Social-DeepWriter: An iterative retrieval-augmented framework for strategic social media content generation

Thieu Ngoc Mai, Bui Khac Hoai Nam, Vu Viet Hoang*,
Duong Van Linh, Nguyen Van Khoe

Viettel Artificial Intelligence and Data Services Center, Viettel Group, 7 Ton That Thuyet, Cau Giay, Hanoi, Vietnam.
*Corresponding author: hoangvv8@viettel.com.vn
Received 9 Sep. 2025; Revised 11 Nov. 2025; Accepted 10 Dec. 2025; Published 25 Dec. 2025.
DOI: https://doi.org/10.54939/1859-1043.j.mst.108.2025.136-142

ABSTRACT

Social media has become a critical domain for strategic communication, influencing public perception and supporting both civil and military operations. In high-tempo information environments, traditional manual content creation is often too slow and resource-intensive to meet the demands of real-time engagement. While large language models (LLMs) such as GPT-4 offer the capability to generate human-like text at scale, their reliance on static training data limits their contextual relevance, factual accuracy, and responsiveness to evolving mission needs. To overcome these limitations, this paper introduces Social-DeepWriter, an AI-enabled framework for the automatic generation of mission-aligned social media content. Built upon the Deep Research paradigm, Social-DeepWriter enhances traditional Retrieval-Augmented Generation (RAG) by incorporating iterative query refinement, multi-hop retrieval, and content evaluation, mirroring the layered reasoning of human analysts. We evaluate how factors such as retrieval quality, prompt design, and generation constraints influence the informativeness, coherence, and strategic fit of generated posts. Our findings highlight the potential of Social-DeepWriter to support dual-use communication scenarios, including military public affairs, psychological operations, and rapid-response campaigns, where accuracy, adaptability, and scalability are essential.

Keywords: Social media content generation; Large language models; Retrieval-augmented generation; Deep research; AI for public affairs.

1. INTRODUCTION

Social media has become a strategic communication battlefield, influencing public perception, shaping narratives, and supporting both civil and military operations. In modern information environments, speed, accuracy, and adaptability in content generation are essential. Manual content creation, while precise, is inherently slow and resource-intensive, posing limitations during high-tempo information campaigns or crisis response scenarios [1].

Recent breakthroughs in natural language processing (NLP), particularly large language models (LLMs) such as GPT-4, have enabled the automatic generation of fluent, human-like text [2]. However, these models are fundamentally limited by their static training data, which often renders them outdated, hallucination-prone, or context-insensitive in real-world applications [3].

This is a critical drawback in operational contexts where content must reflect current facts, mission objectives, and rapidly evolving information.

To address this, Retrieval Augmented Generation (RAG) technology, a class of AI systems that augment generation with live document retrieval, has emerged [4]. Traditional RAG systems retrieve documents once based on a static query and immediately generate a response. Although this approach helps reduce hallucination to some extent, it often struggles to retrieve sufficiently comprehensive information for queries or topics that require in-depth, multi-hop research. In such cases, a single-pass retrieval is inadequate, especially when the initial query fails to capture the

full scope of the information needed. Deep Research [5] extends the RAG framework by incorporating iterative querying, result evaluation, and refinement. The system actively assesses whether the retrieved documents sufficiently address the query. If not, it reformulates the query, searches again, or narrows its focus, emulating how human analysts conduct layered research. This enables the generation of responses that are not only coherent and factually grounded but also contextually complete and aligned with the information objectives of the task.

Based on the Deep Research idea, this paper proposes a Deep Writer that enables an AI framework for automatically generating social media posts aligned with mission goals. The system integrates five components: (1) topic-based query decomposition, (2) document retrieval from trusted sources, (3) summarization and alignment to communication intent, (4) evaluation of whether sufficient information is provided and (5) post-generation in social media format. We investigate how variables such as retrieval quality, prompt engineering, and generation constraints affect the informativeness, coherence, and strategic fit of the generated posts.

