Grenze International Journal of Engineering and Technology, June Issue

Open Deep Research: A Modular Graph-based Framework for Automated Research Synthesis using Large Language Models

Nishant Tekdiwal[1], Shubham Solanke[2], Harshal Malani[3], Saubiya Adnas[4] and Padma Adane[5]
[1-5]Dept. of Data Science RCOEM, Nagpur, India
Email: tekdiwalna@rknec.edu, solankesp@rknec.edu, malanihj@rknec.edu, adnass@rknec.edu, adanep@rknec.edu

***Abstract*— With the increasing amount of digital data, it is much more difficult and takes longer to do research than ever before. Data sources are susceptible to bias due to their limited scale and lengthy validation cycles. Hence, they may be biased. In this paper, they describe a new kind of modular research automation system called Open Deep Research, which uses Artificial Intelligence (AI) and Large Language Models (LLMs) in combination with Graph-based multi-agent architecture to automate research processes from start to finish. There are three interconnected modules of Open Deep Research: a Section Graph that provides users with resolvable report structures, whereas Research consists of graphical data that allows researchers to retrieve and verify information from various sources. Additionally, the Writer graph is designed to enable scientists in Oregon to authenticate documents with appropriate citations from multiple accurate sources using tools such as HTML or PDFs. Using highly developed and transparent (NLP) techniques, Open Deep Research uses agentic workflows to achieve end-to-end research automation in a highly scalable, trackable manner.' Open Deep Research is modular in its approach, allowing it to be extended and maintaining integrity through the creation of content that is referenced. It has been shown that implementing Open Deep Research can significantly reduce the time required for research while still providing clear documentation, making it an appropriate choice among students/researchers and professionals conducting complex interdisciplinary research. LLaMA-3.3, knowledge synthesis, graph-based architecture, research automation, and large language models are all included in the index terms.**

***Index Terms*— Large language models, research automation, graph-based architecture, multi-agent systems, knowledge syn-thesis, LLaMA-3.3.**

I. INTRODUCTION

While information is more readily available by digital means, detailed research is now more challenging because the information has become more complex to research. Re-searchers are inundated by the volume of academic papers, articles, data sets, and online information to sift through for new and useful insights. Traditional approaches to research are often time-consuming, labor intensive and prone to human er-rors and inconsistencies. With the proliferation of information sources, there is a need for flexible and automated solutions that can keep up with academic standards and speed academic research along with increased accuracy.

Grenze ID: 01.GIJET.12.2.329_26

Large Language Models (LLM) are now a crucial tool for automating knowledge tasks, thanks to recent advances in Artificial Intelligence (AI). GPT-4, LLaMA-3.3 and Claude stand out as some of the models that showcase their natural language understandings and the ability to do reasoning tasks and summarization. Most of the current research instruments that are available for AI research are limited to a specific step in the research process, for example, incorporating ex-ternal context elements, rather than providing full end-to-end automation with transparency and traceability.

Open Deep Research is suggested as the basis for a holistic automated research synthesis methodology. This paper pro-vides an overview. Current research automation systems are often found to have their limitations, which are typically addressed by systems that are restricted to certain functions in research. To overcome these limitations there are three major breakthroughs. The first step towards scaling up the execution is to divide the research tasks into isolated and interconnected agents through an architecture of modular graphs. This makes it easier to run the code sequentially. Our second goal is to introduce a multi-source retrieval system which integrates mul-tiple data sources, ranking methods and verification methods to enhance the reliability and quality of information. We propose to develop a ”research sync tool” that will give outputs that are transparent and verifiable, with a citation structure and traceability.

The three primary components of the Coordination module of the system are Section Graph, Research Grasp and Writer

Spec.[B]. Each of them functions as a distinct entity, with sig-nificant collaboration. This will enable flexibility in conducting research as well as making it transparent from the frame of the questions to the final report. The user receives directions as it passes through.

