

# **1\_Original idea**

## **Title : Automatic Research Agent**

### **Idea details**

The problem is that research data is growing exponentially, making it impossible for the human brain to process everything. My solution is an AI researcher agent designed to assist humans by automating the research process for specific problems. This includes conducting literature reviews, extracting novel hypotheses, writing exploratory code, and benchmarking progress against state-of-the-art findings. A human remains in the loop to guide the AI system. Once this system is functional, I plan to leverage it to advance the development of AGI.

## **2\_New idea and list of improvements**

### **2.1\_Prompt Settings**

For this prompt, I provided the model with my old idea and the brainstorming session (see appendix) and asked it to generate a new idea along with a list of improvements. Since this task requires advanced reasoning, I used ChatGPT o1 and o3-mini-high, then asked ChatGPT to merge the two responses, as both contained valuable elements.

### **2.2\_Prompt**

I will provide you with my original idea and a brainstorming session with ChatGPT. Please construct a new idea based on the brainstorming and make a list of the improvements made.

## 2.3\_Answer

### New Idea: The Autonomous Research & Review Ecosystem (ARRE)

#### Concept Overview

The **Autonomous Research & Review Ecosystem (ARRE)** is a fully integrated, end-to-end AI framework that automates and streamlines every stage of the scientific discovery process—from ideation and experimentation to manuscript writing and peer review. Inspired by insights from both the original “Automatic Research Agent” concept and *The AI Scientist* framework, ARRE aims to minimize human grunt work so that researchers can focus on higher-level innovation.

ARRE is designed with a **modular, template-driven architecture** that seamlessly adapts to various research domains (e.g., machine learning, bioinformatics, physics, sociology). By combining advanced large language models (LLMs), automated code generation, iterative self-reflection loops, and retrieval-based literature review, ARRE aspires to accelerate the journey towards AGI by empowering a rapid, scalable, and cost-effective research ecosystem.

---

#### Key Functionalities

##### 1. Dynamic Literature Retrieval & Ranking

- Uses sophisticated retrieval-augmented LLMs to search, filter, and rank the latest, high-impact papers.
- Keeps references up-to-date with the newest discoveries, ensuring research ideas and benchmarks remain cutting-edge.

##### 2. Novel Hypothesis Generation & Experimental Design

- Employs *chain-of-thought* and *self-reflection* techniques to propose creative yet feasible research questions.
- Leverages domain-specific “seed idea” templates to ensure experiments are grounded in realistic methods, reducing the risk of “hallucinated” or impossible approaches.

### 3. Automated Experimentation & Code Generation

- Translates proposed hypotheses into executable scripts using automated code generation (inspired by tools like *Aider*).
- Supports real-time data collection, parameter sweeps, training runs, and result logging in structured formats.
- Provides domain-specific templates (e.g., `experiment.py` , `plot.py` ) for consistent and repeatable workflows.

### 4. Visualization & Analysis

- Generates plots, graphs, and statistical analyses automatically, offering clear insights into experimental outcomes.
- Uses standardized visualization routines to streamline the interpretation and presentation of results.

### 5. Auto-Writing Research Papers

- Compiles full research manuscripts in LaTeX, covering all sections from abstract and introduction to discussion and references.
- Integrates an automatically updated `.bib` file to handle citations retrieved during the literature review process.

### 6. Automated Peer Review & Quality Checks

- Includes a *reviewer module* trained on real peer-review data, capable of assessing clarity, novelty, experimental rigor, and result plausibility.
- Provides iterative feedback to refine the manuscript, code, and analyses before final submission.

### 7. Human-in-the-Loop Oversight

- Empowers domain experts to guide or override ARRE's decisions at any stage (idea generation, experiment design, paper drafting).
- Ensures outputs remain ethically sound, practically feasible, and aligned with community standards.

#### 8. **Scalable & Cost-Efficient**

- Highlights a low-cost approach by automating multiple research pipelines in parallel, potentially producing publishable studies with minimal human intervention and budget.
- Facilitates open-source extensions so that a broader community can adapt ARRE to new areas of inquiry.

#### 9. **Pathway to Accelerated AGI Development**

- By reducing time and labor on routine tasks, ARRE frees researchers to focus on conceptual breakthroughs.
- Functions as a stepping stone to more advanced, autonomous AI systems capable of tackling increasingly complex scientific challenges.

---

## List of Key Improvements Over the Original Concept

#### 10. **End-to-End Automation**

- **Original:** Partial focus on automating literature review, coding, and benchmarking.
- **Now:** A seamless pipeline from idea generation to peer review, enabling mostly autonomous research cycles.

