Novelty Assessment Report

Paper: WebWeaver: Structuring Web-Scale Evidence with Dynamic Outlines for Open-Ended Deep Research
PDF URL: https://openreview.net/pdf?id=MtNCJjlrKt
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-30

Abstract

This paper tackles \textbf{open-ended deep research (OEDR)}, a complex challenge where AI agents must synthesize vast web-scale information into insightful reports. Current approaches are plagued by dual-fold limitations: static research pipelines that decouple planning from evidence acquisition and monolithic generation paradigms that include redundant, irrelevant evidence, suffering from hallucination issues and low citation accuracy. To address these challenges, we introduce \textbf{WebWeaver}, a novel dual-agent framework that emulates the human research process. The planner operates in a dynamic cycle, iteratively interleaving evidence acquisition with outline optimization to produce a comprehensive, citation-grounded outline linking to a memory bank of evidence. The writer then executes a hierarchical retrieval and writing process, composing the report section by section. By performing targeted retrieval of only the necessary evidence from the memory bank via citations for each part, it effectively mitigates long-context issues and citation hallucinations. Our framework establishes a new state-of-the-art across major OEDR benchmarks, including DeepResearch Bench, DeepConsult, and DeepResearchGym. These results validate our human-centric, iterative methodology, demonstrating that adaptive planning and focused synthesis are crucial for producing comprehensive, trusted, and well-structured reports.

Core Task Landscape

This paper addresses: **Synthesizing Web-Scale Information into Comprehensive Research Reports**
A total of **11 papers** were analyzed and organized into a taxonomy with **8 categories**.

Taxonomy Overview
The research landscape has been organized into the following main categories:
- **AI-Driven Deep Research Systems**
- **Domain-Specific Multimodal Foundation Models**
- **Automated Domain-Specific Report Generation**
- **Database and Application Report Tools**
- **Social Media and Web-Scale Data Monitoring**
- **Web-Scale Discovery and Open Science Infrastructure**

Complete Taxonomy Tree
- Synthesizing Web-Scale Information into Comprehensive Research Reports Survey Taxonomy
- AI-Driven Deep Research Systems
  - Dynamic Multi-Agent Research Frameworks ★ (2 papers)
  - [0] WebWeaver: Structuring Web-Scale Evidence with Dynamic Outlines for Open-Ended Deep Research (Anon et al., 2026) View paper
  - [2] Webthinker: Empowering large reasoning models with deep research capability (Li Xiaoxi, 2025) View paper
  - Geo-Temporal Research Systems (1 papers)
  - [5] A Vision for Geo-Temporal Deep Research Systems: Towards Comprehensive, Transparent, and Reproducible Geo-Temporal Information Synthesis (Martins, 2025) View paper
- Domain-Specific Multimodal Foundation Models
  - Medical Imaging Foundation Models (1 papers)
  - [1] Towards generalist foundation model for radiology by leveraging web-scale 2d&3d medical data (Chaoyi Wu, 2025) View paper
  - Biological Sequence Understanding Models (1 papers)
  - [3] Rna-gpt: Multimodal generative system for rna sequence understanding (Xiao, 2024) View paper
- Automated Domain-Specific Report Generation (1 papers)
  - [6] Automatically Generating Macro Research Reports from a Piece of News (Hu Wen-xin, 2022) View paper
- Database and Application Report Tools (2 papers)
  - [10] A report generator for database and web applications (Woei-Kae Chen, 2012) View paper
  - [11] RESEARCH AND DEVELOPMENT OF INFORMATION SYSTEM FOR GENERATION OF FLEXIBLE STATISTICAL REPORTING (Saykha S.-A. Khajimuradova, 2024) View paper
- Social Media and Web-Scale Data Monitoring (1 papers)
  - [8] Social media data aggregation and mining for internet-scale customer relationship management (Stephen Wan, 2015) View paper
- Web-Scale Discovery and Open Science Infrastructure (3 papers)
  - [4] Web scale discovery services (Vaughan, 2011) View paper
  - [7] Open science at Web-scale: optimising participation and predictive potential: consultative report (Lyon, 2009) View paper
  - [9] Open science at web-scale: Optimising participation and predictive potential (Elizabeth Lyon, 2009) View paper

