Evaluating Self-Generated Documents for Enhancing Retrieval-Augmented Generation with Large Language Models

Anonymous ACL submission

Abstract

In retrieval-augmented generation systems, the integration of self-generated documents (SGDs) alongside retrieved content has emerged as a promising strategy for enhancing the performance of large language model. However, previous research primarily focuses on optimizing the use of SGDs, with the inherent properties of SGDs remaining underexplored. Therefore, this paper conducts a comprehensive analysis of different types of SGDs and experiments on various knowledge-intensive tasks. We develop a taxonomy of SGDs grounded in Systemic Functional Linguistics (SFL) to compare the influence of different SGD categories. Our findings offer key insights into what kinds of SGDs most effectively contribute to improving LLM's performance. The results and further fusion methods based on SGD categories also provide practical guidelines for taking better advantage of SGDs to achieve significant advancements in knowledge-driven QA tasks with RAG.

1 Introduction

Recent advancements in large language models (LLMs) have revolutionized natural language processing (NLP) with their impressive ability to understand and generate human-like text (Chang et al., 2023; Ram et al., 2023; Brown et al., 2020; Ouyang et al., 2022). Despite these strides, LLMs remain susceptible to factual inaccuracies and hallucinations, hindering their real-world reliability (Maynez et al., 2020; Zhou et al., 2021; Ji et al., 2023). Retrieval-augmented generation (RAG) has emerged as a promising solution by incorporating external knowledge into the generation process, thereby grounding responses in verifiable information (Lewis et al., 2020; Ram et al., 2023; Kandpal et al., 2023).

While retrieval systems significantly enhance LLM capabilities, recent research points to the potential of self-generated documents (SGDs) as a powerful addition to the RAG framework (Zhang et al., 2023; Liu et al., 2022b; Sun et al., 2023; Yu et al., 2023). SGDs, autonomously generated by the model, offer a dynamic way to supplement retrieved content and enrich the information available for downstream tasks.

Despite the growing interest in SGDs, previous related work has mostly focused on how to better utilize the information in SGDs (Liu et al., 2022a; Fang et al., 2022; Yu et al., 2023), such as designing specific methods to handle conflicting knowledge (Zhang et al., 2023), while neglecting the impact of the inherent features of SGDs. Intuitively, if we have a more thorough understanding of these features and know which ones are more likely to improve the performance of RAG, the potential of SGDs will be fundamentally increased. This paper aims to present a comprehensive exploration of SGDs within RAG by exploring three core research questions:

- RQ1: To what extent do self-generated documents enhance the performance of RAG? (Section 5)
- RQ2: Which types of self-generated documents are most effective across different scenarios? (Section 6)
- RQ3: How can self-generated documents be effectively integrated with external sources (e.g., Wikipedia) based on their properties to further enhance the performance? (Section 7)

To address these questions, we explore RQ1, RQ2, and RQ3 in Sections 5, 6, and 7, respectively. We base our approach on **Systemic Functional Linguistics (SFL)** theory, which provides a systematic framework for analyzing language through its three core metafunctions: ideational, interpersonal, and textual. These metafunctions elucidate how the text constructs meaning, manages social relationships,

and organizes discourse. Leveraging SFL, we develop a taxonomy that categorizes self-generated documents by their granularity, tone, and structure, aligning each document type with distinct communicative functions as articulated by SFL theory.

In summary, our contributions are as follows:

- The design of a novel taxonomy for SGDs based on SFL theory, categorizing documents by their functional roles, linguistic attributes, and impact on RAG performance.
- A series of experiments evaluating the effectiveness of different SGD types across a variety of knowledge-intensive tasks, including open-domain question answering, multi-hop question answering, long-form question answering, and fact verification.
- Practical insights and guidelines for optimizing the integration of SGDs within RAG frameworks, demonstrating how tailored document types can maximize task performance and efficiency.

Our findings reveal which types of self-generated documents are most effective in different contexts, providing actionable recommendations for designing RAG systems that fully leverage the potential of SGDs. By examining the interplay between retrieval mechanisms and document generation, we contribute to the development of more robust and adaptable RAG systems capable of achieving state-of-the-art results across a broad spectrum of knowledge-intensive NLP tasks.[1]

2 Systemic Functional Linguistics Theory

To build an effective classification system for self-generated documents (SGDs) that enhances Retrieval-Augmented Generation (RAG) systems, we ground our taxonomy in Systemic Functional Linguistics (SFL). SFL, developed by Michael Halliday (Halliday and Hasan, 1989), views language as a dynamic tool shaped by social contexts, emphasizing its role in meaning-making rather than static grammatical rules. This framework helps us categorize SGDs based on their communicative functions, aligning document types with specific retrieval and generation tasks to improve RAG performance.

At the heart of SFL are three metafunctions —ideational, interpersonal, and textual— which operate together to convey meaning (Gebhard and Accurso, 2020; Bilal, 2012). These metafunctions guide how language expresses ideas, manages social relations, and structures discourse.

Ideational Metafunction focuses on how language represents knowledge and organizes experiences (Rajab Dorri, 2020; Hutabarat et al., 2020; Koussouhon and Dossoumou, 2015). This includes everything from detailed factual information to abstract concepts, depending on the communicative goal.

Interpersonal Metafunction handles the relational dynamics of communication, including tone, authority, and interaction with the audience (Shekhani, 2022; Cheng, 2023; Chueasuai, 2017). It defines whether the document is formal or conversational and shapes how trust or authority is established.

Textual Metafunction ensures the logical flow and coherence of the document, structuring information for clarity and accessibility (Sameer et al., 2020; Matthiessen, 1992; Koutchadé, 2017). This function is crucial for guiding readers through complex or information-dense content.

By applying SFL to SGDs, we can classify documents not only based on the information they provide but also on how they engage readers and structure their content. This approach helps develop a nuanced taxonomy that aligns with the varying communicative demands of different tasks in RAG systems.

3 Taxonomy of Self-Generated Documents

We classify self-generated documents using Systemic Functional Linguistics (SFL), as introduced in Section 2. Our classification framework is built upon the three core metafunctions of SFL: Ideational, Interpersonal, and Textual.

The Ideational Metafunction pertains to how language represents experiences and conveys knowledge. We classify documents by their granularity, reflecting the level of detail provided. Fine-grained documents offer intricate, detailed information—such as that found in technical documentation and research papers. Coarse-grained documents, on the other hand, provide broader overviews, akin to summaries and opinion pieces. This distinction determines how deeply a document explores its subject, aligning with its purpose as a vehicle for knowledge transmission.

[1] Our code and data will be released to the community to facilitate future research.

Type	Interpersonal	Ideational	Textual	Example
AFS	Authorative	Fine-Grain	Structured	Research Paper
AFU	Authorative	Fine-Grain	Unstructured	Case Studies
ACS	Authorative	Coarse-Grain	Structured	News Article
ACU	Authorative	Coarse-Grain	Unstructured	Opinion Pieces
CFS	Conversational	Fine-Grain	Structured	How-to Guides
CFU	Conversational	Fine-Grain	Unstructured	Personal Essays
CCS	Conversational	Coarse-Grain	Structured	Travel Blogs
CCU	Conversational	Coarse-Grain	Unstructured	Social Media Posts

Table 1: Classification of document types based on the three SFL theory metafunctions: interpersonal, ideational, and textual, including representative examples for each category.

The Interpersonal Metafunction focuses on how language manages relationships between the writer and reader, with tone being the key feature. Authoritative documents use formal language to establish credibility and expertise, common in academic papers and reports. In contrast, conversational documents adopt an informal tone, engaging the reader in a more personal manner, as seen in blogs or how-to guides. This classification highlights the social interaction within documents, emphasizing how tone affects audience engagement.

The Textual Metafunction governs how information is organized. We classify documents by their structure. Structured documents follow a clear, logical format with headings and subheadings, ensuring clarity and navigation, typical in instructional materials and reports. Unstructured documents use a more free-flowing narrative, common in creative writing or informal blog posts, where the emphasis is on storytelling or reflection. This dimension reflects how documents present and guide readers through content.

Overall, considering the three metafunctions, we define eight distinct document types, as illustrated in Table 1. For instance, a fine-grained, structured, authoritative document (e.g., a research paper) delivers detailed, organized, and credible information. Conversely, a coarse-grained, unstructured, conversational document (e.g., a social media post) provides an engaging, high-level overview.

4 Experiment Settings

We evaluate different experiment settings across four diverse, knowledge-intensive tasks in a zero-shot setting. Following prior work (Trivedi et al., 2022), we limit our experiments to the first 500 samples from each dataset to reduce computational overhead. As done in previous research (Yu et al., 2024), we adopt a zero-shot evaluation approach. The tasks span open-domain question answering, multi-hop reasoning, fact verification, and long-form question answering, using well-established datasets: TriviaQA (TQA), HotpotQA, FEVER, and ELI5. Table 2 summarizes the datasets, task types and evaluation metrics.

Dataset	Task Type	Evaluation Metric
TQA	Open-Domain Question Answering	EM
HotpotQA	Multi-Hop Question Answering	F_1
FEVER	Fact Verification	EM
ELI5	Long-Form Question Answering	F_1

Table 2: Summary of datasets, task types, and evaluation metrics for TQA, HotpotQA, FEVER, and ELI5.

4.1 Tasks and Datasets

Open-Domain Question Answering We evaluate open-domain QA on the TriviaQA (TQA) (Joshi et al., 2017) dataset. TQA contains questions q alongside answers o, extracted from supporting Wikipedia documents P. We evaluate the dataset using the Exact Match (EM) metric, following the framework used by Lewis et al. (2020).

Multi-Hop Question Answering For multi-hop reasoning, we use the HotpotQA dataset (Yang et al., 2018), which requires reasoning across multiple passages P to arrive at a correct answer. HotpotQA includes over 113K question-answer pairs, many of which necessitate reasoning beyond a single source document. This task is evaluated using the F_1 score, as proposed by Yang et al. (2018), due to the abstractive nature of many responses.

Fact Verification We assess fact verification capabilities using the FEVER dataset (Thorne et al., 2018), part of the KILT benchmark (Petroni et al., 2021). Each sample consists of a claim q that is either supported or refuted by Wikipedia passages. Claims are labeled as "SUPPORTS" or "REFUTES" based on whether they align with or contradict the factual evidence. We report performance using the Exact Match (EM) metric, consistent with prior work (Thorne et al., 2018).