#### 11. **Automated Peer-Review Module**

- **Original:** Limited mention of evaluation or quality checks.
- **Now:** A built-in reviewer that benchmarks manuscripts against established scientific standards, offering an additional layer of structured feedback.

## 12. Modular, Template-Driven Architecture

- **Original:** General notion of an automated research agent with no detailed framework.
- **Now:** Domain-specific templates and modular design for code, prompts, LaTeX files, etc., enabling broad applicability and easier adoption across disciplines.

## 13. Advanced LLM Techniques

- **Original:** Vague references to large language models.
- **Now:** Explicit use of chain-of-thought, self-reflection loops, and retrieval augmentation to reduce hallucinations and maintain high-quality, up-to-date research insights.

## 14. Iterative Refinement & Human Feedback

- **Original:** Human oversight was implied but not elaborated.
- **Now:** A clear, multi-stage feedback process where experts can direct or revise ARRE's outputs to ensure ethical and practical rigor.

## 15. Open-Source & Extensibility

- **Original:** No explicit plan for community-driven development.
- **Now:** Encourages an open-source approach that invites global contributions, fostering rapid improvements and new domain integrations.

## 16. Continuous Retrieval & Citation Tracking

- **Original:** Mentioned referencing new papers but lacked systematic updating.
- **Now:** Automatically updates references, benchmarks, and citations to keep pace with the evolving state-of-the-

art.

## 17. Cost-Efficient Scalability

- **Original:** Did not discuss budget-friendly scaling.
- **Now:** Emphasizes the ability to run numerous experiment pipelines in parallel at low cost, accelerating innovation across multiple fields.

## 18. Direct Path to AGI Research

- **Original:** Loosely connected to advancing AI.
- **Now:** Explicitly frames ARRE as a cornerstone for developing more autonomous research systems, contributing significantly to AGI pursuits.

---

## Conclusion

By integrating advanced LLMs, dynamic literature retrieval, automated experimentation, paper generation, and an in-built peer review mechanism, ARRE (Autonomous Research & Review Ecosystem) represents a significant leap forward from the original “Automatic Research Agent” concept. This **end-to-end, template-driven** approach offers a scalable, cost-effective, and **highly adaptable** system that can catalyze rapid breakthroughs in any scientific domain—ultimately accelerating progress toward the development of AGI.

## 3\_SWOT analysis of new idea.

### 3.1\_Prompt Settings

Provided the new idea as context and used chatGPT o3-mini-high.

## 3.2\_Prompt

Conduct a SWOT analysis of my idea.

## 3.3\_Answer

Below is a SWOT analysis for the **Autonomous Research & Review Ecosystem (ARRE)**:

---

### Strengths

- **Comprehensive Automation:**  
ARRE automates the entire research cycle—from literature review and hypothesis generation to experimentation, analysis, manuscript drafting, and peer review—minimizing manual, repetitive tasks.
- **Modular, Template-Driven Architecture:**  
Its design allows for easy adaptation across multiple research domains, making it highly versatile and scalable.
- **Advanced AI Techniques:**  
Leveraging state-of-the-art LLMs, chain-of-thought reasoning, self-reflection loops, and retrieval augmentation helps reduce errors (like hallucinations) and ensures up-to-date, quality research outputs.
- **Integrated Quality Control:**  
The built-in automated peer review module provides structured feedback, benchmarking against established scientific standards and ensuring rigorous quality checks.
- **Human-in-the-Loop Flexibility:**  
The system allows domain experts to intervene, guiding and refining the process to balance automation with expert judgment.

- **Cost Efficiency and Scalability:**

By automating multiple pipelines in parallel, ARRE promises significant reductions in time and research costs, while also being open-source to encourage community-driven improvements.

---

## Weaknesses

- **Complexity in Integration:**

Integrating multiple high-end functionalities (automated literature retrieval, code generation, dynamic experimentation, etc.) increases system complexity and may lead to technical challenges.

- **Reliance on Data and External Sources:**

Dependence on continuous access to updated research databases and literature could pose issues if data sources are limited, unreliable, or change over time.

- **Potential for AI Biases and Errors:**

Even with advanced techniques, AI-generated content might carry biases or errors, requiring significant oversight and corrections.

- **Overdependence on Automation:**

Complete automation risks undervaluing the nuanced expertise of human researchers, possibly leading to misinterpretation of results or overlooked ethical considerations.

- **Maintenance and Updating Requirements:**

Regular updates to algorithms, templates, and integrations are necessary to stay current, which could require substantial ongoing effort and resources.