Unlike monolithic AI assistants that provide output without any traceability of information sources, Open Deep Research guarantees the traceability of all information to allow users to check its authenticity and to be able to follow an explanation process. The flexible structure allows academic institutions, research centers and individual researchers to concentrate on customization in the context of community-based development and on the specific subject. This system can reduce manual work and improve the accuracy of facts, which allows users to conduct accurate analysis without infringing on academic integrity.

II. Related Work

A. Traditional Research Automation Systems

Early research automation systems primarily relied on clas-sical information retrieval and statistical natural language processing techniques. Methods such as TextRank [1] uti-lized graph-based ranking algorithms inspired by PageRank to identify important keywords and sentences within documents. Similarly, stemming algorithms such as Porter Stemmer [2] improved document indexing and retrieval efficiency through lexical normalization. Although these approaches improved keyword extraction and document retrieval, they lacked deep semantic under-standing and were unable to perform comprehensive research synthesis across multiple heterogeneous sources.

B. Deep Learning for Academic Text Processing

The emergence of transformer-based architectures signifi-cantly improved natural language understanding capabilities in academic text processing. BERT [3] introduced bidirectional contextual embeddings, enabling improved semantic represen-tation of scientific text and better understanding of technical terminology.

Cohan et al. [6] proposed discourse-aware summarization models specifically designed for long scientific documents. Their work demonstrated the importance of incorporating document structure and rhetorical relationships into neural summarization systems for academic literature.

While these systems improved summarization quality, they primarily focused on isolated tasks such as classification or summarization rather than complete end-to-end research automation.

C. Large Language Models for Research Assistance

Recent Large Language Models (LLMs), including GPT-4

[4] and LLaMA [5], have demonstrated strong capabilities in reasoning, summarization, and technical content generation. These models are increasingly used in research-oriented ap-plications due to their ability to synthesize information from multiple sources and generate coherent long-form text.

Commercial systems such as OpenAI Deep Research [8] and JigsawStack-based workflows [7] provide partial automa-tion for literature search and report generation. However, most existing systems operate as monolithic pipelines with limited transparency, restricted citation traceability, and minimal mod-ularity.

Community discussions on platforms such as Reddit [9] and Together AI [10] further highlight the growing demand for open, modular, and verifiable research automation systems capable of exposing intermediate reasoning and source prove-nance.

D. Multi-Agent and Graph-Based Architectures

Graph-based multi-agent systems have recently emerged as effective approaches for decomposing complex reasoning tasks into smaller specialized components. Frameworks such as LangGraph and LangFlow support stateful workflows in which multiple agents collaboratively perform planning, retrieval, reasoning, and synthesis tasks.

These systems demonstrate the advantages of modular or-chestration and persistent state management. However, existing implementations generally lack integrated mechanisms for claim-level citation tracking, adaptive report structuring, and credibility-aware multi-source retrieval.

III. MOTIVATION

The motivation behind developing Open Deep Research arises from the significant difficulties researchers face in today's digital era. The exponential growth of academic pa-pers, technical reports, online articles, and datasets has made comprehensive literature review and research synthesis in-creasingly time-consuming and challenging. Manual research processes often require weeks or even months of effort, involv-ing repeated searching, reading, cross-verification, and note-taking. These traditional methods are not only labor-intensive but are also prone to human errors, oversight of important sources, and unintentional bias in source selection. Moreover, many existing AI-powered research tools provide only partial automation, such as literature search or text summarization, without offering end-to-end support from topic structuring to final report generation with proper citations.

There is a strong and growing need for an automated, transparent, and modular system that can handle the entire research workflow while strictly maintaining academic standards, traceability of sources, and intellectual integrity. Open Deep Research was developed to address these challenges by combining the power of Large Language Models with a graph-based multi-agent architecture, aiming to significantly reduce research time while improving the quality, reliability, and reproducibility of research outputs.