Our findings aim to contribute to dual-use communication technologies, offering potential applications in military public affairs, psychological operations, and rapid-response information campaigns, where autonomy, scalability, and message accuracy are mission-critical. Our research paper is organized into four sections: (1) Introduction: In this section, we cover current technologies and approaches, along with their characteristics and limitations. (2) Methodology: We provide a detailed description of our proposed model and solution. (3) Experiments: We conduct experiments to compare the results of our solution against current solutions. Additionally, we perform analyses on various aspects to highlight the significance of our research findings. (4) Conclusion: We summarize the main results of the study and discuss several directions for future research.

2. METHODOLOGY

2.1. Social-DeepWriter framework architecture

Figure 1 illustrates the system architecture of the proposed framework. Specifically, the core idea is to extend traditional RAG systems with an iterative process through multi-stage processing and reflective evaluation. The system architecture consists of six interconnected components that work synergistically to produce contextually relevant and strategically aligned social media content.

Let Q_0 represent the initial user query, D denote the corpus of retrievable documents, and P be the final generated social media post. The Social-DeepWriter framework can be formally defined as:

$$P = f_{\{Social-DeepWriter\}}(Q_0, D) \quad (1)$$

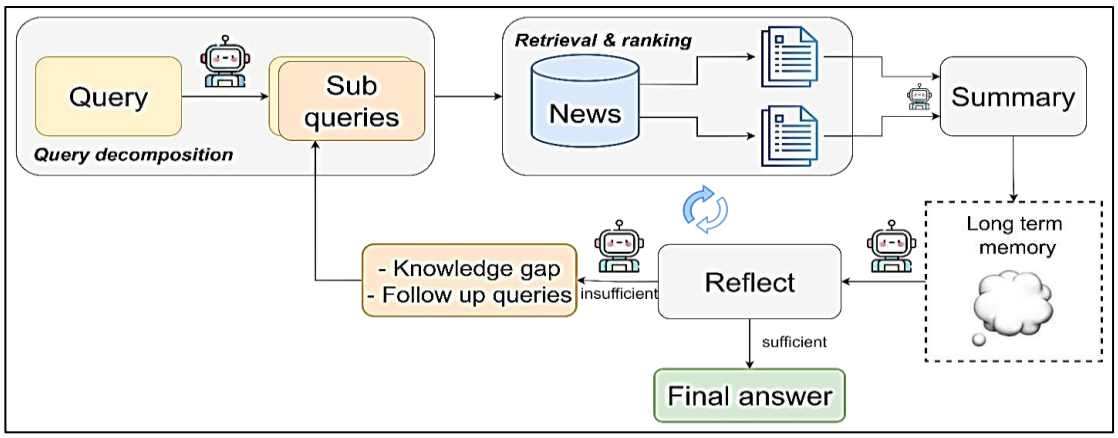

Figure 1. *The overview architecture of Social-DeepWriter, which includes five main sequence components, including query decomposition, document retrieval and ranking, summarization, long-term memory update, reflection and finalization.*

2.2. Query decomposition

Social-DeepWriter begins by decomposing a complex input query into a set of focused sub-queries. This step is critical for ensuring comprehensive coverage when addressing multifaceted topics. Decomposing such queries improves retrieval precision and recall by targeting specific dimensions of the broader research question.

Inspired by the works in multi-hop QA systems [6, 7], the query decomposition module transforms a complex input query Q_0 into a set of focused sub-queries $Q = \{q_1, q_2, \dots, q_n\}$ to ensure comprehensive information coverage. This decomposition follows a hierarchical approach that identifies the key information dimensions within the original query. Given an input query Q_0, the decomposition function is defined as:

$$Q = Decompose(Q_0) = \{q_i\}_{i=1}^{n} \tag{2}$$

where each sub-query q_i addresses a specific aspect of Q_0. The decomposition leverages a prompt-based approach using the LLM model:

$$q_i = LLM\left(prompt_{decomp}, Q_0, context_i\right) \tag{3}$$

2.3. Document retrieval and ranking

For each sub-query $q_i \in Q$, the system retrieves relevant documents from trusted sources using a hybrid retrieval approach combining semantic similarity and keyword matching:

$$R_i = Retrieve(q_i, D, k) = d_{i,1}, d_{i,2}, \dots, d_{i,k} \tag{4}$$

where R_i represents the top-k retrieved documents for the query q_i, ranked by relevance score:

$$score(d, q_i) = \alpha \cdot sim_{semantic}(d, q_i) + \beta \cdot sim_{keyword}(d, q_i) \tag{5}$$

Parameters α and β weight semantic similarity and keyword matching, respectively, where α + β = 1. The semantic similarity is computed using pre-trained embeddings [8].