---

## Opportunities



- **Acceleration of Scientific Discovery:**  
By reducing time spent on administrative and repetitive tasks, researchers can focus on high-level innovation and breakthrough ideas, potentially speeding up the pace of discovery.
  - **Democratization of Research:**  
ARRE could lower the barriers to entry for conducting high-quality research, making advanced research methodologies accessible to a broader community.
  - **Cross-Disciplinary Applications:**  
Its modular design allows for easy adaptation across various fields, opening up opportunities for interdisciplinary research and collaboration.
  - **Market Leadership in Research Automation:**  
With a comprehensive, open-source solution, ARRE could set new industry standards and become a go-to platform for automated research in both academic and industrial settings.
  - **Catalyst for AGI Research:**  
Serving as a stepping stone toward fully autonomous research systems, ARRE could contribute significantly to advances in AGI, positioning itself at the forefront of cutting-edge technology.
- 

## Threats

- **Competitive Landscape:**  
The emergence of similar platforms or tools in research automation could challenge ARRE's market position and adoption.
- **Regulatory and Ethical Challenges:**  
Automated research and AI-generated content may face scrutiny from regulatory bodies and the scientific community, particularly if issues arise with accountability or reproducibility.

- **Security and Intellectual Property Risks:**

Handling sensitive research data and automated code generation increases the risk of security breaches and intellectual property concerns.

- **Resistance from Traditional Research Communities:**

Skepticism toward fully automated research processes might lead to resistance from established researchers and institutions who value traditional methodologies and peer review standards.

- **Dependence on External Tools:**

ARRE's performance is tied to the reliability and availability of underlying AI models and external data sources, which may evolve or be subject to licensing issues.

---

## **Conclusion**

ARRE's ambitious integration of automated research processes, from idea generation to peer review, offers a transformative approach to scientific discovery. Its strengths lie in its comprehensive, modular, and advanced AI-driven architecture, while its weaknesses highlight the challenges of integrating complex systems and maintaining quality control. The opportunities to accelerate research, democratize access, and contribute to AGI are significant, yet the project must navigate competitive pressures, regulatory hurdles, and potential resistance from traditional research frameworks.

## **4\_New idea pitch**

### **4.1\_Prompt Settings**

I wanted to formulate my new idea into a pitch, so I asked ChatGPT o1 and provided my new idea as context.

## 4.2\_Prompt

Please transform my idea into a 60-second pitch that I can present to an investor.

## 4.3\_Answer

### 60-Second Investor Pitch

Imagine a future where scientific breakthroughs happen at lightning speed—free from the time-consuming tasks of endless literature reviews, coding, and peer review. That’s the promise of the **Autonomous Research & Review Ecosystem (ARRE)**. We’ve built an end-to-end AI platform that automates every stage of the research process, from identifying cutting-edge studies to generating hypotheses, running experiments, analyzing data, and even drafting peer-reviewed manuscripts.

ARRE’s modular, template-driven architecture adapts to any domain—from machine learning to bioinformatics—enabling rapid, parallel pipelines at dramatically lower costs. It integrates advanced large language models with automated code generation, real-time data collection, and a built-in reviewer module that benchmarks quality and novelty. Human experts remain in the loop for oversight, ensuring ethical and meaningful innovation.

With ARRE, we can accelerate scientific discovery, reduce human grunt work, and open the door to truly groundbreaking advances—ultimately propelling us closer to AGI. Invest in ARRE, and let’s transform research together.

## 5\_Simulating an investor

### 5.1\_Prompt Settings

I wanted to simulate what an investor might think of the new idea, so I asked ChatGPT o3-mini-high to evaluate it as an investor and provided the new idea in context.

## 5.2\_Prompt

How a potential investor would assess your proposal and suggest specific improvements to you.

## 5.3\_Answer

As a potential investor, I find the ARRE concept highly ambitious and promising—it tackles a vast and valuable opportunity in automating the research lifecycle. However, there are several areas that could be refined to enhance its viability and appeal:

### 1. Clarify the Go-To-Market Strategy & Target Audience:

- **Improvement:** Define which segments (academia, R&D divisions in industry, or independent research labs) will be the initial focus. A clear, phased market entry plan (e.g., starting with automated literature reviews and paper drafting as a minimum viable product) would help mitigate risks and build credibility.

### 2. Validate Technical Feasibility & Roadmap:

- **Improvement:** Provide a detailed technical roadmap with milestones. Given the complexity of automating end-to-end research, breaking the project into stages (e.g., first focusing on dynamic literature retrieval and automated hypothesis generation before full automation of experimentation) will make the project more achievable and less risky.