Narrative

Core task: synthesizing web-scale information into comprehensive research reports. The field encompasses several distinct branches that reflect different approaches to handling large-scale information synthesis. AI-Driven Deep Research Systems focus on autonomous agents and multi-agent frameworks that orchestrate complex research workflows, often combining retrieval, reasoning, and iterative refinement. Domain-Specific Multimodal Foundation Models develop specialized architectures for fields like radiology or genomics, where domain expertise must be encoded into the model itself. Automated Domain-Specific Report Generation and Database and Application Report Tools address more structured synthesis tasks, generating reports from databases or application logs. Social Media and Web-Scale Data Monitoring targets real-time streams and social platforms, while Web-Scale Discovery and Open Science Infrastructure emphasizes indexing, search, and open-access scholarly communication. These branches vary in their emphasis on autonomy versus structure, domain specialization versus generality, and real-time monitoring versus retrospective synthesis.

Particularly active lines of work include dynamic multi-agent systems that decompose research into subtasks and coordinate specialized agents, as well as domain-specific foundation models that integrate multimodal data for expert-level synthesis. WebWeaver[0] sits squarely within the AI-Driven Deep Research Systems branch, specifically among Dynamic Multi-Agent Research Frameworks. It shares this space with WebThinker[2], which similarly emphasizes iterative reasoning and web-scale retrieval for comprehensive report generation. Compared to WebThinker[2], WebWeaver[0] appears to place greater emphasis on orchestrating multiple specialized agents rather than relying on a single reasoning loop. This contrasts with approaches like Geo-Temporal Deep Research[5], which targets spatiotemporal analysis, or domain-specific models such as Generalist Radiology Foundation[1] and RNA-GPT[3], which prioritize vertical depth over horizontal web-scale breadth. The central tension across these branches remains balancing autonomy and control, depth and coverage, and domain expertise with general-purpose reasoning.

Related Works in Same Category

The following **1 sibling papers** share the same taxonomy leaf node with the original paper:

1. Webthinker: Empowering large reasoning models with deep research capability

Authors: Li Xiaoxi, Jin, Jiajie, Xiaoxi Li, Dong, et al. (21 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Large reasoning models (LRMs), such as OpenAI-o1 and DeepSeek-R1, demonstrate impressive long-horizon reasoning capabilities. However, their reliance on static internal knowledge limits their performance on complex, knowledge-intensive tasks and hinders their ability to produce comprehensive research reports requiring synthesis of diverse web information. To address this, we propose WebThinker, a deep research agent that empowers LRMs to autonomously search the web, navigate among web pages, and...

Relationship Analysis

Both papers belong to the Dynamic Multi-Agent Research Frameworks category, employing specialized agents with iterative planning and evidence synthesis for comprehensive report generation. They overlap in their use of dynamic search-planning cycles and hierarchical writing strategies to synthesize web-scale information into research reports. However, WebWeaver emphasizes a dual-agent architecture with explicit outline optimization and citation-grounded memory retrieval, while WebThinker focuses on empowering large reasoning models with deep web exploration capabilities and autonomous think-search-and-draft strategies, incorporating reinforcement learning for tool utilization optimization.

Contributions Analysis

Overall novelty summary. WebWeaver introduces a dual-agent framework for open-ended deep research, combining a planner that iteratively refines outlines with evidence acquisition and a writer that performs hierarchical synthesis. The paper positions itself within the Dynamic Multi-Agent Research Frameworks leaf of the taxonomy, which contains only two papers total. This represents a relatively sparse research direction within the broader field of AI-driven research systems, suggesting the work addresses an emerging rather than saturated problem space.

The taxonomy reveals that AI-Driven Deep Research Systems branch into dynamic multi-agent approaches versus geo-temporal systems, with WebWeaver belonging to the former. Neighboring branches include Domain-Specific Multimodal Foundation Models (medical imaging, biological sequences) and Automated Domain-Specific Report Generation, which handle structured synthesis tasks. WebWeaver's emphasis on web-scale generality and agent orchestration distinguishes it from domain-specific models and static report generators, though it shares conceptual ground with systems emphasizing iterative reasoning and retrieval coordination.