2.4. Summarization

Retrieved documents are processed through an extractive-abstractive summarization pipeline to distill relevant information while preserving factual accuracy and strategic alignment [9]. Specifically, this module reduces token overhead, facilitates cross-document comparison, and improves the clarity and granularity of information stored in memory.

For each document set R_i, the summarization function produces condensed knowledge by employing a prompt-based LLM to generate the coherent summary:

$$s_i = LLM(prompt_{summ}, q_i, R_i) \tag{6}$$

2.5. Long-term memory update

The long-term memory module M maintains a persistent knowledge base that accumulates information across iterations. The summarized knowledge is stored in a long-term memory module that persistently accumulates information across iterations. This memory plays a central role in enabling Social-DeepWriter to reason over multiple sub-queries, integrate distributed facts, and build context-aware outputs using prompting-based LLM. For each sub-query $q_i \in Q$, the corresponding summary s_i is integrated into M using a structured update mechanism. The update function is defined as:

$$M_t = Update\ (M_{t-1}, s_i, q_i, metadata_i) \tag{7}$$

where M_t represents the memory state at iteration t, s_i is the summarized content from the document set R_i, and $metadata_i$ includes contextual attributes such as query relevance, source credibility, and timestamp. The update process employs a prompt-based LLM to ensure that new information is coherently merged, avoiding redundancy and preserving factual consistency:

$$M_t = LLM(prompt_{mem}, M_{t-1}, s_i, q_i) \tag{8}$$

This approach draws inspiration from memory-augmented architectures [10], enabling the system to retain and synthesize distributed facts across multiple sub-queries. The memory module supports reasoning by providing a global context, reducing redundant retrievals, and facilitating cross-query integration. The design mirrors episodic memory systems in reflective agents, ensuring that the accumulated knowledge remains accessible for subsequent reflection and content generation.

2.6. Reflection and finalization

The reflection module evaluates the sufficiency of the accumulated knowledge in M to address the original query Q_0, employing metacognitive reasoning to assess completeness and relevance. Using a prompt-based LLM, the system analyzes the memory state:

$$Eval_{score} = LLM(prompt_{rejlect}, M_t, Q_0) \quad (9)$$

where $Eval_{score}$ quantifies the coverage, factual accuracy, and strategic alignment of the stored information. If gaps are detected (e.g., incomplete sub-topic coverage or missing evidence), the system generates refined sub-queries q_i' to re-enter the query–retrieve–summarize cycle:

$$Q' = LLM(prompt_{refine}, M_t, Q_0) \quad (10)$$

This iterative process continues until $Eval_{score}$ exceeds a predefined threshold, ensuring comprehensive and mission-aligned content. Upon satisfaction, the finalization step synthesizes the memory contents into a coherent social media post P:

$$P = LLM(prompt_{finalize}, M_t, Q_0, constraint) \quad (11)$$

where *constraints* include platform-specific requirements (e.g., token limits). This reflective and iterative approach ensures that the generated content is factually grounded, contextually complete, and strategically aligned with the mission objectives. This reflection module operationalizes metacognitive reasoning [11], allowing the system to identify its own limitations and take corrective actions.

