilities and limitations of LLM-based ideation systems. It is hoped that this work will inspire future research in improving these systems, particularly in fostering greater novelty, promoting ethical considerations, and advancing human-AI collaboration in scientific discovery. By tackling these challenges, LLMs can become even more effective tools for accelerating research and generating innovative ideas.

References

- [1] Wang L, Ma C, Feng X, et al 2024 A survey on large language model based autonomous agents. *Frontiers of Computer Science* vol 18(6) p 186345.
- [2] Xi Z, Chen W, Guo X, et al 2023 The rise and potential of large language model based agents: A survey arXiv: 230907864
- [3] Zhang Y, Chen X, Jin B, Wang S, Ji S, Wang W and Han J 2024 A comprehensive survey of scientific large language models and their applications in scientific discovery arXiv: 240610833
- [4] Boiko D A, MacKnight R and Gomes G 2023 Emergent autonomous scientific research capabilities of large language models arXiv: 230405332
- [5] Shinn N, Cassano F, Gopinath A, Narasimhan K and Yao S 2023 Reflexion: Language agents with verbal reinforcement learning 37th Conference on Neural Information Processing Systems NeurIPS p 12.
- [6] Liang T, He Z, Jiao W, et al 2023 Encouraging divergent thinking in large language models through multi-agent debate arXiv: 230519118

- [7] Lu L C, Chen S J, Pai T M, et al 2024 LLM discussion: Enhancing the creativity of large language models via discussion framework and role-play arXiv: 240506373
- [8] Srinivasan P 2003 Text mining: Generating hypotheses from MEDLINE Journal of the American Society for Information Science and Technology vol 55(5) pp 396–413
- [9] Spangler S, Wilkins A D, Bachman B J, et al 2014 Automated hypothesis generation based on mining scientific literature Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining pp. 1877-1886.
- [10] de Haan R, Tiddi I and Beek W 2021 Discovering research hypotheses in social science using knowledge graph embeddings The Semantic Web ESWC 2021 Lecture Notes in Computer Science (Vol 12731) Springer Cham
- [11] Liu N F, Lin K, Hewitt J, Paranjape A, Bevilacqua M, Petroni F and Liang P 2023 Lost in the middle: How language models use long contexts Transactions of the Association for Computational Linguistics vol 12 pp 157–173
- [12] Nadkarni R K, Wadden D, Beltagy I, Smith N A, Hajishirzi H and Hope T 2021 Scientific language models for biomedical knowledge base completion: An empirical study arXiv: 210609700
- [13] Wang Q, Downey D, Ji H and Hope T 2023 SCIMON: Scientific inspiration machines optimized for novelty arXiv preprint arXiv: 230514259
- [14] Baek J, Jauhar S K, Cucerzan S and Hwang S J 2024 ResearchAgent: Iterative research idea generation over scientific literature with large language models arXiv: 240407738
- [15] Yang Z, Du X, Li J, Zheng J, Poria S and Cambria E 2024 Large language models for automated open-domain scientific hypotheses discovery 62nd Annual Meeting of the Association for Computational Linguistics (ACL 2024)
- [16] Jain M, Deleu T, Hartford J, Liu C H, Hernandez-Garcia A and Bengio Y 2023 GFlowNets for AI-driven scientific discovery Digital Discovery vol 2(2023) pp 557–577
- [17] Qi B, Zhang K, Li H, Tian K, Zeng S, Chen Z R and Zhou B 2023 Large language models are zero shot hypothesis proposers In Instruction Workshop @ NeurIPS 2023 arXiv: 231105965
- [18] Chris L, Cong L, Robert T, et al 2024 The AI Scientist: Towards fully automated open-ended scientific discovery arXiv preprint arXiv: 240806292 2024b
- [19] Yao S, Zhao J, Yu D, Du N, Shafran I, Narasimhan K and Cao Y 2023 ReAct: Synergizing reasoning and acting in language models arXiv: 221003629
- [20] Wei J, Wang X, Schuurmans D, et al 2022 Chain-of-thought prompting elicits reasoning in large language models In Advances in Neural Information Processing Systems vol 35 pp 24824–24837
- [21] Fricke S 2018 Semantic scholar Journal of the Medical Library Association: JMLA vol 106(1) p 145.
- [22] Liu Y, Chen S, Cheng H et al 2024 CoQuest: Exploring research question co-creation with an LLM-based agent CHI Conference on Human Factors in Computing Systems (CHI '24) p 18.
- [23] Gu X and Krenn M 2024 Generation and human-expert evaluation of interesting research ideas using knowledge graphs and large language models arXiv: 240517044
- [24] Zhao P, Xing Q, Dou K, et al 2024 From words to worth: Newborn article impact prediction with LLM arXiv: 240803934
- [25] Shin H, Choi S, Lim H, et al 2024 Towards an evaluation of LLM-generated inspiration by developing and validating inspiration scale 1st HEAL Workshop at the CHI Conference on Human Factors in Computing Systems p 16
- [26] Beasley J M T, Korn D R, Tucker N N, et al 2024 ExEmPLAR (Extracting Exploring and Embedding Pathways Leading to Actionable Research): A user-friendly interface for knowledge graph mining Bioinformatics vol 40(1) p 779