Among 19 candidates examined across three contributions, no clearly refuting prior work was identified. The core dual-agent framework examined 9 candidates with 0 refutations, the dynamic research cycle examined 7 candidates with 0 refutations, and the memory-grounded synthesis examined 3 candidates with 0 refutations. This suggests that within the limited search scope, the specific combination of dual-agent orchestration, iterative outline refinement, and citation-driven hierarchical writing appears relatively unexplored, though the individual components may have precedents in related work.

Based on the top-19 semantic matches examined, WebWeaver's approach appears novel in its specific architectural choices, particularly the separation of planning and writing agents with citation-grounded memory. However, the limited search scope and sparse taxonomy leaf mean this assessment reflects novelty within a narrow comparison set rather than exhaustive field coverage. The sibling paper WebThinker likely represents the closest conceptual neighbor, warranting careful comparison of architectural distinctions.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: WebWeaver dual-agent framework for open-ended deep research

Description: The authors propose WebWeaver, a dual-agent system comprising a planner and a writer. The planner iteratively interleaves evidence acquisition with outline optimization to produce a citation-grounded outline, while the writer performs hierarchical retrieval and section-by-section synthesis to compose the final report.

This contribution was assessed against **9 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. An Efficient Dual-Agent Framework for Generating and Evaluating Synthetic Aviation Safety Reports using Large Language Models

URL: View paper

Brief Assessment

Aviation Safety Reports[23] focuses on generating and evaluating synthetic aviation safety reports using a dual-agent framework with specialized generator and evaluator agents. This differs fundamentally from WebWeaver's dual-agent system (planner and writer) designed for open-ended deep research involving web-scale evidence acquisition, dynamic outline optimization, and hierarchical report synthesis. The application domains and agent roles are distinct.

2. Probabilistic Economy. Unified Market Theory

URL: View paper

Brief Assessment

Probabilistic Economy[31] discusses economic market theory and does not address dual-agent frameworks, automated research systems, or AI-driven report generation. The candidate paper operates in an entirely different domain (economics) with no overlap with WebWeaver's contributions to open-ended deep research.

3. The Landscape of Medical Agents: A Survey

URL: View paper

Brief Assessment

Medical Agents Landscape[30] is a survey paper about medical AI agents, not a framework for automated research report generation. It does not present a dual-agent system for web-scale evidence synthesis.

4. A Hierarchical Tree-based approach for creating Configurable and Static Deep Research Agent (Static-DRA)

URL: View paper

Brief Assessment

Static-DRA[27] uses a hierarchical tree-based static workflow with supervisor, independent, and worker agents, fundamentally different from WebWeaver's dynamic dual-agent framework where a planner iteratively interleaves evidence acquisition with outline optimization, and a writer performs hierarchical retrieval and synthesis.

5. S3-Net: A Self-Supervised Dual-Stream Network for Radiology Report Generation.

URL: View paper

Brief Assessment

S3-Net[28] focuses on radiology report generation from medical images using self-supervised learning, not on dual-agent frameworks for web-scale research or automated report generation from web evidence.

6. Enhancing Research Productivity Through Agentic AI Workflows: A Multi-Agent Framework for Intelligent Research Assistance

URL: View paper

Brief Assessment

Agentic Research Workflows[29] focuses on literature review automation for academic research (literature discovery, source assessment, report generation for academic papers), while the original paper addresses open-ended deep research requiring web-scale information synthesis with dynamic outline optimization and citation-grounded writing. The candidate's dual-agent system (search agent + drafting agent) serves a fundamentally different purpose than WebWeaver's planner-writer architecture for OEDR tasks.

7. A Composable Agentic System for Automated Visual Data Reporting

URL: View paper

Brief Assessment

Composable Agentic Reporting[26] focuses on automated visual data reporting with a multi-stage pipeline for visualization and interactive reports, not on open-ended deep research with web-scale evidence synthesis and citation-grounded outline generation.