Long-Form Question Answering For long-form QA, we use the ELI5 dataset (Fan et al., 2019), which contains 270K threads from the Reddit forum "Explain Like I'm Five." These questions require elaborate, multi-sentence answers, often demanding a higher level of abstraction and reasoning. Given the complexity of the responses, we evalu-

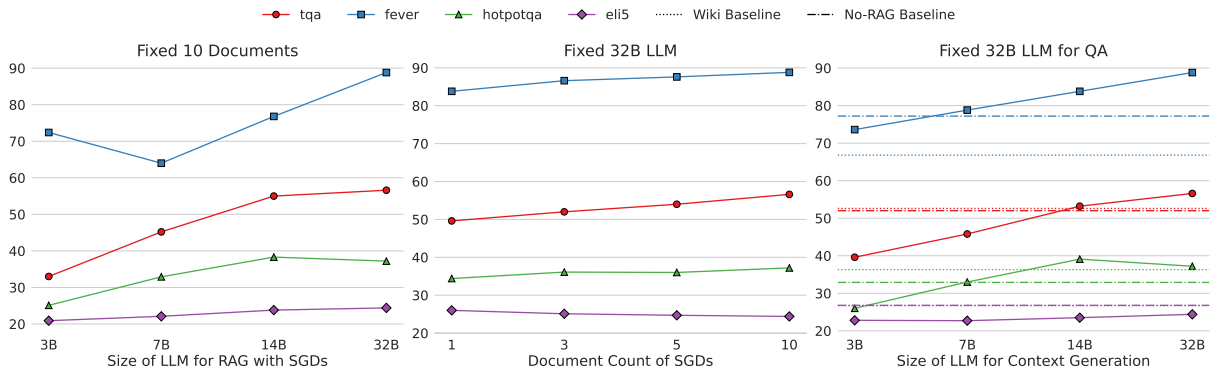

Figure 1: Performance factor analysis examining the effects of base model size, document count, and context generation model size on RAG performance across different tasks. **Left:** Impact of varying the base model size when using the same Qwen2.5 model for both context generation and question answering, with the number of self-generated documents fixed at 10. **Middle:** Impact of varying the number of self-generated documents, using the Qwen2.5-32B-Instruct model as the base model. **Right:** Impact of varying the context generation model size while keeping the question answering model fixed at Qwen2.5-32B-Instruct and the document count at (n=10).

ate performance using the F_1 metric, focusing on the generative short-answer setting, where models must provide concise yet informative responses.

4.2 Large Language Models

To investigate the impact of model size on the effectiveness of self-generated documents, as explored in Section 5, we leverage the Qwen2.5 series of large language models (Team, 2024). This model family offers a wide range of readily available and computationally feasible model sizes, spanning from 0.5B to 72B parameters. This diverse selection allows for a more comprehensive analysis of the relationship between model scale and performance, compared to other mainstream models like LLaMA, which provides a more limited set of parameter variants (8B, 70B, and 405B in LLaMA 3.1; 1B and 3B in LLaMA 3.2) (Dubey et al., 2024). Moreover, the Qwen2.5 series exhibits strong performance, significantly outperforming the LLaMA 3.1 series across comparable parameter sizes, as reported in (Team, 2024).

To ensure controlled experimental conditions and isolate the impact of model size (Section 5), we exclusively employ models from the Qwen2.5 series. Our investigation focuses on the 3B, 7B, 14B, and 32B variants, balancing computational constraints with model performance. We primarily utilize the largest computationally feasible model within this series, Qwen2.5-32B-Instruct, as the backbone model for both question answering and context generation. To analyze the specific influence of context generation model size, we fix the QA model as Qwen2.5-32B-Instruct and vary the size of the context generation model, leveraging the selected range of Qwen2.5 Instruct models. These experiments are illustrated in Figure 1.

4.3 Implementation Details

Our implementation follows a two-step workflow: generating diverse Wiki context shots and transforming them into the target type to produce SGDs of various types. The purpose of these context shots is to ensure the generated documents align closely with the desired type, enabling the model to provide more tailored, high-quality contexts for RAG systems.

For shot generation, we follow the methodology outlined in *GenRead* (Yu et al., 2023), utilizing the `bge-large-en-v1.5` model (Xiao et al., 2023) to embed both questions and documents. This process yields a set of diverse "shots," each consisting of a query paired with its corresponding top-ranked Wikipedia context. However, as the retrieved contexts are Wikipedia articles, their inherent factual, encyclopedic style may not align with our desired SGD types.

To generate SGD with diverse types, we first employ GPT-4o to transform these Wikipedia contexts into the target types, and then use them to guide our backbone model to generate SGDs. We prompt GPT-4o with a shot (query-context pair) and a task-specific prompt designed to elicit a specific document style (e.g., conversational, fine-grained), prompting it to generate a style-transformed document under a zero-shot setting. These transformed shots are then used to prompt our backbone model, Qwen2.5-32B-Instruct, for SGD generation. A

manual evaluation of 120 sampled SGDs revealed that an average of 94.2% of generated documents closely matched the target types in terms of tone, granularity, and structure. Details of the manual analysis are given in Appendix D. This strong alignment confirms that the transformations effectively achieved the intended variations. Detailed prompts for GPT-4o style transformation and Qwen2.5-32B-Instruct context generation are provided in Appendices E.1 and E.2, respectively. Further implementation details are provided in Appendix C.

5 Impact of Self-Generated Documents

To determine the extent to which SGDs enhance the performance of RAG, we investigate the impact of three key factors. In this section, the *"base model"* refers to the language model used for both question answering and generating the SGDs, while the *"context generation model"* refers to the language model used only to generate the documents for QA. We first examine how the size of the base model affects performance. Next, we analyze the impact of varying the number of self-generated documents used. Finally, we isolate the effect of the context generation model size by fixing the question answering model size and only changing the size of the context generation model. Our analysis leverages the Qwen2.5 series Instruct models (3B, 7B, 14B, and 32B) and explores various settings across four knowledge-intensive tasks, as presented in Table 2. We report the corresponding performance metrics for each dataset. Detailed experimental results are provided in Appendix B.

5.1 Scaling the Base Model

Overall Trends Our results indicate that the size of the base model plays a crucial role in enhancing performance across all tasks. As the model size used for both self-generation and question answering increases, we observe a general trend of performance improvement across all datasets (Figure 1, left). This demonstrates the substantial benefits of scaling model parameters when incorporating SGDs within a RAG framework.

Detailed Analysis While larger base models generally lead to better performance, the magnitude of these gains is not uniform across all tasks. Notably, the relative improvement from increasing model size diminishes as we scale up. This is particularly evident in tasks that demand complex reasoning and factual synthesis, such as HotpotQA. This suggests that for such tasks, simply increasing the base model size may not yield proportional benefits, potentially indicating a plateau in the model's ability to effectively process and utilize the expanded knowledge capacity.

5.2 Varying Document Number

Overall Trends Increasing the number of self-generated documents (n) generally improves performance across most tasks (Figure 1, middle). This trend suggests that a greater number of task-specific documents enhances the availability of relevant information, positively impacting overall model performance.

Detailed Analysis However, the impact of document number (n) is not consistent across all tasks. For those heavily reliant on factual verification, such as FEVER, increasing the number of SGDs leads to substantial performance gains. Conversely, in long-form question answering tasks like ELI5, which rely more on creative reasoning, providing a larger number of documents can sometimes result in a slight performance decrease. This discrepancy suggests that for tasks demanding creative reasoning, carefully selecting a smaller number of highly relevant and high-quality documents is more crucial than simply increasing the quantity of input.

5.3 Scaling the Context Generation Model

Overall Trends Our results demonstrate the significant influence of the context generation model size on overall RAG performance (Figure 1, right). As the model size specifically used for context generation increases, performance generally trends upward across all tasks, indicating the effectiveness of larger models for producing high-quality, task-relevant context.

Detailed Analysis Larger context generation models consistently outperform smaller models, often exceeding both the No-RAG setting and the performance achieved by leveraging Wikipedia as an external knowledge source. This trend becomes increasingly pronounced as the model size scales up, highlighting the crucial role of model capacity in generating high-quality, task-relevant context. This finding aligns with previous studies, which often utilize large LLMs for context generation, demonstrating their effectiveness in this role (Yu et al., 2023). Our analysis further reveals that this conclusion holds true specifically when the context generation LLM is sufficiently large. While

smaller models may lack the capacity for generating highly effective documents, larger models can produce superior context that significantly benefits RAG performance.

In conclusion, our analyses demonstrate that self-generated documents can substantially enhance the performance of RAG systems. Within this framework, utilizing larger LLMs for both question answering and self-generation generally leads to significant performance improvements. Furthermore, leveraging larger models specifically for context generation is crucial for producing high-quality, task-relevant documents that outperform both pre-existing knowledge bases like Wikipedia and solely relying on the model's internal knowledge. Optimizing the number of generated documents based on specific task requirements, particularly for creative reasoning tasks, also plays a significant role in maximizing performance.

Having established the positive impact of self-generated documents on RAG performance, we next investigate which specific types of SGDs are most effective. This will allow us to tailor document generation to the specific needs of different knowledge-intensive tasks.

6 Exploration of SGD Types

To investigate which types of self-generated documents are most effective across different scenarios, we analyze their performance based on the taxomony in Section 3. The performance of different self-generated document types across tasks is presented in Table 3 and aggregated by their metafunctional dimensions in Table 4.

6.1 Metafunction Analysis

Textual Metafunction Unstructured SGDs consistently outperform their structured counterparts across all tasks, achieving the highest average score (Table 4). It suggests that the flexibility of unstructured content allows the model to better extract and utilize relevant information, potentially enabling more nuanced and adaptable understanding.

Interpersonal Metafunction While both authoritative and conversational tones achieve similar average performance, we observe task-dependent differences. Conversational documents are slightly more effective for long-form QA tasks like ELI5, potentially fostering a more engaging and accessible interaction for knowledge exploration. On the other hand, authoritative documents perform slightly better for factual tasks like FEVER, likely due to the emphasis on precision and reliability associated with a formal tone in factual verification scenarios.

Ideational Metafunction Fine-grain and coarse-grain documents demonstrate comparable average performance overall. However, we see task-specific variations. Fine-grained documents excel in tasks requiring detailed factual information (e.g., TQA), likely benefiting from the richness of detail and specific references provided. In contrast, the latter perform better in open-ended tasks where broader overviews are more beneficial (e.g., ELI5), potentially due to their ability to provide a more concise and digestible overview of the topic.