3. EXPERIMENTS

3.1. Experiment setup

Dataset: The experiments were conducted using a corpus of 16,000 Vietnamese documents spanning six domains: Economy, Politics, Health, Travel, Life, and Technology, to ensure broader coverage of real-world social media contexts. The dataset was curated from trusted sources, including news archives, government publications, and verified social media posts, ensuring high-quality and diverse content for retrieval. We evaluated the Social-DeepWriter framework against baseline models, including traditional Retrieval-Augmented Generation (RAG) systems across various LLMs size such as Qwen3-8b, Qwen3-14b, and Qwen3-32b. Two retrieval methods were tested: BM25 (keyword-based) and BGE (embedding-based) for document ranking

Evaluation metrics: We assessed the generated social media posts across five dimensions—Consistency, Coherence, Relevance, Fluency, and Depth, using a 1–5 scale. Evaluations were performed by both an LLM (Gemini 2.5 Flash) and a panel of human annotators with expertise in social media communication and strategic messaging.

3.2. Main results

Table 1 presents the overall performance of Social-DeepWriter compared to baselines, evaluated by the LLM (Gemini 2.5 Flash). Social-DeepWriter consistently outperformed traditional RAG and standalone LLMs across all model sizes and retrieval methods. The best performance was achieved with Qwen3-14b using BM25 retrieval, yielding an average score of 4.387 across metrics, a 2.17% improvement over traditional RAG.

Table 1. *Overall performance comparison.*

LLM	Retrieval	Social-DeepWriter	Traditional RAG	Improvement
Qwen3-8B	BM25	4.266	4.184	+ 1.96%
Qwen3-14B	BM25	4.387	4.294	+ 2.17%
Qwen3-32B	BM25	4.342	4.299	+ 1.00%
Qwen3-8B	BGE	4.300	3.922	+ 9.64%
Qwen3-14B	BGE	4.306	4.280	+ 0.61%
Qwen3-32B	BGE	4.324	4.312	+ 0.28%

The results indicate consistent improvements across all configurations, with the most significant gains observed in the BGE + Qwen3-8B combination (+9.64%), suggesting that Social-DeepWriter's iterative retrieval approach particularly benefits semantic retrieval methods in resource-constrained scenarios.

Furthermore, to even validate the results, we also have humans evaluate generated posts, ensuring that the trends observed in automated evaluations align with real-world audience perceptions, particularly regarding clarity, engagement, and strategic alignment.

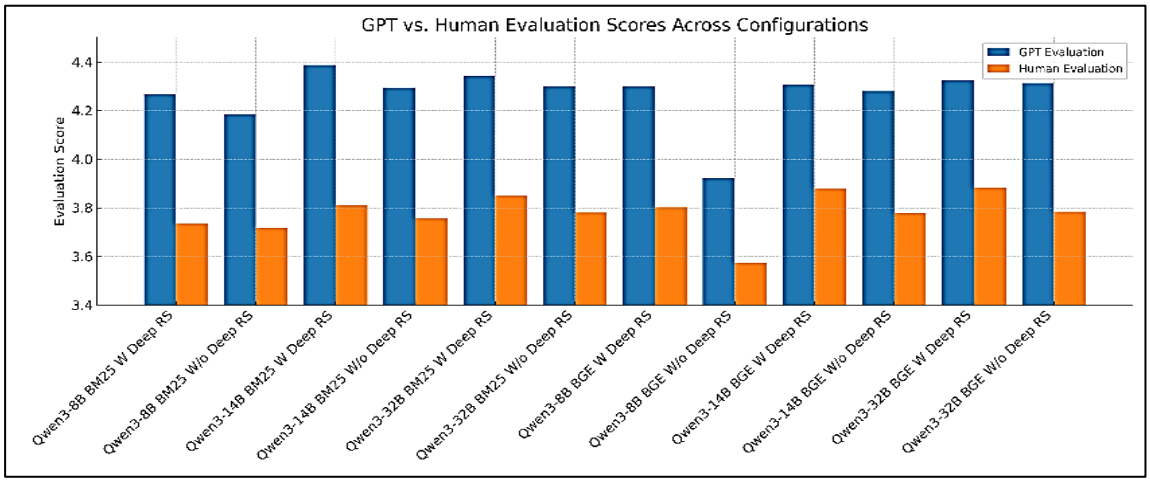

Figure 2. *Social-DeepWriter: GPT-based vs human evaluation scores by model size.*

Figure 2 illustrates the comparative results from GPT-based and human evaluations, revealing clear patterns in how Deep Research contributes to the quality of generated social media content.