8. Reflections & Resonance: Two-Agent Partnership for Advancing LLM-based Story Annotation

URL: View paper

Brief Assessment

Reflections and Resonance[24] focuses on story annotation and narrative reconstruction tasks, not open-ended deep research or web-scale evidence synthesis for report generation.

9. Aviation safety qa dataset for extracting knowledge from incident reports

URL: View paper

Brief Assessment

Aviation Safety QA[25] focuses on creating a question-answering dataset for extracting safety facts from aviation incident reports, not on developing a dual-agent framework for automated research report generation. The candidate addresses domain-specific QA dataset construction, while the original contribution concerns iterative planning-writing architectures for comprehensive research synthesis.

Contribution 2: Dynamic research cycle with iterative evidence acquisition and outline optimization

Description: The authors introduce a planning mechanism that iteratively interleaves searching for evidence with optimizing the outline, allowing emergent findings to reshape the research direction. This contrasts with static outline-guided or search-then-outlining approaches that decouple planning from discovery.

This contribution was assessed against **7 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Knowledge acquisition for visual question answering via iterative querying

URL: View paper

Brief Assessment

Iterative Querying VQA[17] focuses on visual question answering through iterative queries to external knowledge sources, not on research report generation with outline optimization. The domains and tasks are fundamentally different.

2. Development of an evidence‑based review with recommendations using an online iterative process

URL: View paper

Brief Assessment

Evidence-Based Review Process[18] focuses on medical literature review methodology with iterative author review cycles, not on AI agent systems that dynamically interleave web searching with outline optimization for open-ended research tasks.

3. Deep research agents: A systematic examination and roadmap

URL: View paper

Brief Assessment

Deep Research Agents[12] focuses on systematic examination of deep research agent architectures, including search engines, tool use, and workflow optimization. While it discusses dynamic workflows and planning strategies, it does not specifically address the iterative

interleaving of evidence acquisition with outline optimization for research report generation that is central to the original paper's contribution.

4. A proposed evidence-guided algorithm for the adjustment and optimization of multi-function articulated ankle-foot orthoses in the clinical setting

URL: View paper

Brief Assessment

Ankle-Foot Orthoses Algorithm[14] addresses iterative adjustment of medical devices (AFOs) based on observed gait deviations, not research planning or outline optimization. The iterative process here involves clinical device tuning, not evidence-guided research methodology.

5. Pace-of-life syndromes: a framework for the adaptive integration of behaviour, physiology and life history

URL: View paper

Brief Assessment

Pace-of-Life Syndromes[15] addresses biological frameworks for integrating behavior, physiology, and life history in evolutionary ecology, not AI research methodologies or iterative planning systems for information synthesis.

6. Bayes-entropy collaborative driven agents for research hypotheses generation and optimization

URL: View paper

Brief Assessment

Bayes-Entropy Agents[13] focuses on scientific hypothesis generation through Bayesian inference and entropy-driven refinement, not on web-scale evidence gathering with dynamic outline optimization for report writing.

7. SplitWise Regression: Stepwise Modeling with Adaptive Dummy Encoding

URL: View paper

Brief Assessment

SplitWise Regression[16] focuses on stepwise regression with adaptive dummy encoding for numeric predictors, not on research planning or evidence acquisition workflows. The candidate addresses statistical modeling, while the original addresses research agent design.

Contribution 3: Memory-grounded hierarchical synthesis with citation-driven retrieval

Description: The authors design a writing process where the writer constructs the report section by section, retrieving only relevant evidence from a structured memory bank using citations embedded in the outline. This approach addresses long-context challenges and reduces hallucinations by focusing on pertinent evidence for each section.

This contribution was assessed against **3 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Hybrid Augmented Reasoning Interpretation (HARI) Framework for Massive Scientific Literature Semantic Retrieval Analysis

URL: View paper

Brief Assessment

HARI Framework[22] focuses on scientific literature semantic retrieval using knowledge graphs and multi-level recall approaches, not on hierarchical report writing with citation-driven evidence retrieval from a memory bank for open-ended research tasks.