6.2 Task-Specific Analysis

Fact Verification In Fact Verification (FEVER), unstructured, authoritative, and fine-grained documents perform best. This suggests that for tasks requiring high accuracy and reliable information, a combination of content flexibility, a formal tone, and detailed factual content proves most effective.

Multi-hop QA For Multi-hop QA (HotpotQA), unstructured documents outperform structured ones, and fine-grained content slightly outperforms coarse-grained content. This indicates that for tasks involving complex reasoning over multiple information sources, a flexible content structure and a higher level of detail are beneficial.

Open-Domain QA In Open-Domain QA (TQA), fine-grained and authoritative documents demonstrate a slight advantage. This finding aligns with the need for precision and detailed information in answering factual questions based on a broad knowledge base.

Long-Form QA In Long-Form QA (ELI5), conversational and unstructured documents perform best. This suggests that for tasks focused on generating accessible and user-friendly explanations, a more engaging and flexible style is crucial.

6.3 Key Insights

Our analysis reveals several key insights regarding the effectiveness of different SGD types. First, unstructured documents generally demonstrate superior performance across a range of tasks. The flexibility offered by unstructured content appears to allow the model to more effectively extract and synthesize relevant information, ultimately leading to better overall results.

Docs Type	Interpersonal	Ideational	Textual	TQA	HOTPOTQA	FEVER	ELI5	Average
AFS	Authoritative	Fine-Grain	Structured	55.4	37.5	83.4	20.9	49.3
AFU	Authoritative	Fine-Grain	Unstructured	**58.4**	**39.9**	**85.2**	22.8	**51.6**
ACS	Authoritative	Coarse-Grain	Structured	56.8	38.7	83.6	22.0	50.3
ACU	Authoritative	Coarse-Grain	Unstructured	56.4	39.6	85.0	23.5	51.1
CFS	Conversational	Fine-Grain	Structured	57.4	36.4	81.8	24.0	49.9
CFU	Conversational	Fine-Grain	Unstructured	57.8	39.5	84.2	25.1	**51.6**
CCS	Conversational	Coarse-Grain	Structured	56.0	35.2	82.0	25.4	49.6
CCU	Conversational	Coarse-Grain	Unstructured	56.6	38.7	83.8	26.4	51.4
WIKI	Authoritative	Flexible	Unstructured	52.6	36.3	66.8	**26.7**	45.6

Table 3: Performance comparison of various document types across tasks. Each document type is categorized based on its interpersonal, ideational, and textual functions. The best overall performance for each dataset is highlighted in **bold**, and the best performance within the category of SGDs is underlined.

Meta Function	Dimension	TQA	HOTPOTQA	FEVER	ELI5	Average
IDEATIONAL	FINE-GRAIN	57.2	38.3	83.6	23.2	50.6
	COARSE-GRAIN	56.4	38.1	83.6	24.3	50.6
INTERPERSONAL	AUTHORITATIVE	56.8	38.9	84.3	22.3	50.6
	CONVERSATIONAL	57.0	37.4	83.0	25.2	50.7
TEXTUAL	STRUCTURED	56.4	37.0	82.7	23.1	49.8
	UNSTRUCTURED	57.3	39.4	84.6	24.5	51.4

Table 4: Aggregated performance across different metafunction dimensions (Ideational, Interpersonal, and Textual) for various datasets. Each metafunction dimension is further divided into fine-grain and coarse-grain (for Ideational), authoritative and conversational (for Interpersonal), and structured and unstructured (for Textual). The best performance within each metafunction dimension is underlined.

Second, the optimal tone and level of detail for SGDs are highly task-dependent. Long-Form QA tasks, such as ELI5, benefit from a conversational tone and coarse-grained content, creating a more engaging and accessible experience for users seeking explanations. Conversely, factual tasks, like FEVER, are better served by authoritative, fine-grained documents, reflecting the need for precision and reliability in such scenarios.

Finally, our findings strongly emphasize the importance of tailoring SGDs to specific task requirements. Aligning document characteristics with specific task demands is crucial for maximizing the benefits of SGDs in RAG systems.

Building on these findings, we next delve into integrating SGDs with readily available external knowledge sources, specifically Wikipedia. This exploration aims to further enhance RAG performance by combining the strengths of generated and retrieved knowledge.

7 Integrating SGDs with External Knowledge

This section explores how SGDs can be effectively integrated with external knowledge sources, focusing on their combination with Wikipedia within the RAG framework. We evaluate two strategies for merging SGDs with Wiki documents: *Direct Mix* and *Style-Transformation Mix*. Table 5 shows the performance of these strategies for the document types that achieved top results in at least one task from Table 3, across different tasks.

Direct Mix The Direct Mix strategy directly concatenates the top k ranked Wiki documents, denoted as $W = \{W_1, W_2, \ldots, W_k\}$, with the first n SGDs from a predefined taxonomy, represented as $S = \{S_1, S_2, \ldots, S_n\}$. The final mixed document set D_{Mixed} is given by:

$$D_{\text{Mixed}} = \text{Concat}\left(\bigcup_{i=1}^{k} W_i, \bigcup_{j=1}^{n} S_j\right)$$

As shown in Table 5, the Direct Mix strategy successfully combines broad factual knowledge from Wiki documents with task-specific insights from SGDs. However, performance can be affected by stylistic inconsistencies between the Wiki content and the SGDs, which may lead to variability in tone and structure.

Style-Transformation Mix The Style-Transformation Mix strategy addresses the stylistic mismatch by transforming the retrieved Wiki documents to align with the style of the

Docs Type	Mix Strategy	Interpersonal	Ideational	Textual	TQA	HOTPOTQA	FEVER	ELI5	Average
WIKI	-	Authoritative	Flexible	Unstructured	52.6	36.3	66.8	**26.7**	45.6
GENREAD MIX	Direct Mix	Authoritative	Flexible	Unstructured	57.6	39.1	87.0	24.7	52.1
AFU	Direct Mix	Authoritative	Fine-Grain	Unstructured	$56.8_{\downarrow 1.6}$	$41.0_{\uparrow 1.1}$	$84.4_{\downarrow 0.8}$	$23.1_{\uparrow 0.3}$	$51.3_{\downarrow 0.3}$
CFU	Direct Mix	Conversational	Fine-Grain	Unstructured	$57.2_{\downarrow 0.6}$	$41.6_{\uparrow 2.1}$	$85.0_{\uparrow 0.8}$	$25.6_{\uparrow 0.5}$	$52.4_{\uparrow 0.8}$
CCU	Direct Mix	Conversational	Coarse-Grain	Unstructured	$\underline{57.4}_{\uparrow 0.8}$	$\underline{41.8}_{\uparrow 3.1}$	$\underline{85.8}_{\uparrow 2.0}$	$\underline{26.5}_{\uparrow 0.1}$	$\underline{52.9}_{\uparrow 1.5}$
AFU	Style Transformation Mix	Authoritative	Fine-Grain	Unstructured	$\underline{\mathbf{59.4}}_{\uparrow 1.0}$	$\underline{\mathbf{43.3}}_{\uparrow 3.4}$	$\underline{\mathbf{87.6}}_{\uparrow 2.4}$	$22.7_{\downarrow 0.1}$	$53.3_{\uparrow 1.7}$
CFU	Style Transformation Mix	Conversational	Fine-Grain	Unstructured	$58.0_{\uparrow 0.2}$	$41.0_{\uparrow 1.5}$	$86.4_{\uparrow 2.2}$	$25.3_{\uparrow 0.2}$	$52.7_{\uparrow 1.1}$
CCU	Style Transformation Mix	Conversational	Coarse-Grain	Unstructured	$57.6_{\uparrow 1.0}$	$43.0_{\uparrow 4.3}$	$87.4_{\uparrow 3.6}$	$\underline{26.3}_{\downarrow 0.1}$	$\underline{\mathbf{53.6}}_{\uparrow 2.2}$

Table 5: Performance comparison across different generated document types combined with Wiki documents using mix with style transformation strategy for various tasks. Superscript values indicate the difference from Table 3, with increases marked by an upward arrow (↑) and decreases by a downward arrow (↓). The best overall performance for each dataset is highlighted in **bold**, and the best performance within the category of SGDs with the same mix strategy is underlined.

SGDs. The transformation function $T(W_i)$ uses a large language model to adapt the style of each Wiki document W_i to match the SGDs. The final mixed document set is constructed as:

$$D_{\text{Mixed}} = \text{Concat}\left(\bigcup_{i=1}^{k} T(W_i), \bigcup_{j=1}^{n} S_j\right)$$

By aligning the style of the Wiki documents with the SGDs, the Style-Transformation Mix strategy creates a more unified context, enhancing coherence across documents. This stylistic alignment is particularly useful for tasks that require both factual accuracy and nuanced reasoning, leading to improved performance across datasets, as shown in Table 5.

Performance Comparison The results in Table 5 demonstrate that integrating Wiki documents with SGDs improves performance across different tasks and document types. The Style-Transformation Mix strategy shows clear advantages over Direct Mix, emphasizing the importance of stylistic alignment between external knowledge and SGDs for better RAG system performance.

8 Related Work

Generation-Augmented Approaches Recent advances in **generation-augmented methods** leverage LLMs to generate intermediate contexts that enhance task performance by drawing on their pre-trained parametric knowledge (Petroni et al., 2019; Roberts et al., 2020; Chen et al., 2023). The generated contexts span diverse knowledge domains, including background information (Sun et al., 2023; Yu et al., 2023), commonsense reasoning (Liu et al., 2022a), and domain-specific expertise (Feng et al., 2024; Luo et al., 2023), along with chain-of-thought reasoning processes (Wei et al., 2022; Kojima et al., 2022).

Linguistic Frameworks in NLP Linguistic frameworks have been employed to improve the coherence of language model generated content (Patten et al., 1992; Majewska and Korhonen, 2023; Miaschi et al., 2020). However, their application to analyzing self-generated documents in RAG remains underexplored. Our work addresses this by introducing a taxonomy of SGDs grounded in Systemic Functional Linguistics (SFL) (Halliday and Hasan, 1989), optimizing document generation for factual accuracy and engagement in RAG systems.

Hybrid Strategies for Knowledge Integration Hybrid approaches, which combine retrieved and generated content, have shown promise in knowledge intensive tasks (Abdallah and Jatowt, 2024; Zhang et al., 2023; Yu et al., 2023; Shi et al., 2024). However, these methods often overlook how different types of generated documents affect performance. We propose a novel framework that classifies SGDs based on their communicative roles using SFL (Halliday and Hasan, 1989), allowing for more effective integration of generated and retrieved content, resulting in notable performance gains in knowledge-driven NLP tasks.