IV. PROBLEM DOMAIN

The problem domain of this work lies in the field of research automation within academic and professional environments. Researchers across disciplines are required to process and synthesize large volumes of heterogeneous data coming from multiple sources such as scholarly databases, open-access repositories, web articles, technical reports, and PDFs. These sources often vary greatly in quality, credibility, format, and recency. Key challenges in this domain include handling bias present in data sources, managing information overload, en-suring factual accuracy, maintaining proper academic citations, and producing coherent, well-structured research documents. Additionally, interdisciplinary research further complicates the process as it demands integration of knowledge from multi-ple fields. Traditional manual approaches and current partial automation tools struggle to provide efficient, verifiable, and scalable solutions for comprehensive literature synthesis. This creates a pressing need for intelligent systems that can auto-mate the entire research pipeline while preserving transparency and academic rigor.

V. PROBLEM DEFINITION

Given a user-provided research query Q, the system must automatically generate a complete, well-structured, and aca-demically acceptable research report. Specifically, the system is expected to:

- Dynamically create a logical hierarchical section structure suitable for the given research topic and user require-ments.
- Retrieve relevant and credible information from multiple heterogeneous sources using advanced search and infor-mation retrieval techniques.
- Verify and rank the retrieved information based on rele-vance, credibility, and recency.
- Synthesize the collected information into coherent, original-sounding narrative text while maintaining factual accuracy.
- Provide fine-grained citations at the claim or sentence level, ensuring full traceability of all information used.

- Produce the final report in a clear, professional format that meets academic standards.

The entire process must be transparent, modular, and scalable, allowing users to trace back any statement to its original source while preserving academic integrity throughout the research synthesis pipeline.

VI. Statement

A. Context and Motivation

The exponential growth of digital academic literature has fundamentally altered the landscape of scholarly research. As of 2024, over 2.5 million peer-reviewed papers are published annually across disciplines, with preprint servers such as arXiv and bioRxiv contributing tens of thousands of additional docu-ments each month [12].

Researchers, students, and knowledge workers are consequently confronted with an overwhelming volume of heterogeneous information that must be retrieved, evaluated, synthesized, and presented—often under significant time constraints and with limited institutional support.

B. Formal Problem Definition

Let Q denote a natural language research query issued by a user u with domain expertise level $\delta \in [0, 1]$ and a target output format $F \in \{\text{survey, comparative analysis, technical report, literature review}\}$.

The research synthesis problem is formally defined as follows:

Given:

- A research query Q with associated parameters (δ, F, L), where L represents length and structural constraints.
- A heterogeneous corpus of information sources $D = \{d_1, d_2, \ldots, d_N\}$, drawn from academic databases, web APIs, and document repositories.
- A quality threshold τ defining the minimum acceptable standard of factual accuracy, citation completeness, and structural coherence.

The objective is to produce a research report R such that:

$$R = \arg\max_{\hat{R}} \mathbb{E}\left[\text{Quality}(\hat{R}, Q, D)\right] \quad (1)$$

where the Quality $(\cdot)$ function is a composite measure de-fined over three primary dimensions:

$$\text{Quality}(R, Q, D) = \lambda_1 \cdot \text{Acc}(R) + \lambda_2 \cdot \text{Cov}(R, Q) + \lambda_3 \cdot \text{Coh}(R) \quad (2)$$

Here, Acc(R) denotes factual accuracy, defined as the proportion of verifiable claims in R that can be traced to authoritative sources within D. Cov(R, Q) denotes coverage, defined as the degree to which R addresses all semantically relevant sub-topics of Q. Coh(R) denotes structural and narrative coherence, measured by the logical consistency of section ordering, terminology uniformity, and stylistic conti-nuity across the report.

The weights $\lambda_1, \lambda_2, \lambda_3 \geq 0$ satisfy $\lambda_1 + \lambda_2 + \lambda_3 = 1$, and can be tuned to reflect user or institutional priorities.