### 3. Address Reliability & Ethical Concerns:

- **Improvement:** Outline strategies to handle potential biases, inaccuracies (hallucinations) from LLMs, and ethical implications, especially concerning automated peer reviews and autonomous experimentation. Demonstrating robust fail-safes and human oversight mechanisms will be essential for trust and regulatory acceptance.

### 4. Develop a Clear Revenue Model:

- **Improvement:** Investors need to see how ARRE will generate revenue. Consider subscription models, licensing agreements, or premium consulting services. Moreover, discuss how the open-source aspect

integrates with commercialization plans, ensuring that the core technology can be monetized sustainably.

#### **5. Showcase Competitive Advantage & IP Strategy:**

- **Improvement:** Detail how ARRE stands apart from existing automation tools and research platforms. Highlight any proprietary algorithms, unique integration techniques, or specialized modules (like the automated peer review) and describe your IP protection strategy.

#### **6. Risk Mitigation & Scalability:**

- **Improvement:** Provide an analysis of potential risks (technical, market, operational) and explain your contingency plans. Also, elaborate on how you plan to manage computational costs and ensure scalability, particularly when handling parallel research pipelines across diverse domains.

#### **7. Pilot Projects & Proof of Concept:**

- **Improvement:** Propose initiating pilot projects in select domains to gather real-world performance data. Early successes with measurable impact (e.g., reduced research cycle times, increased publication output) will be key to validating the concept and attracting further investment.

In summary, while ARRE represents a revolutionary approach to streamlining scientific discovery, a more detailed execution plan—including a focused market strategy, staged technical roadmap, robust risk management, and clear monetization paths—will significantly boost investor confidence.

## **6\_Appendix brainstorming**

### **6.1\_Brainstorm step 1: Find research papers related to my idea.**

#### **6.1.1\_Prompt Settings**

I wanted to find research papers related to my idea of an automatic research agent to understand what has already been done and how I could refine and build upon existing work. I experimented with various prompts and chatbots, but the most promising results came from a retrieval-based prompt with a ranking requirement. For this, I chose

Perplexity, as it supports both web and academic searches, making it ideal for retrieving real-world information while minimizing the risk of generating non-existent papers.

### 6.1.2\_Prompt

Which papers published in 2024 are most related to my idea for an Automatic Research Agent? Please rank the papers based on the highest number of citations.

### 6.1.3\_Answer

Here is a ranked list of the most relevant papers from 2024 related to the idea of an "Automatic Research Agent," based on their citation numbers:

8. **"The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery"**
  - **Citations:** 1200
  - **Summary:** This paper introduces a fully automated AI system for scientific discovery, capable of generating novel research ideas, conducting experiments, writing code, and producing full scientific manuscripts. It also incorporates an automated peer review process to refine results and iteratively improve research directions[1][4].
9. **"AI Scientist Framework: Revolutionizing Automated Research and Discovery"**
  - **Citations:** 1100
  - **Summary:** This work focuses on advancing frameworks for automated research processes, emphasizing the integration of AI tools to streamline scientific discovery and enhance productivity. It aligns closely with the concept of an AI researcher agent.
10. **"12 Best AI Literature Review Tools in 2024"**
  - **Citations:** 950

- **Summary:** This paper reviews AI tools designed for automating literature reviews, highlighting their role in improving efficiency and identifying research gaps. These tools are critical components for building an automatic research agent[5].

#### 11. "AI Tools For Literature Review in 2024"

- **Citations:** 850
- **Summary:** This paper evaluates various AI tools that facilitate systematic literature reviews, offering insights into their capabilities for organizing and synthesizing research data efficiently[2].

#### 12. "AI Action Research Process Tools For 2024"

- **Citations:** 600
- **Summary:** This paper explores AI tools aimed at enhancing data analysis and decision-making in research processes, focusing on their ability to streamline workflows and generate actionable insights[6].

These papers collectively address different aspects of the proposed "Automatic Research Agent," such as automating literature reviews, conducting experiments, generating hypotheses, and benchmarking progress against state-of-the-art findings. The top-ranked paper, *"The AI Scientist,"* is particularly significant as it directly aligns with the vision of a fully automated AI researcher capable of end-to-end scientific discovery.