The results reveal that although GPT-based and human evaluations exhibit broad agreement in identifying performance trends, they emphasize different dimensions of quality. GPT-based assessments place primary weight on factual correctness, precision of evidence integration, and structural coherence, resulting in greater sensitivity to retrieval quality, particularly evident in the sharper performance distinctions between BM25 and BGE. By contrast, human evaluators tend to value clarity, contextual appropriateness, and overall readability, demonstrating greater tolerance for minor factual omissions when the narrative remains coherent and contextually aligned.

This divergence arises from their inherent evaluative frameworks: GPT-based scoring operates within a fixed, prompt-driven rubric, whereas human judgment is shaped by experiential context, domain familiarity, and audience expectations. These findings support the adoption of a dual-validation strategy: leveraging GPT-based evaluation for scalable, fact-oriented performance monitoring, complemented by human evaluation for targeted, audience-centered quality assurance in mission-critical communication scenarios.

3.3. Detailed analysis

Table 2 and table 3 provide a breakdown of performance across individual metrics for Social-DeepWriter compared to traditional RAG. Social-DeepWriter excelled in **Relevance** (4.89 for Qwen3-14b with BGE), reflecting its ability to align content with mission objectives through iterative retrieval and reflection. **Depth** remained a challenge, with scores slightly lower due to occasional oversimplification during summarization.

Table 2. *Social-DeepWriter performance (BM25 retrieval).*

Model	Overal	Consistency	Coherence	Relevance	Fluency	Depth
Qwen3-8B	4.275	3.915	4.713	4.266	4.926	3.553
Qwen3-14B	4.379	3.947	4.844	4.635	4.958	3.510
Qwen3-32B	4.342	4.160	4.860	4.480	4.950	3.240

Table 3. *Traditional RAG performance (BM25 retrieval).*

Model	Overall	Consistency	Coherence	Relevance	Fluency	Depth
Qwen3-8B	4.184	4.070	4.610	3.980	4.930	3.330
Qwen3-14B	4.294	4.032	4.809	4.191	4.926	3.511
Qwen3-32B	4.299	4.090	4.830	4.450	4.930	3.190

Despite this, the trade-off between conciseness and depth appears to benefit Relevance and Coherence, key priorities in high-tempo strategic communication settings.

Overall, the data confirm that Social-DeepWriter's main strengths lie in Relevance and Coherence, while future refinements could target Depth to further enhance the richness and granularity of generated content.

4. CONCLUSIONS

Social-DeepWriter demonstrates significant advancements in generating mission-aligned social media content, outperforming traditional RAG and standalone LLMs across all evaluated metrics. The framework's iterative query refinement, multi-hop retrieval, and reflective evaluation enable it to produce content that is not only factually accurate and coherent but also strategically tailored to communication objectives. Future work will focus on enhancing the summarization module to preserve nuanced information and integrating real-time feedback mechanisms to further improve adaptability in dynamic information environments. Social-DeepWriter represents a robust, scalable solution for autonomous content generation, with potential to transform strategic communication in both civil and military contexts.

REFERENCES

[1]. Starbird, K. and T. Wilson, *"Disinformation as collaborative work: Surfacing the participatory nature of strategic information operations"*, Proceedings of the ACM on Human-Computer Interaction, 4, CSCW2, 1–26, (2020).

[2]. OpenAI, *"GPT-4 Technical Report"*, Technical Report, (2023).

[3]. Ji, Z., N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, and P. Fung, *"Survey of hallucination in natural language generation"*, ACM Computing Surveys, 55, 12, 1–38, (2023).

[4]. Lewis, P., E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, and S. Riedel, *"Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks"*, Advances in Neural Information Processing Systems, 33, 9459–9474, (2020).

[5]. Wu, J., J. Zhu, and Y. Liu, *"Agentic reasoning: Reasoning LLMs with tools for deep research"*, arXiv preprint arXiv:2502.04644, (2025).