VII. Innovative Content

A. Overview of Novel Contributions

Open Deep Research introduces a collection of technically substantive innovations that collectively distinguish it from prior work in research automation, multi-agent orchestration, and LLM-based knowledge synthesis. These innovations span system architecture, retrieval methodology, citation manage-ment, prompt engineering strategy, and context management.

This section elaborates each contribution in detail, articulating both its technical design and its practical significance.

B. Summary of Innovative Contributions

Table I summarizes the six primary innovations of Open Deep Research and their relationship to specific challenges identified in the problem statement.

TABLE I: SUMMARY OF KEY INNOVATIVE CONTRIBUTIONS

Innovation	Description	Challenge Addressed
Graph-Based Multi-Agent Orchestration	Stateful directed graph of independent agents with check-point recovery	Scalability, modularity
Credibility- Weighted Retrieval	Three-component relevance scoring with recency decay and source authority signals	Source quality variance
Claim-Level Citation Tracking	Provenance graph linking every report claim to its sup-porting passage	Citation integrity, hallucination
Format-Aware Dynamic Structuring	Query-conditioned hierarchical report outline generation	Structural adaptability
Semantic Chunking	Coherence-aware chunk boundary detection with sliding overlap	Context preservation
Asynchronous Parallel Execution	Critical-path scheduling of concurrent agent execution	Computational efficiency

VIII. DESIGN

A. Architectural Overview

Open Deep Research is an open-source modular architec-ture, consisting of three tiers, which function as separate but connected nodes in graph-based architecture. A graphic representation of the overall architecture design of the Open Deep Research system and data transfer among modules can be seen in Figure.Besides, users' input through the user input interface also demonstrates the interaction among modules for generating the research report.

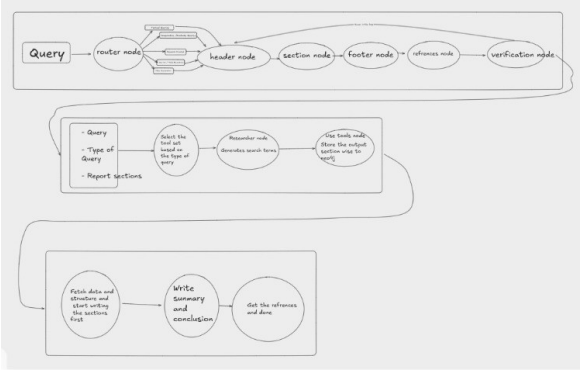

Fig. 1. Modular Graph-based Architecture of Open Deep Research

This architecture consists of three layers of functionalities that convert the user's input into complete research reports. First, there is the user interface layer designed as the front-end layer in which the user enters the research type to explore, unique structural requirements of the research report and report generation process monitored by the user in real time. This interface has been built using Streamlit and offers users to interact with its controls without regardless of their technical knowledge level.

The orchestrator engine works to coordinate the activities of the three main modules of the graphs using asynchronous work- flows. The back-end layer offers communication among modules, controls APIs of third parties including OpenAI, Groq, and Tavily, and helps create seamless development of reports, even if one or more modules experience any delays in developing the output.

Finally, the data layer serves as the long-term storage area of the final report along with two interim reports.

B. Section Graph Module

Depending on the research topic and user preferences, the Section Graph is utilized as the structural planning agent to generate hierarchical report layouts. A Research Question Q is represented by the Section of the Graph in this manner:

$$S = \{s_1, s_2, \ldots, s_n\} \tag{3}$$

s_i is a subset of each section. $s_i = \{ss_{i,1}, ss_{i,2}, \ldots, ss_{i,m}\}$. To maintain a well-organized research content, the structure is dynamically modified through adaptive logic that adjusts to the complexity of the question and user feedback.