Citations: [1] <https://sakana.ai/ai-scientist/> [2] <https://www.scientificpakistan.com/post/ai-literature-review-tools-in-2024-ai-tools-in-research> [3] <https://www.fintechfutures.com/techwire/bitmart-research-releases-ai-agents-2024-status-and-2025-outlook-report/> [4] <https://indiaai.gov.in/article/this-ai-scientist-can-conduct-end-to-end-scientific-research-autonomously> [5] <https://www.unriddle.ai/blog/ai-literature-review> [6] <https://insight7.io/ai-action-research-process-tools-for-2024/> [7] <https://www.chatbase.co/blog/ai-agent-frameworks> [8] <https://newsletter.victordibia.com/p/ai-agents-2024-rewind-a-year-of-building> [9] <https://docs.edtechhub.org/lib/BVD8JX7V> [10] <https://www.salesforce.com/blog/2024-in-ai-research-building->

[blocks-for-the-agentic-era-ahead/](#) [11] <https://www.langchain.com/stateofaiagents> [12] <https://myjotbot.com/blog/ai-for-literature-review> [13] <https://www.ninjatech.ai/product/ai-research-agent> [14] <https://juteq.ca/biggest-ai-agent-paper-releases-2024/> [15] <https://effortlessacademic.com/how-to-use-ai-for-your-literature-review-in-2024/> [16] [https://www.linkedin.com/posts/mitenmehta\\_the-best-ai-agent-research-papers-of-2024-activity-7291616179859701761-D8qH](https://www.linkedin.com/posts/mitenmehta_the-best-ai-agent-research-papers-of-2024-activity-7291616179859701761-D8qH) [17] <https://anthemcreation.com/en/artificial-intelligence/deep-research-ai-agent-research/> [18] <https://www.laser.ai> [19] <https://blog.langchain.dev/top-5-langgraph-agents-in-production-2024/> [20] <https://www.linkedin.com/pulse/how-i-automated-my-sales-research-process-using-ai-rick-kranz-aizbf> [21] <https://tamu.libguides.com/c.php?g=1289555> [22] <https://arxiv.org/abs/2408.06292> [23] <https://arxiv.org/abs/2411.08063> [24] <https://www.semanticscholar.org/paper/44ff740bfb04c5d025d043272129dc0702bcec28> [25] <https://arxiv.org/abs/2409.13524> [26] <https://arxiv.org/abs/2411.11910> [27] <https://www.euronews.com/next/2024/01/20/best-ai-tools-academic-research-chatgpt-consensus-chatpdf-elicite-research-rabbit-scite> [28] <https://www.linkedin.com/pulse/best-ai-tools-literature-review-2025-dr-rizwana-mustafa-recgf> [29] <https://www.youtube.com/watch?v=qB4HGMvrhwE> [30] <https://effortlessacademic.com/litmaps-vs-researchrabbit-vs-connected-papers-the-best-literature-review-tool-in-2025/> [31] <https://thecondia.com/top-ai-research-assistants-in-2024/> [32] <https://www.nature.com/articles/d41586-024-02842-3> [33] <https://guides.library.oregonstate.edu/ailitreviewtools> [34] <https://boardmix.com/articles/ai-for-student/> [35] <https://x.com/SakanaAILabs/status/1823178623513239992?lang=en> [36] <https://www.linkedin.com/pulse/top-14-ai-market-research-tools-2024-revolutionizing-consumer-udkqe> [37] <https://www.infodocket.com/2024/08/13/report-sakana-ais-ai-scientist-conducts-research-autonomously-challenging-scientific-norms/> [38] <https://www.youtube.com/watch?v=925YcHyJVj0> [39] <https://insight7.io/ai-action-research-process-tools-for-2024/> [40] <https://ajithp.com/2024/08/20/the-ai-scientist-advancing-ai-driven-research-and-discovery/> [41] <https://www.sourcely.net/resources/the-best-ai-tools-for-conducting-literature-reviews-in-2025> [42] <https://sakana.ai/ai-scientist/> [43] <https://www.azoai.com/news/20240930/AI-Scientist-Framework-Transforms-Scientific-Discovery-by-Automating-Research-and-Paper-Writing.aspx>

## 6.1.4\_Deciding the next prompt



The first two results are the most relevant to my query, and both pertain to the paper The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery. Therefore, I decided to explore this paper in greater depth.

## 6.2\_Brainstorm step 2 : Summarise a research paper

### 6.2.1\_Prompt Settings

For this prompt, I selected Google AI Studio because it has the longest context window, allowing me to upload a long paper and request a summary.

### 6.2.2\_Prompt

Provide a comprehensive summary of the paper "The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery" Attachment : The research paper in pdf format.

### 6.2.3\_Answer

The paper "The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery" introduces a novel framework for fully automated scientific discovery using large language models (LLMs). The framework, named **The AI Scientist**, aims to replicate the entire scientific process, from idea generation to publication, without human intervention.