9 Conclusion

This study demonstrates that the type of self-generated document (SGD) notably influences the effectiveness of Retrieval-Augmented Generation (RAG) systems. We show that tailoring SGD types to specific tasks leads to performance improvements across diverse knowledge-intensive tasks (RQ1 & RQ2). Additionally, our exploration of strategies for integrating SGDs with external knowledge, such as Wikipedia, reveals further performance gains, highlighting the potential of combining generated and retrieved information (RQ3).

Limitations

This study primarily focuses on the Qwen2.5-32B-Instruct model due to computational constraints. Exploring the effectiveness of SGDs with larger models like GPT-4, especially in specialized domains requiring extensive knowledge, remains a promising direction for future work.

Ethical Considerations

In this research, we evaluate the role of self-generated documents (SGD) by models for enhancing the RAG system. One concern that arises is the inherent uncertainty regarding the correctness of the self-generated content. This issue is also prevalent in various methods such as self-generated data augmentation, self-reflection, and so on, as models may produce inaccurate or misleading information. While we have ensured the quality of the style of SGDs through human evaluation, their factual accuracy cannot be fully guaranteed.

References

Abdelrahman Abdallah and Adam Jatowt. 2024. Generator-retriever-generator approach for open-domain question answering. *Preprint*, arXiv:2307.11278.

Hafiz Bilal. 2012. Analysis of thank you m'am: Halliday's metafunctions.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. *Preprint*, arXiv:2005.14165.

Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, Wei Ye, Yue Zhang, Yi Chang, Philip S. Yu, Qiang Yang, and Xing Xie. 2023. A survey on evaluation of large language models. *Preprint*, arXiv:2307.03109.

Liang Chen, Yang Deng, Yatao Bian, Zeyu Qin, Bingzhe Wu, Tat-Seng Chua, and Kam-Fai Wong. 2023. Beyond factuality: A comprehensive evaluation of large language models as knowledge generators. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 6325–6341, Singapore. Association for Computational Linguistics.

Shi Cheng. 2023. A review of interpersonal metafunction studies in systemic functional linguistics (2012–2022). *Journal of World Languages*.

Pasakara Chueasuai. 2017. The interpersonal metafunction and translation of power relations: A case study of fifty shades of grey. *MANUSYA*, 23 (2017):1–22.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, Arun Rao, Aston Zhang, Aurelien Rodriguez, Austen Gregerson, Ava Spataru, Baptiste Roziere, Bethany Biron, Binh Tang, Bobbie Chern, Charlotte Caucheteux, Chaya Nayak, Chloe Bi, Chris Marra, Chris McConnell, Christian Keller, Christophe Touret, Chunyang Wu, Corinne Wong, Cristian Canton Ferrer, et al. 2024. The llama 3 herd of models. *Preprint*, arXiv:2407.21783.

Angela Fan, Yacine Jernite, Ethan Perez, David Grangier, Jason Weston, and Michael Auli. 2019. Eli5: Long form question answering. *Preprint*, arXiv:1907.09190.

Yuwei Fang, Shuohang Wang, Yichong Xu, Ruochen Xu, Siqi Sun, Chenguang Zhu, and Michael Zeng. 2022. Leveraging knowledge in multilingual commonsense reasoning. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 3237–3246.

Shangbin Feng, Weijia Shi, Yuyang Bai, Vidhisha Balachandran, Tianxing He, and Yulia Tsvetkov. 2024. Knowledge card: Filling llms' knowledge gaps with plug-in specialized language models. *Preprint*, arXiv:2305.09955.

Meg Gebhard and Kathryn Accurso. 2020. *Systemic Functional Linguistics*.

Michael A.K. Halliday and Rugaiya Hasan. 1989. Language, context, and text: Aspects of language in a social-semiotic perspective.

Elisabet Hutabarat, Herman Herman, Dumaris Silalahi, and Partohap Sihombing. 2020. An analysis of ideational metafunction on news jakarta post about some good covid-19 related news. *VELES Voices of English Language Education Society*, 4:142–151.

Gautier Izacard, Mathilde Caron, Lucas Hosseini, Sebastian Riedel, Piotr Bojanowski, Armand Joulin, and Edouard Grave. 2021. Unsupervised dense information retrieval with contrastive learning. *Trans. Mach. Learn. Res.*, 2022.

Gautier Izacard, Patrick Lewis, Maria Lomeli, Lucas Hosseini, Fabio Petroni, Timo Schick, Jane Dwivedi-Yu, Armand Joulin, Sebastian Riedel, and Edouard Grave. 2022. Few-shot learning with retrieval augmented language models. *arXiv preprint arXiv:2208.03299*.

Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12):1–38.

Mandar Joshi, Eunsol Choi, Daniel Weld, and Luke Zettlemoyer. 2017. TriviaQA: A large scale distantly supervised challenge dataset for reading comprehension. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1601–1611, Vancouver, Canada. Association for Computational Linguistics.

Nikhil Kandpal, Haikang Deng, Adam Roberts, Eric Wallace, and Colin Raffel. 2023. Large language models struggle to learn long-tail knowledge. *Preprint*, arXiv:2211.08411.

Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2022. Large language models are zero-shot reasoners. *arXiv preprint arXiv:2205.11916*.

Léonard Koussouhon and Ashani Dossoumou. 2015. Exploring ideational metafunction in helon habila's oil on water: A re-evaluation and redefinition of african women's personality and identity through literature. *International Journal of Applied Linguistics and English Literature*, 4:129–136.

Innocent Sourou Koutchadé. 2017. Exploring textual metafunction in selected nigerian news reports: Linguistic description and interpretation. *Advances in Social Sciences Research Journal*, 4(16).

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. In *Advances in Neural Information Processing Systems*, volume 33, pages 9459–9474. Curran Associates, Inc.

Jiacheng Liu, Alisa Liu, Ximing Lu, Sean Welleck, Peter West, Ronan Le Bras, Yejin Choi, and Hannaneh Hajishirzi. 2022a. Generated knowledge prompting for commonsense reasoning. *Preprint*, arXiv:2110.08387.

Jiacheng Liu, Alisa Liu, Ximing Lu, Sean Welleck, Peter West, Ronan Le Bras, Yejin Choi, and Hannaneh Hajishirzi. 2022b. Generated knowledge prompting for commonsense reasoning. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3154–3169, Dublin, Ireland. Association for Computational Linguistics.

Ziyang Luo, Can Xu, Pu Zhao, Xiubo Geng, Chongyang Tao, Jing Ma, Qingwei Lin, and Daxin Jiang. 2023. Augmented large language models with parametric knowledge guiding. *Preprint*, arXiv:2305.04757.

Olga Majewska and Anna Korhonen. 2023. Verb classification across languages. *Annual Review of Linguistics*, 9(Volume 9, 2023):313–333.

Christian Matthiessen. 1992. Interpreting the textual metafunction. *Advances in Systemic Linguistics: Recent Theory and Practice*, pages 37–81.

Joshua Maynez, Shashi Narayan, Bernd Bohnet, and Ryan McDonald. 2020. On faithfulness and factuality in abstractive summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1906–1919, Online. Association for Computational Linguistics.

Alessio Miaschi, Dominique Brunato, Felice Dell'Orletta, and Giulia Venturi. 2020. Linguistic profiling of a neural language model. In *Proceedings of the 28th International Conference on Computational Linguistics*. International Committee on Computational Linguistics.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. *Preprint*, arXiv:2203.02155.

Terry Patten, Michael L. Geis, and Barbara D. Becker. 1992. Toward a theory of compilation for natural language generation. *Computational Intelligence*, 8(1):77–101.

Fabio Petroni, Aleksandra Piktus, Angela Fan, Patrick Lewis, Majid Yazdani, Nicola De Cao, James Thorne, Yacine Jernite, Vladimir Karpukhin, Jean Maillard, Vassilis Plachouras, Tim Rocktäschel, and Sebastian Riedel. 2021. Kilt: a benchmark for knowledge intensive language tasks. *Preprint*, arXiv:2009.02252.

Fabio Petroni, Tim Rocktäschel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, Alexander H. Miller, and Sebastian Riedel. 2019. Language models as knowledge bases? *Preprint*, arXiv:1909.01066.

Elham Rajab Dorri. 2020. Textual metafunction and translation: A systemic functional linguistic perspective. *Applied Linguistics Research Journal*, 4.

Ori Ram, Yoav Levine, Itay Dalmedigos, Dor Muhlgay, Amnon Shashua, Kevin Leyton-Brown, and Yoav Shoham. 2023. In-context retrieval-augmented language models. *Transactions of the Association for Computational Linguistics*, 11:1316–1331.

Adam Roberts, Colin Raffel, and Noam Shazeer. 2020. How much knowledge can you pack into the parameters of a language model? *Preprint*, arXiv:2002.08910.

Imad Sameer, Hazim Al Dilaimy, Al Dilaimy, and Linguistics. 2020. Ideational aspect of systemic functional grammar in bush's and al-assad's first inaugural speech. 12:1203–1218.

Hawsar Shekhani. 2022. *Halliday's Interpersonal Metafunction Analysis of Entrepreneurs' Public Speeches in English and Central Kurdish: A Comparative Study*. Ph.D. thesis.

Zhengliang Shi, Shuo Zhang, Weiwei Sun, Shen Gao, Pengjie Ren, Zhumin Chen, and Zhaochun Ren. 2024. Generate-then-ground in retrieval-augmented generation for multi-hop question answering. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7339–7353, Bangkok, Thailand. Association for Computational Linguistics.

Zhiqing Sun, Xuezhi Wang, Yi Tay, Yiming Yang, and Denny Zhou. 2023. Recitation-augmented language models. *Preprint*, arXiv:2210.01296.

Qwen Team. 2024. Qwen2.5: A party of foundation models.

James Thorne, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal. 2018. Fever: a large-scale dataset for fact extraction and verification. *Preprint*, arXiv:1803.05355.

Harsh Trivedi, Niranjan Balasubramanian, Tushar Khot, and Ashish Sabharwal. 2022. Interleaving retrieval with chain-of-thought reasoning for knowledge-intensive multi-step questions. *arXiv preprint arXiv:2212.10509*.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, brian ichter, Fei Xia, Ed Chi, Quoc V Le, and Denny Zhou. 2022. Chain-of-thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems*, volume 35, pages 24824–24837. Curran Associates, Inc.