2. SurveyG: A Multi-Agent LLM Framework with Hierarchical Citation Graph for Automated Survey Generation

URL: View paper

Brief Assessment

SurveyG[20] focuses on survey paper generation using hierarchical citation graphs to organize research papers by their structural relationships. The candidate does not address the specific challenge of long-context report writing with section-by-section evidence retrieval that WebWeaver tackles for open-ended deep research tasks.

3. The Next Phase of Scientific Fact-Checking: Advanced Evidence Retrieval from Complex Structured Academic Papers

URL: View paper

Brief Assessment

Scientific Fact-Checking[19] focuses on evidence retrieval from scientific papers for fact-checking, not on hierarchical report generation with citation-driven synthesis for open-ended research tasks.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] WebWeaver: Structuring Web-Scale Evidence with Dynamic Outlines for Open-Ended Deep Research View paper
- [1] Towards generalist foundation model for radiology by leveraging web-scale 2d&3d medical data View paper
- [2] Webthinker: Empowering large reasoning models with deep research capability View paper
- [3] Rna-gpt: Multimodal generative system for rna sequence understanding View paper
- [4] Web scale discovery services View paper
- [5] A Vision for Geo-Temporal Deep Research Systems: Towards Comprehensive, Transparent, and Reproducible Geo-Temporal Information Synthesis View paper
- [6] Automatically Generating Macro Research Reports from a Piece of News View paper
- [7] Open science at Web-scale: optimising participation and predictive potential: consultative report View paper
- [8] Social media data aggregation and mining for internet-scale customer relationship management View paper
- [9] Open science at web-scale: Optimising participation and predictive potential View paper
- [10] A report generator for database and web applications View paper
- [11] RESEARCH AND DEVELOPMENT OF INFORMATION SYSTEM FOR GENERATION OF FLEXIBLE STATISTICAL REPORTING View paper

- [12] Deep research agents: A systematic examination and roadmap View paper
- [13] Bayes-entropy collaborative driven agents for research hypotheses generation and optimization View paper
- [14] A proposed evidence-guided algorithm for the adjustment and optimization of multi-function articulated ankle-foot orthoses in the clinical setting View paper
- [15] Pace-of-life syndromes: a framework for the adaptive integration of behaviour, physiology and life history View paper
- [16] SplitWise Regression: Stepwise Modeling with Adaptive Dummy Encoding View paper
- [17] Knowledge acquisition for visual question answering via iterative querying View paper
- [18] Development of an evidenceâ□□based review with recommendations using an online iterative process View paper
- [19] The Next Phase of Scientific Fact-Checking: Advanced Evidence Retrieval from Complex Structured Academic Papers View paper
- [20] SurveyG: A Multi-Agent LLM Framework with Hierarchical Citation Graph for Automated Survey Generation View paper
- [21] GarmentAligner: Text-to-Garment Generation via Retrieval-augmented Multi-level Corrections View paper
- [22] Hybrid Augmented Reasoning Interpretation (HARI) Framework for Massive Scientific Literature Semantic Retrieval Analysis View paper
- [23] An Efficient Dual-Agent Framework for Generating and Evaluating Synthetic Aviation Safety Reports using Large Language Models View paper
- [24] Reflections & Resonance: Two-Agent Partnership for Advancing LLM-based Story Annotation View paper
- [25] Aviation safety qa dataset for extracting knowledge from incident reports View paper
- [26] A Composable Agentic System for Automated Visual Data Reporting View paper
- [27] A Hierarchical Tree-based approach for creating Configurable and Static Deep Research Agent (Static-DRA) View paper
- [28] S3-Net: A Self-Supervised Dual-Stream Network for Radiology Report Generation. View paper
- [29] Enhancing Research Productivity Through Agentic AI Workflows: A Multi-Agent Framework for Intelligent Research Assistance View paper
- [30] The Landscape of Medical Agents: A Survey View paper
- [31] Probabilistic Economy. Unified Market Theory View paper