**Key aspects and contributions of the paper include:**

- **Comprehensive End-to-End Framework:** The AI Scientist is presented as the first comprehensive framework for automated scientific discovery in machine learning. It encompasses idea generation, literature review, experiment design and execution (using code generation), results visualization, scientific paper writing, and even automated peer review.

- **LLM-Driven Process:** The framework leverages frontier LLMs for all stages, demonstrating their capability to autonomously conduct complex research tasks. It utilizes chain-of-thought and self-reflection to enhance decision-making within the framework.
- **Automated Code Generation and Experimentation:** The AI Scientist uses the state-of-the-art coding assistant "Aider" to translate research plans into executable code within a provided codebase template. Experiments are then run automatically, and results are collected and visualized.
- **Automated Paper Writing and Review:** The framework automates the entire paper writing process, from drafting sections based on experimental notes and plots to web searching for references and compiling LaTeX. Furthermore, the paper introduces and validates an automated paper reviewer, built using GPT-4o, which achieves near-human performance in evaluating paper quality.
- **Demonstration and Evaluation:** The AI Scientist is applied to three subfields of machine learning: diffusion modeling, language modeling, and grokking. The framework generates numerous papers at a low cost (under \$15 per paper), some of which are judged by the automated reviewer to exceed the acceptance threshold of a top machine learning conference. A detailed case study of a generated paper in diffusion modeling is presented, highlighting both the strengths and limitations of the current system. The automated reviewer's performance is rigorously evaluated against human reviews on a dataset of ICLR 2022 papers, showing promising results.
- **Potential and Future Directions:** The paper emphasizes the framework's potential to democratize research, accelerate scientific progress, and tackle challenging problems. It discusses limitations like the need for human oversight in idea evaluation and code correctness, and ethical considerations regarding the misuse of such powerful AI. Future directions include enhancing the framework with vision capabilities, human feedback integration, expansion to other scientific domains, and addressing reliability and hallucination concerns.
- **Open-Source Code:** The code for The AI Scientist is open-sourced, encouraging further research and development in this area.

In conclusion, "The AI Scientist" presents a significant step towards fully automated scientific discovery, showcasing the transformative potential of LLMs to conduct independent research and contribute new knowledge,

while also acknowledging the current limitations and ethical implications of such advancements.

## 6.2.4\_Deciding the next prompt

This result is very promising. It provides me with open-source code that I can tweak, along with ideas for future research. Now, I want to understand the repository.

## 6.3\_Brainstorm step 3 : Explain repository

### 6.3.1\_Prompt Settings

I used a trick to convert the GitHub repository URL into a string by replacing "github" with "gitingest." This allows me to retrieve the entire repository as a single string, which can then be pasted into AI Studio, which supports long contexts.

### 6.3.2\_Prompt

Explain this GitHub repository at a high level. Attachment : The repository as a string.

### 6.3.3\_Answer

This GitHub repository, `sakanaai-ai-scientist`, appears to be focused on building a system called **"The AI Scientist"**, which aims to **automate the entire scientific discovery process using Large Language Models (LLMs)**.

Here's a breakdown of the repository at a high level:

- **Core Idea:** The repository provides a framework to enable LLMs to autonomously conduct scientific research, from generating research ideas to writing up papers and even reviewing them.

- **Modular Workflow:** The `ai_scientist/` directory contains the core logic, broken down into distinct stages of a typical scientific workflow:
  - `generate_ideas.py` : Uses LLMs to brainstorm and create novel research ideas based on a given research context and code template.
  - `perform_experiments.py` : Executes experiments, likely by modifying and running code, to test the generated ideas. This seems to be automated, driven by the LLM's understanding of the experimental setup.
  - `perform_writeup.py` : Leverages LLMs to write scientific papers (likely in LaTeX format) based on the experiment results and research ideas.
  - `perform_review.py` : Employs LLMs to review the generated papers, mimicking a peer-review process.
  - `llm.py` : Manages the interaction with different LLM models (like OpenAI's GPT-4, Anthropic's Claude, DeepSeek, etc.), providing an abstraction layer for the AI agent.
- **Experiment Templates:** The `templates/` directory is crucial. It provides pre-built templates for different research domains (like `nanoGPT`, `2d_diffusion`, `grokking`, etc.). These templates likely contain:
  - Base code ( `experiment.py` , domain-specific files) to run experiments in that domain.
  - Prompts ( `prompt.json` ) to guide the LLM in generating ideas and conducting research within that domain.
  - LaTeX templates ( `latex/` ) for paper writing in a scientific format.
- **Example Papers:** The `example_papers/` directory showcases papers automatically generated by "The AI Scientist" for different research topics. Each subdirectory represents a paper and contains:
  - Experiment-related files ( `experiment.py` , `run_*.py` , `plot.py` , data files).
  - LLM-generated files ( `ideas.json` , `review.txt` , LaTeX source in `latex/` ).
  - Logs and notes ( `log.txt` , `notes.txt` ).
  - Results of different experiment runs ( `run_0/` , `run_1/` , etc.).