Shitao Xiao, Zheng Liu, Peitian Zhang, and Niklas Muennighoff. 2023. C-pack: Packaged resources to advance general chinese embedding. *Preprint*, arXiv:2309.07597.

Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William W. Cohen, Ruslan Salakhutdinov, and Christopher D. Manning. 2018. Hotpotqa: A dataset for diverse, explainable multi-hop question answering. *Preprint*, arXiv:1809.09600.

Wenhao Yu, Dan Iter, Shuohang Wang, Yichong Xu, Mingxuan Ju, Soumya Sanyal, Chenguang Zhu, Michael Zeng, and Meng Jiang. 2023. Generate rather than retrieve: Large language models are strong context generators. *Preprint*, arXiv:2209.10063.

Yue Yu, Wei Ping, Zihan Liu, Boxin Wang, Jiaxuan You, Chao Zhang, Mohammad Shoeybi, and Bryan Catanzaro. 2024. Rankrag: Unifying context ranking with retrieval-augmented generation in llms. *Preprint*, arXiv:2407.02485.

Michael Zhang and Eunsol Choi. 2021. Situatedqa: Incorporating extra-linguistic contexts into qa. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 7371–7387.

Yunxiang Zhang, Muhammad Khalifa, Lajanugen Logeswaran, Moontae Lee, Honglak Lee, and Lu Wang. 2023. Merging generated and retrieved knowledge for open-domain qa. *Preprint*, arXiv:2310.14393.

Chunting Zhou, Graham Neubig, Jiatao Gu, Mona Diab, Francisco Guzmán, Luke Zettlemoyer, and Marjan Ghazvininejad. 2021. Detecting hallucinated content in conditional neural sequence generation. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1393–1404, Online. Association for Computational Linguistics.

A Pseudocode

Algorithm 1 Direct Mix Strategy for Document Mixing

Input: Set of ranked Wiki documents W, set of unordered self-generated documents (SGDs) S from taxonomy category T, number of top Wiki documents k, number of SGDs n

Output: Mixed document set $\mathcal{M}$

Initialize empty list $\mathcal{M}$ to store the mixed document set
Select Top Wiki documents:
Retrieve and select the top k Wiki documents from W based on rank order
Add selected Wiki documents to $\mathcal{M}$
Select SGDs from Taxonomy Category:
Select the first n SGDs from S, where S belongs to a specific category T in the SGD taxonomy (e.g., *Authoritative-Coarse-Grained-Structured*)
Add selected SGDs to $\mathcal{M}$
Final Mixed Document:

$$D_{\text{Mixed}} = \text{Concat}(D_{\text{SGD}}, D_{\text{Wiki}})$$

return $\mathcal{M}$

Algorithm 2 Mix with Style Transformation Strategy

Input: Set of ranked Wiki documents W, set of unordered self-generated documents (SGDs) S from taxonomy category T, number of top Wiki documents k, number of SGDs n, transformation function $T(W_i)$

Output: Mixed document set $\mathcal{M}$

Initialize empty list $\mathcal{M}$ to store the mixed document set
Select Top Wiki documents:
for each W_i in the top k Wiki documents from W **do**
    Apply style transformation: $T(W_i)$
    Add transformed document $T(W_i)$ to $\mathcal{M}$
end for
Select SGDs from Taxonomy Category:
Select the first n SGDs from S, where S belongs to a specific category T in the SGD taxonomy (e.g., *Authoritative, Conversational, etc.*)
Add selected SGDs to $\mathcal{M}$
Final Mixed Document:

$$D_{\text{Mixed}} = \text{Concat}\left(\bigcup_{i=1}^{k} T(W_i), \bigcup_{j=1}^{n} S_j\right)$$

return $\mathcal{M}$

B Detailed Results

Dataset	Qwen2.5-32B-Instruct	Qwen2.5-14B-Instruct	Qwen2.5-7B-Instruct	Qwen2.5-3B-Instruct
TQA	56.6	55.0	45.2	33.0
FEVER	88.8	76.8	64.0	72.4
HOTPOTQA	37.2	38.3	32.9	25.1
ELI5	24.4	23.8	22.1	20.9

Table 6: Performance metrics across different datasets and model sizes under the self-generated documents (SGDs) setting, with $n = 10$ documents. Both the QA and context generation models are from the Qwen2.5 series, ranging from 3B to 32B parameters.

Info Source	Qwen2.5-32B-Instruct	Qwen2.5-14B-Instruct	Qwen2.5-7B-Instruct	Qwen2.5-3B-Instruct	Wiki	No-RAG
TQA	56.6	53.2	45.8	39.6	52.6	52.0
FEVER	88.8	83.8	78.8	73.6	66.8	77.2
HOTPOTQA	37.2	39.1	33.0	26.0	36.3	32.9
ELI5	24.4	23.5	22.7	22.8	26.7	26.8

Table 7: Performance comparison across different information sources (Qwen2.5 models of varying sizes, Wiki, and No-RAG) on multiple datasets. This table highlights the impact of model size and information source on retrieval-augmented tasks like TQA, FEVER, HotpotQA, and ELI5.

Dataset	n=10	n=5	n=3	n=1
TQA	56.6	54.0	52.0	49.6
FEVER	88.8	87.6	86.6	83.8
HOTPOTQA	37.2	36.0	36.1	34.4
ELI5	24.4	24.7	25.1	26.0

Table 8: Performance across various document numbers (n) for different datasets. The results show how the number of self-generated documents impacts model performance, with n values set to 10, 5, 3, and 1.

Type	Interpersonal	Ideational	Textual	TQA	HOTPOTQA	FEVER	ELI5	Average
AFS	Authoritative	Fine-Grain	Structured	56.0↑0.6	38.3↑0.8	73.6↓9.8	22.0↑1.1	47.5↓1.8
AFU	Authoritative	Fine-Grain	Unstructured	56.8↓1.6	41.0↑1.1	84.4↓0.8	23.1↑0.3	51.3↓0.3
ACS	Authoritative	Coarse-Grain	Structured	55.8↓1.0	39.5↑0.8	84.8↑1.2	22.7↑0.7	50.7↑0.4
ACU	Authoritative	Coarse-Grain	Unstructured	56.2↓0.2	40.8↑1.2	86.6↑1.6	23.8↑0.3	51.8↑0.7
CFS	Conversational	Fine-Grain	Structured	56.4↓1.0	39.0↑2.6	78.2↓3.6	24.8↑0.8	49.6↓0.3
CFU	Conversational	Fine-Grain	Unstructured	57.2↓0.6	41.6↑2.1	85.0↑0.8	25.6↑0.5	52.4↑0.8
CCS	Conversational	Coarse-Grain	Structured	56.4↑0.4	40.1↑4.9	84.0↑2.0	25.7↑0.3	51.6↑2.0
CCU	Conversational	Coarse-Grain	Unstructured	57.4↑0.8	**41.8**↑3.1	85.8↑2.0	26.5↑0.1	**52.9**↑1.5
WIKI	Authoritative	Flexible	Unstructured	52.6	36.3	66.8	**26.7**	45.6
GENREAD MIX	Authoritative	Flexible	Unstructured	**57.6**	39.1	**87.0**	24.7	52.1

Table 9: Performance comparison across different generated document types combined with Wiki documents for various tasks. Superscript values indicate the difference from Table 3, with increases marked by an upward arrow (↑) and decreases by a downward arrow (↓). The best overall performance for each dataset is highlighted in **bold**, and the best performance within the category of SGDs with the same mix strategy is underlined.

Meta Function	Dimension	TQA	HOTPOTQA	FEVER	ELI5	Average
IDEATIONAL	FINE-GRAIN	56.6↓0.6	40.0↑1.7	80.3↓3.3	23.9↑0.7	50.2↓0.4
	COARSE-GRAIN	56.4±0.0	40.6↑2.5	85.3↑1.7	24.7↑0.4	51.8↑1.2
INTERPERSONAL	AUTHORITATIVE	56.2↓0.6	39.9↑1.0	82.4↓1.9	22.9↑0.6	50.4↓0.2
	CONVERSATIONAL	56.8↓0.2	40.6↑3.2	83.2↑0.2	25.6↑0.4	51.6↑0.9
TEXTUAL	STRUCTURED	56.2↓0.2	39.2↑2.2	80.1↓2.6	23.8↑0.7	49.8±0.0
	UNSTRUCTURED	56.9↓0.4	41.3↑1.9	85.4↑0.8	24.8↑0.3	51.6↑0.2

Table 10: Aggregated performance results across different dimensions of the metafunctions (Ideational, Interpersonal, and Textual) on various datasets. Superscript values indicate the difference from Table 4, with increases marked by an upward arrow (↑) and decreases by a downward arrow (↓). The best performance within each metafunction dimension is underlined.

Docs Type	Interpersonal	Ideational	Textual	TQA	HOTPOTQA	FEVER	ELI5	Average
AFS	Authoritative	Fine-Grain	Structured	56.6↑1.2	37.5±0.0	87.6↑4.2	21.3↑0.4	50.8↑1.5
AFU	Authoritative	Fine-Grain	Unstructured	**59.4**↑1.0	**43.3**↑3.4	87.6↑2.4	22.7↓0.1	53.3↑1.7
ACS	Authoritative	Coarse-Grain	Structured	56.8±0.0	42.4↑3.7	88.8↑5.2	22.2↑0.2	52.6↑2.3
ACU	Authoritative	Coarse-Grain	Unstructured	57.0↑0.6	41.9↑2.3	**90.4**↑5.4	23.3↓0.2	53.2↑2.1
CFS	Conversational	Fine-Grain	Structured	57.0↓0.4	39.3↑2.9	86.0↑4.2	24.2↑0.2	51.6↑1.7
CFU	Conversational	Fine-Grain	Unstructured	58.0↑0.2	41.0↑1.5	86.4↑2.2	25.3↑0.2	52.7↑1.1
CCS	Conversational	Coarse-Grain	Structured	57.4↑1.4	41.8↑6.6	86.6↑4.6	25.6↑0.2	52.9↑3.3
CCU	Conversational	Coarse-Grain	Unstructured	57.6↑1.0	43.0↑4.3	87.4↑3.6	26.3↓0.1	**53.6**↑2.2
WIKI	Authoritative	Flexible	Unstructured	52.6	36.3	66.8	**26.7**	45.6
GENREAD MIX	Authoritative	Flexible	Unstructured	57.6	39.1	87.0	24.7	52.1

Table 11: Performance comparison across different generated document types combined with Wiki documents using mix with style transformation strategy for various tasks. Superscript values indicate the difference from Table 3, with increases marked by an upward arrow (↑) and decreases by a downward arrow (↓). The best overall performance for each dataset is highlighted in **bold**, and the best performance within the category of SGDs with the same mix strategy is underlined.