This module utilizes research format options such as litera-ture review, comparative analysis, survey paper and technical report. Nevertheless, it is customized to meet the needs of each inquiry. A research question for emerging technologies may contain sections on the current state, recent advancements, challenges, and future prospects, or it may include sections for each item being compared, summarized, undated, but recommended. Other examples are listed below.

The Section Graph will consider both the purpose and target audience of the research. Additionally, Technical reports or policy briefs have varying structure from an academic paper. Why? This module will analyze the query, along with any other parameters provided by the user, and create an organizational structure to guide further research and writing.

C. Research Graph Module

The Research Graph is a sophisticated tool for gathering information, capable of producing complex query expansions (fexpand) and accessing data from various sources. (Q_{expanded}) is the result of calling the fexpand module from s_i.

$$Q_{\text{expanded}} = f_{\text{expand}}(s_i, \text{context})(Q_1, Q_2, \ldots, Q_K) \quad (4)$$

By utilizing the high level section topic (s_i), the expan-sion process generates numerous sub-queries that can retrieve relevant information from various sources using semantic understanding.

The process of retrieving data from various sources by combining results using different APIs and databases can be mathematically described as follows:

$$R = \bigcup_{j=1}^{K} \text{retrieve}(Q_j, \{API_1, API_2, \ldots, API_m\}) \quad (5)$$

In order to retrieve documents, they are ranked and filtered using relevance scores that consider semantic similarity or source credibility.

$$\text{Relevance}(d, q) = \alpha \cdot \text{semantic\_sim}(d, q) + \beta \cdot \text{credibility}(d) \quad (6)$$

To ensure the quality of documents is taken into account when retrieving them, semantic similarity (0.7) and credibility (0.3) are given weights to *alpha* and */beta* respectively.

The level of semantic similarity in a document is determined
by how well it answers the specific question, and the credibility of that document depends on several factors, including the authority of its source, the publication site, citation pattern, etc. to ensure that any documents that are returned are of high quality and come from authoritative sources.

The Research Graph retains metadata for all sources that retrieve documents, including the URL, location, published date, author, and retrieval time. This information is stored in the Research Data Structure (RGD).

D. Writer Graph Module

The Writer Graph is a method of converting recovered data into logical, citation-supported, and natural text. For every section of research data, the module generates R_i. Hence,

$$C_i = f_{\text{synthesize}}(R_i, s_i, style_{\text{params}}) \quad (7)$$

The synthesis function employs LLaMA-3.3-70B to gener-ate and paraphrase content in a manner that is consistent with the original style and correct in the information provided. This module is tasked with synthesising information from various sources and viewpoints, some written in more detail than others, and some in different writing styles.

Each statement produced contains citation metadata that provides a direct link from the claims to the sources that support them: (i.e.

$$C_i = \{(text_1, [cite_1, cite_2, \ldots]), (text_2, [cite_3, \ldots]), \ldots\} \quad (8)$$

At the claim level or sentence level, citation tracking is done to identify and correlate individual statements or assertions with the sources they are taken from. The Writer Graph is capable of handling excessive information, which can prevent the duplication of similar data from various sources. If various sources offer conflicting information, the module will exhibit a range of opinions or identify areas of doubt, while still maintaining intellectual credibility in the summation.

This module is designed to ensure uniformity of terminol-ogy, notation and style throughout the report. Technical termi-nology is consistently employed, abbreviations are elucidated when introduced for the first time, and the narrative remains coherent, even in parts that reference other sources.

E. Integration and Orchestration

All three modules communicate asynchronously through an integration layer. Assuming modules can function simultane-ously, the system's efficiency is increased:.

$$workflow = Section \rightarrow Research \rightarrow Writer \qquad (9)$$

Although writing is a process that moves from research to writing, it also involves varying levels of complexity. While the Section Graph is in the process of finalizing later sections of the outline, the Research Flow can initiate work on the initial chapters. The Writer Graph is capable of beginning to compile the final research results, while the Research Fig. is still gathering data for later parts of the paper.