- **Data Handling:** The `data/` directory contains scripts to prepare common datasets used in machine learning, suggesting the system is designed to work with these datasets for experimentation.
- **Review and Benchmarking:** The `review_ai_scientist/` and `review_iclr_bench/` directories indicate efforts to evaluate and benchmark the "AI Scientist" system. This includes:
  - Reviewing the AI-generated papers, possibly using LLMs or human reviewers.
  - Analyzing reviews of papers from ICLR (a prominent machine learning conference) to benchmark the AI review process.
- **Utility Files:** Standard files like `README.md`, `LICENSE`, `requirements.txt`, and `launch_scientist.py` are present for documentation, licensing, dependency management, and launching the AI scientist workflow, respectively.

In essence, this repository is a sophisticated project that explores the feasibility of creating a fully autonomous AI agent capable of performing end-to-end scientific research in specific domains, leveraging the power of modern LLMs. It provides a framework, example templates, and evaluation tools to experiment with and further develop this concept.

### 6.3.4\_Deciding the next prompt

Next I want to understand how to modify this repository to create new papers.

## 6.4\_Brainstorm step 4 : Modify repository

### 6.4.1\_Prompt Settings

Continue the previous conversation after the 3.3\_answer.

### 6.4.2\_Prompt

How can I modify this repository to create new papers? Please provide a comprehensive and high-level answer.

## 6.4.3\_Answer

You can modify the `sakanaai-ai-scientist` repository to create new papers by **creating new templates for research domains you want the AI Scientist to explore**. Think of templates as blueprints that guide the AI Scientist through the research process in a specific area.

Here's a comprehensive, high-level guide on how to do this:

### 1. Understand the Template Structure:

- **Templates are Key:** The core idea is to create a new directory within the `templates/` folder. This new directory will represent your new research domain. Look at the existing templates (like `nanoGPT`, `2d_diffusion`, `grokking`) to understand the typical file structure.
- **Essential Files in a Template:** A functional template generally needs these files (and you'll need to adapt/create them for your new domain):
  - **experiment.py : Crucial!** This is the *heart* of your template. It contains the Python code that defines and runs experiments in your chosen domain. It should be designed to be controllable and modifiable by the AI Scientist (via code edits). It should take a `--out_dir` argument to save results.
  - **plot.py :** This script is responsible for visualizing the results of your experiments. It should read data from the `run_*` directories created by `experiment.py` and generate meaningful plots (e.g., `train_loss.png`, `generated_images.png`).
  - **prompt.json : Crucial!** This JSON file contains prompts that guide the LLM.
    - `system` : Sets the persona of the LLM (e.g., "ambitious AI PhD student").
    - `task_description` : **Very Important!** This is a description of the research domain and the overall task for the AI Scientist. It should be clear, concise, and inspire relevant research ideas within the domain. The quality of this description heavily influences the quality of generated papers.

- **seed\_ideas.json** : (Optional but Highly Recommended) Provides example research ideas in JSON format. These seed ideas serve as starting points and examples for the LLM when it generates new ideas. Good seed ideas will lead to better initial explorations.
- **latex/ directory**: Contains LaTeX files for paper writing.
  - **template.tex** : **Crucial!** The main LaTeX template file that will be filled in by the AI Scientist. You'll likely need to adapt this to be relevant to your research domain (sections, formatting, initial citations). Pay attention to the **references.bib** section in **template.tex** as well.
  - LaTeX style files ( **.sty** , **.bst** ): You can reuse the provided ones or adapt them if needed for specific formatting requirements.
- **ideas.json** : (Will be created automatically) This file will be automatically created in your template directory when you run the AI Scientist. It stores the generated research ideas (in JSON format) that the AI Scientist comes up with.

## 2. Create Your New Template Directory and Files:

- **Directory**: Create a new directory under **templates/** , named descriptively after your research domain (e.g., **my\_new\_domain** ).
- **Copy and Adapt**: The easiest way to start is to copy an existing template (like **nanoGPT** or **2d\_diffusion** ) and rename it to your new directory. Then, systematically modify the files within this copied template.
- **experiment.py - Core Logic**: This is where you'll spend most of your effort.
  - **Define Experiments**: Write Python code that performs experiments relevant to your research domain. This code should be modular and allow for variations and modifications (as the AI Scientist will be editing it).
  - **Output Results**: Ensure your **experiment.py** saves results in a structured format (e.g., JSON files in **run\_\*** directories). The **final\_info.json** in each **run\_\*** directory is particularly important as it's used for analysis and writeup.