Meta Function	Dimension	TQA	HOTPOTQA	FEVER	ELI5	Average
IDEATIONAL	FINE-GRAIN	57.8$_{\uparrow 0.6}$	40.3$_{\uparrow 2.0}$	86.9$_{\uparrow 3.3}$	23.4$_{\uparrow 0.2}$	52.1$_{\uparrow 1.5}$
	COARSE-GRAIN	57.2$_{\uparrow 0.8}$	42.3$_{\uparrow 4.2}$	88.3$_{\uparrow 4.7}$	24.3$_{\pm 0.0}$	53.0$_{\uparrow 2.4}$
INTERPERSONAL	AUTHORITATIVE	56.5$_{\downarrow 0.3}$	40.3$_{\uparrow 1.4}$	84.2$_{\downarrow 0.1}$	23.2$_{\uparrow 0.9}$	51.0$_{\uparrow 0.4}$
	CONVERSATIONAL	57.5$_{\uparrow 0.5}$	41.3$_{\uparrow 3.9}$	86.6$_{\uparrow 3.6}$	25.3$_{\uparrow 0.1}$	52.7$_{\uparrow 2.0}$
TEXTUAL	STRUCTURED	57.0$_{\uparrow 0.6}$	40.2$_{\uparrow 3.2}$	87.2$_{\uparrow 4.5}$	23.3$_{\uparrow 0.2}$	51.9$_{\uparrow 2.1}$
	UNSTRUCTURED	56.9$_{\downarrow 0.4}$	41.1$_{\uparrow 1.7}$	83.7$_{\downarrow 0.9}$	24.9$_{\uparrow 0.4}$	51.6$_{\uparrow 0.2}$

Table 12: Aggregated performance results across different dimensions of the metafunctions (Ideational, Interpersonal, and Textual) on various datasets. Superscript values indicate the difference from Table 4, with increases marked by an upward arrow (↑) and decreases by a downward arrow (↓). The best performance within each metafunction dimension is underlined.

C Implementation Details

This study excludes the Natural Questions (NQ) dataset due to several issues identified in previous work. Notably, a substantial portion of NQ questions (approximately 16.5%) are time-sensitive, leading to potential evaluation inconsistencies when temporal context is not explicitly provided (Zhang and Choi, 2021; Yu et al., 2023). Additionally, using different versions of Wikipedia for retrieval can result in significant performance fluctuations (Izacard et al., 2022; Yu et al., 2023). Further, GenRead's error analysis revealed data quality issues in NQ, including annotation mistakes, incomplete answers, and retrieval inconsistencies (Yu et al., 2023). Given these challenges, incorporating NQ would hinder our goal of rigorous and accurate assessment.

For experiments related to RQ1, we follow the methodology outlined in *GenRead* (Yu et al., 2023), using the `bge-large-en-v1.5` model (Xiao et al., 2023) for embedding both questions and documents, ensuring consistency across experimental settings. Document generation prompts are also aligned with those used in *GenRead* to maintain methodological uniformity.

However, for generating different types of self-generated documents (SGDs), we deviate from *GenRead*'s shot-based generation strategy. We adopt a GPT-4o-style approach, adapting generated documents based on Wiki contexts retrieved using Contriever (Izacard et al., 2021). In this setup, GPT-4o is prompted to generate style-transformed documents adhering to predefined styles, based on the provided Wiki context under a zero-shot setting. We provide GPT-4o with the Wiki context and a task-specific prompt to produce documents in the desired format.

To assess the impact of document number, we experiment with (n = 1, 3, 5, 10) documents per setting. During document generation, we set the temperature parameter to 0.95 to encourage diversity in the output. For the question answering inference stage, we use greedy decoding with the temperature set to 0 for deterministic outputs and improved reproducibility.

D Manual Evaluation

To assess whether the self-generated documents (SGDs) align with their expected types, we conducted a manual evaluation of 120 document samples from the TQA dataset, covering the eight SGD types outlined in Table 1. We selected 15 samples from each SGD type and manually evaluated how well each document matched its designated interpersonal, ideational, and textual metafunctions. The success rates are presented in Table 13.

Type	Interpersonal	Ideational	Textual	Success Rate (%)
AFS	Authoritative	Fine-Grain	Structured	86.7
AFU	Authoritative	Fine-Grain	Unstructured	86.7
ACS	Authoritative	Coarse-Grain	Structured	93.3
ACU	Authoritative	Coarse-Grain	Unstructured	100.0
CFS	Conversational	Fine-Grain	Structured	93.3
CFU	Conversational	Fine-Grain	Unstructured	93.3
CCS	Conversational	Coarse-Grain	Structured	100.0
CCU	Conversational	Coarse-Grain	Unstructured	100.0

Table 13: Success rates of self-generated document types classified based on the three SFL metafunctions: interpersonal, ideational, and textual.

Each document was evaluated for its consistency with the SFL metafunctions, and the results indicate high success rates across all types, with some achieving perfect alignment. These results underscore the reliability of the SGD taxonomy in producing documents that align with their expected communicative roles.

E Prompts

We provide detailed prompts used throughout the experiment. The prompt for shot generation can be found in Section E.1, while the prompt for generating context documents is presented in Section E.2. Finally, the prompt used for the QA tasks is outlined in Section E.3. These prompts ensure consistency and transparency across the different stages of our experimental setup.

E.1 Prompt for shot generation

Prompt E.1: TQA, HotpotQA, Eli5 - AFS

Given the context: {context} and the query: {query}, transform the context into an authoritative, fine-grain, structured document.
The tone should be authoritative, maintaining a formal and objective style.
The content should be fine-grain, providing detailed explanations and in-depth analysis of each aspect.
The structure should be clear with proper headings and sections to logically organize the content.

Prompt E.2: TQA, HotpotQA, Eli5 - AFU

Given the context: {context} and the query: {query}, transform the context into an authoritative, fine-grain, unstructured document.
The tone should be authoritative, maintaining a formal and objective style.
The content should be fine-grain, providing detailed explanations and in-depth analysis of each aspect.
The structure should be free-flowing, allowing ideas to develop organically without strict headings.

Prompt E.3: TQA, HotpotQA, Eli5-ACS

Given the context: {context} and the query: {query}, transform the context into an authoritative, coarse-grain, structured document.
The tone should be authoritative, maintaining a formal and objective style.
The content should be coarse-grain, offering a broad overview and key points without deep details.
The structure should be clear with proper headings and sections to logically organize the content.

Prompt E.4: TQA, HotpotQA, Eli5 - ACU

Given the context: {context} and the query: {query}, transform the context into an authoritative, coarse-grain, unstructured document.
The tone should be authoritative, maintaining a formal and objective style.
The content should be coarse-grain, offering a broad overview and key points without deep details.
The structure should be free-flowing, allowing ideas to develop organically without strict headings.

Prompt E.5: TQA, HotpotQA, Eli5 - CFS

Given the context: {context} and the query: {query}, transform the context into a conversational, fine-grain, structured document.
The tone should be conversational, using an informal and approachable style.
The content should be fine-grain, providing detailed explanations and in-depth analysis of each aspect.
The structure should be clear with proper headings and sections to logically organize the content.

Prompt E.6: TQA, HotpotQA, Eli5 - CFU

Given the context: {context} and the query: {query}, transform the context into a conversational, fine-grain, unstructured document.
The tone should be conversational, using an informal and approachable style.
The content should be fine-grain, providing detailed explanations and in-depth analysis of each aspect.
The structure should be free-flowing, allowing ideas to develop organically without strict headings.

Prompt E.7: TQA, HotpotQA, Eli5 - CCS

Given the context: {context} and the query: {query}, transform the context into a conversational, coarse-grain, structured document.
The tone should be conversational, using an informal and approachable style.
The content should be coarse-grain, offering a broad overview and key points without deep details.
The structure should be clear with proper headings and sections to logically organize the content.

Prompt E.8: TQA, HotpotQA, Eli5 - CCU

Given the context: {context} and the query: {query}, transform the context into a conversational, coarse-grain, unstructured document.
The tone should be conversational, using an informal and approachable style.
The content should be coarse-grain, offering a broad overview and key points without deep details.
The structure should be free-flowing, allowing ideas to develop organically without strict headings.

Prompt E.9: Fever - AFS

Given the context: {context} and the claim: {query}, transform the context into an authoritative, fine-grain, structured document to support or refute the claim.
The tone should be authoritative, maintaining a formal and objective style.
The content should be fine-grain, providing detailed explanations and in-depth analysis of each aspect.
The structure should be clear with proper headings and sections to logically organize the content.

Prompt E.10: Fever - AFU

Given the context: {context} and the claim: {query}, transform the context into an authoritative, fine-grain, unstructured document to support or refute the claim.
The tone should be authoritative, maintaining a formal and objective style.
The content should be fine-grain, providing detailed explanations and in-depth analysis of each aspect.
The structure should be free-flowing, allowing ideas to develop organically without strict headings.

Prompt E.11: Fever - ACS

Given the context: {context} and the claim: {query}, transform the context into an authoritative, coarse-grain, structured document to support or refute the claim.
The tone should be authoritative, maintaining a formal and objective style.
The content should be coarse-grain, offering a broad overview and key points without deep details.
The structure should be clear with proper headings and sections to logically organize the content.

Prompt E.12: Fever - ACU

Given the context: {context} and the claim: {query}, transform the context into an authoritative, coarse-grain, unstructured document to support or refute the claim.
The tone should be authoritative, maintaining a formal and objective style.
The content should be coarse-grain, offering a broad overview and key points without deep details.
The structure should be free-flowing, allowing ideas to develop organically without strict headings.

Prompt E.13: Fever - CFS

Given the context: {context} and the claim: {query}, transform the context into a conversational, fine-grain, structured document to support or refute the claim.
The tone should be conversational, using an informal and approachable style.
The content should be fine-grain, providing detailed explanations and in-depth analysis of each aspect.
The structure should be clear with proper headings and sections to logically organize the content.