Unusual and unpredictable external APIs can be handled with error handling, retry strategies, and logging to ensure reliable operation. Recovering from checkpoints and avoiding unnecessary computation during iterative refinement is pos-sible with stateful design. When an individual changes one area, they must reprocess it while leaving the rest of the report unchanged.

This orchestration approach regards the research process as an iterative and flexible workflow process.. It is possible for the system to react to intermediate results, adjusting its strategy accordingly rather than following a specific method using only the information available.

IX. IMPLEMENTATION

A. Technology Stack

A group of advanced and easy-to-use software tools and libraries have been chosen based on their performance and dependability. The addition of support for the Groq API in LLaMA-3.3-70B has resulted in improved speed and accuracy when it comes to text generation tasks. In addition, it's very good at reasoning and summarizing tasks so well suited to synthesized research materials.).

Through the use of user-friendly interfaces for visual design and programming, LangFlow offers a comprehensive orches-tration workflow that allows for intricate agent interactions while still providing full management of execution logic.

The stateful workflow structure of LangFlow has a signif-icant advantage in supporting research projects that require contextual continuity between processing steps.

PyPDF software is capable of extracting text from PDF files while maintaining the structure of the text, including section headings, paragraph structures and lists. To obtain a precise context for the information it extracts, PyPDF must be used to understand how source material is organized and flown into synthesized documents.

The Tavily API is essentially' multi-source' Internet information-retrieval system that merges output from multiple search engines and databases into ONE single, easy-to-use interface. A combined search result interface on the Web eliminates the burden of searching for different information, ensuring consistency between query and answer formats across multiple platforms.

B. Query Processing Pipeline

The research process is based on a multi-stage pipeline that transforms user inquiries into detailed research reports. Every step advances the results of previous steps while simul-taneously upholding adaptability by considering intermediate outcomes. First, the focus is on understanding the questions. User input is utilized for semantic analysis to determine the intent, scope, purpose, and domain of the research. The strategy establishes central themes and makes decisions regarding what type of research is being sought (exploratory, comparative, analytical), while also observing the limitations and preferences specified by the user. The Section Graph derives from this knowledge to develop a preliminary report according to the inquiry, ensuring that the content is consistent within the organization and aligned with the intended purpose. Stage two handles information retrieval. For each section defined in stage one, the Research Graph generates semantic sub-questions that respond to specific topics. Multiple queries are processed in parallel using APIs to filter and categorize results according to relevance and credibility. It extracts rele-vant passages from recovered documents and stores them with complete metadata information about the source, context, and retrieval parameters. This stage continues until adequate and high-quality information is gathered for each section.

Stage three performs content synthesis. The Writer Graph processes the retrieved data and uses semantic similarity to cluster shared topics and viewpoints. Unnecessary information is eliminated, and potential contradictions are given special attention. The system then creates cohesive narrative content that incorporates insights from a variety of sources while main-taining proper credit through fine-grained citation tracking. Stylistic consistency is enforced throughout the entire report so that the final document reads as a cohesive whole and not as scattered individual pieces.

In stage four, the final report is verified and delivered. Con-tent is examined for accuracy, ensuring that claims conform to their cited sources and that there are no unproven statements in the report. Citation completeness is also examined, verifying that all factual allegations contain accurate references. The assessment of structural coherence ensures that the sections follow logical progressions throughout the narrative. Once these checks are satisfied, the report is formatted accordingly and made available to the user.

C. Prompt Engineering Strategy

Proper prompt design is crucial for obtaining high-quality output from LLMs. We employ well-crafted prompts utilizing four distinct categories of important factors that guide the model to create relevant output.

This setting establishes the academic scene and research environment. This prepares the "contextual model" to function properly in a formal and technical research setting. Prompts, for example, could require content to be written for a scientific audience with appropriate aptitude in the field of research or require explanations that have graduate-level knowledge of concepts.