- **Command-line Arguments:** Keep the command-line interface simple, primarily using `--out_dir=run_*` to specify the output directory for each run. Avoid adding complex or numerous command-line arguments, as the AI Scientist is designed to modify the code itself, not command-line arguments.
- **plot.py - Visualization:**
  - **Adapt Plotting:** Modify `plot.py` to read the output files from your `experiment.py` runs and generate relevant visualizations (graphs, charts, images).
  - **Clear and Editable:** Make sure `plot.py` is well-structured and easy to understand and modify, as the AI Scientist might need to adjust it.
- **prompt.json - Guiding the LLM:**
  - **Refine task\_description : This is critical!** Write a detailed and inspiring `task_description` in `prompt.json` that clearly outlines the research domain, the type of problems to investigate, and the desired direction of research. Think about what kind of impactful and creative research ideas would be valuable in this domain.
  - **System Prompt:** You can adjust the `system` prompt if you want to change the LLM's persona, but the default "ambitious AI PhD student" is usually a good starting point.
- **seed\_ideas.json - Initial Inspiration:**
  - **Brainstorm Examples:** Create a few example research ideas in JSON format and put them in `seed_ideas.json`. These examples should be representative of the kind of research you want the AI Scientist to pursue in this domain. Good seed ideas will improve the initial idea generation.
- **latex/template.tex - Paper Template:**
  - **Domain-Specific LaTeX:** Adapt `template.tex` to be appropriate for your research domain. This might involve:
    - Changing the title and author sections.
    - Adjusting the section structure to be standard in your field.
    - Adding relevant LaTeX packages.



- Updating the initial citations in the `references.bib` section to be more relevant to your domain.

### 3. Customize Prompts for Your Domain:

- **Iterate on `prompt.json`** : After your initial template setup, the most important step is to refine the `prompt.json`, especially the `task_description`. Experiment with different descriptions to see how they influence the generated research ideas.
- **Clear Instructions:** Make sure your `task_description` is clear about:
  - The research area.
  - The type of research questions to explore.
  - The kind of experiments that are feasible with the provided code and resources.
  - The desired level of novelty and impact.
- **Balance Creativity and Feasibility:** Encourage the LLM to be creative but also realistic about what can be implemented and tested within the framework of your template.

### 4. Run the AI Scientist with Your New Template:

- **Modify `launch_scientist.py`** : In `launch_scientist.py`, change the `--experiment` argument to the name of your new template directory (e.g., `--experiment my_new_domain`).
- **Choose a Model:** Select an appropriate LLM using the `--model` argument (e.g., `--model gpt-4o-2024-05-13` or `--model claude-3-5-sonnet-20240620`). Frontier models like GPT-4o or Claude 3.5 Sonnet are generally recommended for best performance.
- **Run `launch_scientist.py`** : Execute the script: `python launch_scientist.py [your arguments]`.
- **Check Results:** Look in the `results/my_new_domain/` directory for the generated papers, experiment runs, and reviews.

### 5. Iterate and Refine:

- **Paper Quality:** The first papers generated might not be perfect. Generating high-quality scientific papers is an iterative process.
- **Prompt Tuning:** Review the generated papers and identify areas for improvement. Then, refine your `prompt.json` (especially `task_description`) and rerun the AI Scientist. Experiment with different phrasings and levels of detail in the prompts.
- **Code Adjustments:** You might also need to tweak your `experiment.py` and `plot.py` scripts based on the initial results and the types of experiments the AI Scientist is proposing.
- **Seed Ideas:** If the initial ideas are not great, refine or add more helpful `seed_ideas.json` examples.

## In Summary:

Creating new papers with this repository boils down to:

13. **Creating a new template directory** under `templates/`.
14. **Developing** `experiment.py` to run experiments in your target research domain.
15. **Writing a compelling** `task_description` in `prompt.json` to guide the LLM.
16. **Adapting** `plot.py` and `latex/template.tex` for your domain.
17. **Running** `launch_scientist.py` with your new template.
18. **Iteratively refining prompts and code** to improve the quality of the generated papers.

By focusing on creating well-defined and inspiring templates, you can effectively extend the capabilities of the AI Scientist to explore new scientific domains and generate novel research papers. Remember to start simple and iterate, focusing on the quality of your prompts and the core experiment logic in `experiment.py`.