[6]. Ammann, P. J., J. Golde, and A. Akbik, *"Question decomposition for retrieval-augmented generation"*, arXiv preprint arXiv:2507.00355, (2025).

[7]. Ma, X., Y. Gong, P. He, H. Zhao, and N. Duan, *"Query rewriting in retrieval-augmented large*

language models", Proceedings of the Conference on Empirical Methods in Natural Language Processing, 5303–5315, (2023).

[8]. Nguyen, L. T. D., N. D. Nguyen, and K. H. N. Bui, *"An embedding method for sentiment classification across multiple languages"*, Proceedings of the International Conference on Knowledge and Systems Engineering, 1–6, (2021).

[9]. Phan, T. A., N. D. N. Nguyen, and K. H. N. Bui, *"HeterGraphLongSum: Heterogeneous graph neural network with passage aggregation for extractive long document summarization"*, Proceedings of the International Conference on Computational Linguistics, 6248–6258, (2022).

[10]. Liang, X., Y. He, Y. Xia, X. Song, J. Wang, M. Tao, and T. Shi, *"Self-evolving agents with reflective and memory-augmented abilities"*, arXiv preprint arXiv:2409.00872, (2024).

[11]. Yao, S., J. Zhao, D. Yu, N. Du, I. Shafran, K. Narasimhan, and Y. Cao, *"ReAct: Synergizing reasoning and acting in language models"*, International Conference on Learning Representations, (2023).

TÓM TẮT

Mô hình chiến lược sinh nội dung truyền thông phân tích chuyên sâu ứng dụng truy xuất thông tin tăng cường liên tục

Truyền thông mạng xã hội đã trở thành một lĩnh vực quan trọng cho giao tiếp chiến lược, có ảnh hưởng đến nhận thức của công chúng và hỗ trợ các hoạt động cả dân sự lẫn quân sự. Trong các môi trường thông tin có nhịp độ cao, việc tạo nội dung thủ công truyền thống thường quá chậm và tốn nhiều tài nguyên để đáp ứng nhu cầu tương tác theo thời gian thực. Trong khi các mô hình ngôn ngữ lớn (LLM) như GPT-4 có khả năng tạo văn bản giống con người trên quy mô lớn, việc chúng phụ thuộc vào dữ liệu huấn luyện tĩnh đã hạn chế tính phù hợp theo ngữ cảnh, mức độ xác thực thông tin bị nghi ngờ và không có khả năng tùy biến với các nhiệm vụ đa dạng. Để khắc phục những hạn chế này, bài báo này giới thiệu Social-DeepWriter, một khung làm việc hỗ trợ bởi AI để tự động tạo nội dung truyền thông xã hội phù hợp với nhiệm vụ. Được xây dựng dựa trên mô hình "Nghiên cứu Sâu" (Deep Research paradigm), Social-DeepWriter cải thiện Mô hình Tạo sinh tăng cường truy xuất (RAG) truyền thống bằng cách tích hợp tinh chỉnh truy vấn lặp đi lặp lại, truy xuất đa bước và đánh giá nội dung, phản ánh quá trình lập luận nhiều tầng của các nhà phân tích con người. Chúng tôi đánh giá các yếu tố như chất lượng truy xuất, thiết kế câu lệnh và các ràng buộc khi tạo nội dung ảnh hưởng như thế nào đến tính cung cấp thông tin, sự mạch lạc và mức độ phù hợp về mặt chiến lược của các bài đăng được tạo ra. Kết quả của chúng tôi nhấn mạnh tiềm năng của Social-DeepWriter trong việc hỗ trợ các kịch bản giao tiếp lưỡng dụng và các chiến dịch phản ứng nhanh, nơi mà độ chính xác, khả năng thích ứng và khả năng mở rộng là yếu tố thiết yếu.

Từ khoá: Tạo sinh dữ liệu truyền thông mạng xã hội; Mô hình ngôn ngữ lớn; Tạo sinh truy xuất tăng cường; Nghiên cứu sâu; AI cho các vấn đề công cộng.