The task specification describes the specific task that must be completed, such as encapsulating an entire passage, compil-ing data from diverse sources, or obtaining specific categories of information, such as significant discoveries or methodolog-ical particulars. By defining tasks, the model can identify what is required and what is not.

Facts, citation requirements, and style choices are bound by certain constraints. The prompt is unambiguously directed to instruct the model to assert all claims using appropriate references from the sources mentioned and to maintain proper academic writing practices. These limits ensure that unsup-ported data is excluded by the model and prevent the use of incorrect information or deviations from proper academic writing.

The answers are arranged according to the specified out-put format. Such parameters are easily parsed and can be integrated into the bigger workflow. By requesting specific forms of structures using JSON and/or markdown format, it is guaranteed that the necessary information can be gathered and transmitted to later processing stages.

D. Chunking and Context Management

Many advanced methods are used for chunking large re-search documents to ensure completeness while saving com-putational power. By using chunking as a tool to implement semantic-aware segmentation on large research documents, we can ensure that the content's logical boundaries are taken into account and define coherence

$$\text{Chunk}_i = \{t_j \mid j \in [\text{start}_i, \text{end}_i],\ \text{coherence}(t_{\text{start}j}, t_{\text{end}j}) > \theta\} \qquad (10)$$

where: $\theta = 0.75$ represents at least some semantic cohe-sion.The aim of this approach is to categorize the documents in chunks based on logical constraints, keeping in mind the context of related information and refraining from breaking them down into smaller fragments at a fixed size per token.

The chunking algorithm detects natural breaks in the struc-ture of a document, such as section breaks, paragraph breaks or changes in subject area, and allocates chunks to fill an LLM context window while including elaboration. Coherent input to an LLM can result in high-quality output, as described above illustrates.

The context management system expands the concept of contextual information beyond the currently processed chunk. By providing contextual continuity across multiple chunks, the system enables better comprehension of the relationships between sections and improves the consistency and accuracy of research outputs.

X. DISCUSSION

A. Research Contributions

Open Deep Research introduces a transparent and modular framework for automated research synthesis using graph-based multi-agent orchestration. Unlike conventional monolithic AI research systems, the proposed framework decomposes the research workflow into specialized modules responsible for structural planning, information retrieval, and content synthesis.

The primary contribution of this work is the integration of three independent yet interconnected graph modules:

- Section Graph for dynamic report structure generation

- Research Graph for adaptive multi-source retrieval and verification
- Writer Graph for citation-aware academic content syn-thesis

This modular decomposition improves scalability, inter-pretability, and maintainability while enabling asynchronous execution across different stages of the workflow.

A second major contribution is the implementation of claim-level citation tracking. Each generated statement is associated with supporting source metadata, enabling users to trace fac-tual claims back to their original references. This significantly improves transparency and reduces the risk of hallucinated or unverifiable content.

Another important contribution is the credibility-aware re-trieval mechanism that combines semantic relevance, source authority, and recency scoring. Unlike conventional retrieval systems that rely primarily on keyword similarity, the proposed method prioritizes academically reliable and contextually rel-evant information.

The framework also introduces semantic-aware chunk-ing and context preservation strategies that improve long-document processing within bounded LLM context windows. This enables coherent synthesis across large collections of research material.

B. Advantages of the Proposed Framework

The proposed architecture provides several practical advan-tages over existing research automation systems.

Modularity and Extensibility: Each module operates inde-pendently and can be upgraded or replaced without affecting the overall system architecture. This allows future integration of domain-specific agents, fact-checking modules, or alterna-tive language models.

Transparency and Traceability: The system maintains explicit source metadata and citation mappings throughout the pipeline, enabling verification of generated claims and improving academic integrity.

Parallel Processing Capability: Independent sections of a report can be researched and synthesized concurrently, significantly reducing report generation time for large-scale research tasks.