Prompt E.14: Fever - CFU

Given the context: {context} and the claim: {query}, transform the context into a conversational, fine-grain, unstructured document to support or refute the claim.
The tone should be conversational, using an informal and approachable style.
The content should be fine-grain, providing detailed explanations and in-depth analysis of each aspect.
The structure should be free-flowing, allowing ideas to develop organically without strict headings.

Prompt E.15: Fever - CCS

Given the context: {context} and the claim: {query}, transform the context into a conversational, coarse-grain, structured document to support or refute the claim.
The tone should be conversational, using an informal and approachable style.
The content should be coarse-grain, offering a broad overview and key points without deep details.
The structure should be clear with proper headings and sections to logically organize the content.

Prompt E.16: Fever - CCU

Given the context: {context} and the claim: {query}, transform the context into a conversational, coarse-grain, unstructured document to support or refute the claim.
The tone should be conversational, using an informal and approachable style.
The content should be coarse-grain, offering a broad overview and key points without deep details.
The structure should be free-flowing, allowing ideas to develop organically without strict headings.

E.2 Prompt for document generation

Prompt E.17: TQA, HotpotQA, Eli5 - RQ1 - GenRead

Provide a background document from Wikipedia to answer the given question.

{query}

{wiki top1 context}

Provide a background document from Wikipedia to answer the given question.

{query}

{wiki top1 context}

Provide a background document from Wikipedia to answer the given question.

{query}

{wiki top1 context}

Provide a background document from Wikipedia to answer the given question.

{query}

{wiki top1 context}

Provide a background document from Wikipedia to answer the given question.

{query}

{wiki top1 context}

Provide a background document from Wikipedia to answer the given question.

{query}

Prompt E.18: Fever - RQ1 - GenRead

Generate a background document from Wikipedia to support or refute the statement.

Statement: {query}

{wiki top1 context}

Generate a background document from Wikipedia to support or refute the statement.

Statement: {query}

{wiki top1 context}

Generate a background document from Wikipedia to support or refute the statement.

Statement: {query}

{wiki top1 context}

Generate a background document from Wikipedia to support or refute the statement.

Statement: {query}

{wiki top1 context}

Generate a background document from Wikipedia to support or refute the statement.

Statement: {query}

{wiki top1 context}

Generate a background document from Wikipedia to support or refute the statement.

Statement: {query}

Prompt E.19: TQA, HotpotQA, Eli5 - AFS

System: You are tasked with generating a detailed and authoritative response. Maintain a formal and objective tone while providing in-depth analysis. The document should have a structured format with clear headings and sections that logically organize the content, offering precise explanations of each aspect.

User: Given the query: {query}, generate an authoritative, fine-grain, structured document.
The tone should be authoritative, maintaining a formal and objective style.
The content should be fine-grain, providing detailed explanations and in-depth analysis of each aspect.
The structure should be clear with proper headings and sections to logically organize the content.

Prompt E.20: TQA, HotpotQA, Eli5 - AFU

System: You are tasked with generating a detailed and authoritative response in a free-flowing format. The tone should remain formal and objective, while the content offers in-depth analysis. The document should allow ideas to develop naturally without the constraint of rigid headings, encouraging an organic flow of detailed information.

User: Given the query: {query}, generate an authoritative, coarse-grain, structured document.
The tone should be authoritative, maintaining a formal and objective style.
The content should be coarse-grain, offering a broad overview and key points without deep details.
The structure should be clear with proper headings and sections to logically organize the content.

Prompt E.21: TQA, HotpotQA, Eli5 - ACU

System: You are tasked with generating an authoritative, high-level overview in a free-flowing format. The tone should remain formal and objective, while the content offers broad explanations without going into detail. The document should allow ideas to develop organically without strict structure, focusing on the overall flow of the argument.

User: Given the query: {query}, generate an authoritative, coarse-grain, unstructured document.
The tone should be authoritative, maintaining a formal and objective style.
The content should be coarse-grain, offering a broad overview and key points without deep details.
The structure should be free-flowing, allowing ideas to develop organically without strict headings.

Prompt E.22: TQA, HotpotQA, Eli5 - CFS

System: You are tasked with generating a detailed and engaging document. Use a conversational tone that is approachable and informal while maintaining a structured format. The content should provide in-depth explanations, with clear headings and sections that logically guide the reader through the document.

User: Given the query: {query}, generate a conversational, fine-grain, structured document.
The tone should be conversational, using an informal and approachable style.
The content should be fine-grain, providing detailed explanations and in-depth analysis of each aspect.
The structure should be clear with proper headings and sections to logically organize the content.

Prompt E.23: TQA, HotpotQA, Eli5 - CFU

System: You are tasked with generating a detailed and engaging response in a conversational tone. Use an informal and approachable style while allowing the document to flow naturally without strict headings or sections. Provide in-depth explanations while maintaining a free-flowing, organic narrative structure.

User: Given the query: {query}, generate a conversational, fine-grain, unstructured document.
The tone should be conversational, using an informal and approachable style.
The content should be fine-grain, providing detailed explanations and in-depth analysis of each aspect.
The structure should be free-flowing, allowing ideas to develop organically without strict headings.

Prompt E.24: TQA, HotpotQA, Eli5 - CCS

System: You are tasked with generating a conversational, high-level overview. Maintain an informal and approachable tone while organizing the content into clear headings and sections. The content should offer a broad overview, summarizing key points without going into much detail.

User: Given the query: {query}, generate a conversational, coarse-grain, structured document.
The tone should be conversational, using an informal and approachable style.
The content should be coarse-grain, offering a broad overview and key points without deep details.
The structure should be clear with proper headings and sections to logically organize the content.

Prompt E.25: TQA, HotpotQA, Eli5 - CCU

System: You are tasked with generating a conversational, broad overview in a free-flowing format. Use an informal and approachable tone while allowing the content to develop organically without a structured format. The document should provide key points and high-level explanations without strict headings or detailed sections.

User: Given the query: {query}, generate a conversational, coarse-grain, unstructured document.
The tone should be conversational, using an informal and approachable style.
The content should be coarse-grain, offering a broad overview and key points without deep details.
The structure should be free-flowing, allowing ideas to develop organically without strict headings.

Prompt E.26: Fever - AFS

System: You are tasked with generating a detailed, authoritative analysis to support or refute a claim. Maintain a formal and objective tone, providing in-depth analysis and explanations of each aspect. The document should have a structured format with clear headings and sections, logically organizing the reasoning.

User: Given the claim: {claim}, generate an authoritative, fine-grain, structured document to support or refute the claim.
The tone should be authoritative, maintaining a formal and objective style.
The content should be fine-grain, providing detailed explanations and in-depth analysis of each aspect.
The structure should be clear with proper headings and sections to logically organize the content.

Prompt E.27: Fever - AFU

System: You are tasked with generating a detailed, authoritative analysis in a free-flowing format to support or refute a claim. Maintain a formal and objective tone, while allowing the document to develop naturally without rigid headings. The content should provide in-depth explanations and analysis, letting ideas unfold organically.

User: Given the claim: {claim}, generate an authoritative, fine-grain, unstructured document to support or refute the claim.
The tone should be authoritative, maintaining a formal and objective style.
The content should be fine-grain, providing detailed explanations and in-depth analysis of each aspect.
The structure should be free-flowing, allowing ideas to develop organically without strict headings.

Prompt E.28: Fever - ACS

System: You are tasked with generating a broad, authoritative analysis to support or refute a claim. Maintain a formal and objective tone, summarizing the key points without deep details. The document should have clear headings and sections to logically organize the argument.

User: Given the claim: {claim}, generate an authoritative, coarse-grain, structured document to support or refute the claim.
The tone should be authoritative, maintaining a formal and objective style.
The content should be coarse-grain, offering a broad overview and key points without deep details.
The structure should be clear with proper headings and sections to logically organize the content.

Prompt E.29: Fever - ACU

System: You are tasked with generating a high-level, authoritative analysis in a free-flowing format to support or refute a claim. Maintain a formal and objective tone while allowing the content to develop organically without structured sections. Summarize key points without delving into too much detail.

User: Given the claim: {claim}, generate an authoritative, coarse-grain, unstructured document to support or refute the claim.
The tone should be authoritative, maintaining a formal and objective style.
The content should be coarse-grain, offering a broad overview and key points without deep details.
The structure should be free-flowing, allowing ideas to develop organically without strict headings.

Prompt E.30: Fever - CFS

System: You are tasked with generating a detailed, conversational analysis to support or refute a claim. Use an informal and approachable tone while maintaining a structured format with clear headings and sections. Provide in-depth explanations and a logical argument to support or refute the claim.

User: Given the claim: {claim}, generate a conversational, fine-grain, structured document to support or refute the claim.
The tone should be conversational, using an informal and approachable style.
The content should be fine-grain, providing detailed explanations and in-depth analysis of each aspect.
The structure should be clear with proper headings and sections to logically organize the content.

Prompt E.31: Fever - CFU

System: You are tasked with generating a detailed, conversational analysis to support or refute a claim in a free-flowing format. Use an informal and approachable tone while allowing the content to develop naturally without rigid structure. Provide in-depth explanations and analysis while letting ideas flow organically.

User: Given the claim: {claim}, generate a conversational, fine-grain, unstructured document to support or refute the claim.
The tone should be conversational, using an informal and approachable style.
The content should be fine-grain, providing detailed explanations and in-depth analysis of each aspect.
The structure should be free-flowing, allowing ideas to develop organically without strict headings.

Prompt E.32: Fever - CCS

System: You are tasked with generating a conversational, high-level analysis to support or refute a claim. Use an informal and approachable tone while organizing the content into clear headings and sections. Summarize key points without going into much detail, maintaining a broad overview.

User: Given the claim: {claim}, generate a conversational, coarse-grain, structured document to support or refute the claim.
The tone should be conversational, using an informal and approachable style.
The content should be coarse-grain, offering a broad overview and key points without deep details.
The structure should be clear with proper headings and sections to logically organize the content.

Prompt E.33: Fever - CCU

System: You are tasked with generating a conversational, high-level analysis in a free-flowing format to support or refute a claim. Use an informal and approachable tone, allowing ideas to develop organically without rigid structure. Summarize key points without going into too much detail.

User: Given the claim: {claim}, generate a conversational, coarse-grain, unstructured document to support or refute the claim.
The tone should be conversational, using an informal and approachable style.
The content should be coarse-grain, offering a broad overview and key points without deep details.
The structure should be free-flowing, allowing ideas to develop organically without strict headings.

E.3 Prompt for question answering

Prompt E.34: TQA, HotpotQA - No RAG

System: This is a chat between a user and an artificial intelligence assistant. The assistant gives helpful, detailed, and polite answers to the user's questions.