Adaptive Structuring: The framework dynamically gener-ates report structures based on research objectives and output formats rather than relying on static templates.

C. Limitations

Despite its advantages, Open Deep Research has several limitations.

The framework depends heavily on external APIs and third-party language models. Network latency, API rate limits, and service outages can affect system performance and reliability. Although citation tracking improves factual grounding, LLM-generated content may still contain hallucinations or subtle factual inconsistencies. Human verification remains important for high-stakes academic applications.

The system is designed primarily for general-purpose in-terdisciplinary research. Highly specialized domains such as medicine, law, or biotechnology may require domain-specific retrieval systems and fine-tuned language models for optimal performance.

In addition, large-scale deployment can incur significant computational and API costs due to repeated retrieval, em-bedding generation, and long-context synthesis operations.

D. Future Work

Several directions exist for extending the framework. Future work will focus on integrating automated fact-verification modules capable of validating claims against trusted academic databases. Additional improvements include collaborative multi-user research environments, semantic visu-alization systems, and domain-specific fine-tuning for special-ized research applications.

The integration of local inference models, retrieval caching, and optimized workflow scheduling may further reduce opera-tional costs and improve scalability. Future versions may also incorporate multimodal research synthesis involving tables, figures, charts, and audiovisual content.

XI. CONCLUSION

In this paper, we propose a novel modular framework named Open Deep Research, which abstracts and models large language models and the multi-agent architecture based on graphs to make research synthesis more convenient.

The study has three distinct modules: creating an outline, graphing the research and final report. Note: The research process can be automated by Segment, which is transparent and scalable.

The research time can significantly be reduced, and the work will still be academically valid, if an open-source modular framework is used due to the presence of citations, verification of sources, and multiple sources.

Traditional one-unit research tools are not modular and would not allow for the development of domains or the creation of a community-driven approach as with Open Deep Research.
Students, researchers and professionals are using Open Deep Research to make conducting in-depth interdisciplinary research more efficient. Moreover, it also helps in preserving transparency and traceability to ensure academic integrity, and uses artificial intelligence to increase computation efficiency. Upcoming work include ”enhancement of fact-verification mechanisms, expansion of the domains at which. is available and development of collaborative features that make possible
team-based research projects.

ACKNOWLEDGMENT

The authors express sincere gratitude to Dr. Padma D. Adane for her invaluable guidance, expertise, and continuous support throughout this project. We also thank Shri Ramdeob-aba College of Engineering and Management, Nagpur, for providing essential resources, facilities, and an environment conducive to research and innovation.

REFERENCES

[1] R. Mihalcea and P. Tarau, “TextRank: Bringing Order into Texts,” in Proceedings of EMNLP, 2004.
[2] M. F. Porter, “An algorithm for suffix stripping,” Program, vol. 14, no. 3, pp. 130-137, 1980.
[3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understand-ing,” arXiv:1810.04805, 2019.
[4] OpenAI, “GPT-4 Technical Report,” arXiv:2303.08774, 2023.
[5] H. Touvron et al., “LLaMA: Open and Efficient Foundation Language Models,” arXiv:2302.13971, 2023.
[6] A. Cohan et al., “A Discourse-Aware Attention Model for Abstractive Summarization of Long Documents,” arXiv:1804.05685, 2018.
[7] Langflow Blog, “Query, Split, Search, Synthesize: Automating Deep Research with JigsawStack and Langflow,” July 2025.
[8] OpenAI Cookbook, “Deep Research API with the Agents SDK,” November 2024.
[9] Reddit Discussion, “I built Open Source Deep Research: here’s how it works,” April 2025.
[10] Together AI Blog, “Open Deep Research,” April 2025.
[11] A. Vaswani et al., “Attention is All You Need,” in Advances in Neural Information Processing Systems, 2017.
[12] T. Brown et al., “Language Models are Few-Shot Learners,” in Advances in Neural Information Processing Systems, 2020.