{query}. Answer the above question. Your output should strictly be one entity.

Prompt E.35: TQA, HotpotQA - RAG

System: This is a chat between a user and an artificial intelligence assistant. The assistant gives helpful, detailed, and polite answers to the user's questions based on the context. The assistant should also indicate when the answer cannot be found in the context.

{background} User: {query}. Answer the above question. Your output should strictly be one entity.

Prompt E.36: Fever - No RAG

System: This is a chat between a user and an artificial intelligence assistant. The assistant gives helpful, detailed, and polite answers to the user's questions.

Answer the following question with True or False. Is the claim {claim} correct? Your output should be either True or False.

Prompt E.37: Fever - RAG

System: This is a chat between a user and an artificial intelligence assistant. The assistant gives helpful, detailed, and polite answers to the user's questions based on the context. The assistant should also indicate when the answer cannot be found in the context.

{background} User: Answer the following question with True or False. Is the claim {claim} correct? Your output should be either True or False.

Prompt E.38: Eli5 - No RAG

System: This is a chat between a user and an artificial intelligence assistant. The assistant gives helpful, detailed, and polite answers to the user's questions.

{query}

Prompt E.39: Eli5 - RAG

System: This is a chat between a user and an artificial intelligence assistant. The assistant gives helpful, detailed, and polite answers to the user's questions based on the context.

{background} User: {query}

F Case Study

In this section, we present a case study leveraging the Qwen2.5-32B-Instruct model to assess its performance on self-generated documents (SGDs) across various document types, as defined by the full taxonomy in F.2. This analysis explores how different SGD types, classified according to interpersonal tone, ideational content, and textual structure, impact overall task performance. Additionally, we provide a case study evaluating the performance of GPT-4o-generated document shots in F.1, demonstrating the effects of varying document styles on specific tasks.

F.1 Examples of Self-generated Document Shots

Dataset: TQA

Authoritative, Fine-Grain, Structured (AFS)

```
# The Sinatra Doctrine: A Foreign Policy Framework

## Introduction

The Sinatra Doctrine represents a pivotal shift in
Soviet foreign policy, distinguishing itself by
acknowledging the autonomy of Warsaw Pact countries in
 choosing their political paths. Named after Frank
Sinatra's iconic song "My Way," this doctrine marked a
 significant departure from previous Soviet positions
of influence over Eastern Europe.

## Historical Context

### The Soviet Union and Warsaw Pact Dynamics

The Warsaw Pact was a collective defense treaty
established in 1955 among the Soviet Union and seven
Eastern Bloc socialist republics during the Cold War.
Traditionally, the Soviet Union exerted considerable
influence over these states, often dictating their
political and economic directions.

### Shift in Policy

The introduction of the Sinatra Doctrine emerged from
evolving geopolitical realities in the late 20th
century. As pressures for reform and democratization
mounted across Eastern Europe, the Soviet leadership
recognized the need to adapt its foreign policy stance
 to accommodate these changing dynamics.

## Formulation of the Sinatra Doctrine

The Sinatra Doctrine was articulated by Gennadi
Gerasimov, a prominent Soviet government spokesperson,
 during an interview. He referenced Frank Sinatra's
song "I Did It My Way" to illustrate the doctrines
essence: allowing each country to independently choose
 its path without imposed external influence.

### Key Statements

- **Gerasimov's Declaration:** Gerasimov affirmed, "We
 now have the Frank Sinatra doctrine. He has a song, '
I Did It My Way.' So every country decides on its own
which road to take."
- **Acceptance of Political Diversity:** When queried
about the potential rejection of communist parties
within the Soviet bloc, Gerasimov stated, "That's for
sure... political structures must be decided by the
people who live there."

......

## Conclusion

The Sinatra Doctrine marked a significant
transformation in Soviet foreign policy, emphasizing
respect for national sovereignty and political self-
determination. As an acknowledgment of evolving global
 trends, this doctrine played a crucial role in
reshaping the political landscape of Eastern Europe
during the late 20th century. Its legacy persists as a
 testament to the power of ideological evolution in
response to geopolitical changes.
```

Authoritative, Fine-Grain, Unstructured (AFU)

```
In a significant ideological shift during the late 20
th century, the Soviet Union introduced a notable
transformation in its foreign policy approach,
commonly referred to as the "Sinatra Doctrine." This
strategic realignment was articulated by Soviet
Foreign Ministry spokesman Gennadi Gerasimov in a
pivotal moment of diplomatic evolution. The doctrine
was symbolically named after the famed American singer
 Frank Sinatra, specifically invoking his song "My Way
" to encapsulate the essence of this new policy.

The "Sinatra Doctrine" emerged as a clear departure
from previous Soviet policies, which had traditionally
 exerted significant control over the internal and
external affairs of the Warsaw Pact states.
Historically, the Soviet Union maintained a policy of
strict oversight, ensuring that the political and
economic systems of Eastern Bloc countries were
closely aligned with communist ideologies. However,
Gerasimovs declaration marked a decisive shift towards
 acknowledging the sovereignty and autonomy of these
nations.

During an interview, Gerasimov expounded on this new
ideological perspective, emphasizing that the Soviet
Union now recognized the principle of freedom of
choice for all countries, explicitly including its
Warsaw Pact allies. He articulated a vision where each
 nation possessed the authority to determine its own
pathway, unhindered by external pressures. His
statement, "We now have the Frank Sinatra doctrine. He
 has a song, 'I Did It My Way.' So every country
decides on its own which road to take," served as a
metaphorical representation of this newfound respect
for national self-determination.

......

In essence, the "Sinatra Doctrine" laid the foundation
 for a transformative era in international relations.
It embodied a philosophical pivot towards embracing
pluralism, encouraging organic political developments,
 and enhancing diplomatic autonomy within the Eastern
Bloc. Through this doctrine, the Soviet Union signaled
 a strategic retreat from its historically pervasive
interventionist stance, thereby acknowledging and
accommodating the complex realities of a rapidly
changing world order.
```

Authoritative, Coarse-Grain, Structured (ACS)

```
# The Sinatra Doctrine: An Overview of a Decisive
Policy Shift

## Introduction

The Sinatra Doctrine represents a significant shift in
 Soviet foreign policy during the late 20th century.
It emphasizes the autonomy of individual countries,
specifically those within the Warsaw Pact, to
determine their own political paths independently of
Soviet influence.

## Background

The doctrine emerged as a bold move acknowledging the
changing geopolitical landscape. It marks a departure
from previous Soviet policies that enforced strict
adherence to communism across the Eastern Bloc.

## Key Principles

### Autonomy of Nations

- **Recognition of Freedom**: The doctrine asserts the
 right of each nation to choose its political
structure and direction.
- **Rejection of Previous Controls**: It signifies a
move away from the rigid enforcement of communist
ideologies by the Soviet Union.

### Inspiration from Culture
```

- **Cultural Reference**: The doctrine derives its name from Frank Sinatra's famous song, "My Way," symbolizing the freedom of choice.
- **Philosophical Shift**: It underscores a philosophical shift towards respecting individual nation sovereignty.

.....

Conclusion

The Sinatra Doctrine stands as a pivotal moment in international relations, illustrating a refined approach to foreign policy grounded in the respect for national sovereignty. This realignment paved the way for significant political changes across Eastern Europe, marking the beginning of the end for Soviet dominance in the region.

Authoritative, Coarse-Grain, Unstructured (ACU)
In the context of evolving international relations during the late 20th century, a significant shift in Soviet foreign policy was articulated, highlighting a newfound recognition of national autonomy within the Eastern Bloc. This policy, informally dubbed the "Sinatra Doctrine," was articulated by a Soviet spokesperson. It was named after Frank Sinatra's iconic song "I Did It My Way," symbolizing the empowerment of each nation to independently determine their path.

The doctrine emerged as a pivotal change from previous policies, acknowledging the right of countries, particularly those within the Warsaw Pact, to exercise freedom of choice. This recognition extended to allowing them to decide their political structures without external interference, including the potential rejection of communist governments.

The authoritative declaration of this doctrine marked a departure from traditional Soviet influence, endorsing a more decentralized approach to governance among allied states. This strategy encouraged political determination by the populations residing within these nations, emphasizing self-determination as a cornerstone principle.

Through this declaration, the Soviet Union granted unprecedented autonomy to its neighbors, fostering a diplomatic environment where individual states were encouraged to pursue their unique political and social ideals. This transformation in policy underscored a broader movement towards recognizing the diversity of governance models within the international community, ultimately accepting that each country should establish its path in accordance with the will of its people.

Conversational, Fine-Grain, Structured (CFS)
The Sinatra Doctrine: A Conversational Breakdown

Introduction

Hey there! Let's dive into something intriguingly named the "Sinatra Doctrine." This wasn't some grand plan cooked up in a secret Kremlin meeting room, but rather a shift in thinking that signaled a big change in foreign policy. Its named after Frank Sinatras classic song, "My Way." Intrigued? Lets break it down.

What Is the Sinatra Doctrine?

The Basics

So, what exactly are we talking about? The "Sinatra Doctrine" refers to a policy approach by the Soviet Union during the late 1980s. Picture this: instead of the Soviets dictating how the satellites in their orbitlike those in the Warsaw Pactshould operate, they started singing a new tune inspired by "I Did It My Way." Each country could now choose its own path without Moscow pulling the strings.

Why Frank Sinatra?

You might be wondering why Frank Sinatra, of all people, was referenced for a Soviet policy. Well, it's because of his famous song "My Way." The idea was that just like in the song, every country could decide for itself what path to takepolitically, economically, and socially.

The Key Players

Who Said What?

This fresh approach was put into simple words by a Soviet spokesman named Gerasimov. In an interview, he cheekily declared, "We now have the Frank Sinatra Doctrine." It was his way of saying that the Soviets were ready to let countries in the Eastern Bloc find their groove and do things their way.

The Impact on Communism

Now, you might ask, did this mean the Soviets were okay with these countries rejecting communism? Gerasimov's answer was a straightforward "That's for sure." He emphasized that political decisions should be made by the people living in those countriestalking about major vibes of autonomy and independence!

.....

Conclusion

So, there you have itthe Sinatra Doctrine wasnt just about the Soviets letting go of control; it was about acknowledging the right of each nation to carve out its destiny. By evoking Sinatra's "My Way," it symbolized a significant turn in foreign policy, encouraging countries to embrace their sovereignty and decide their own future.

And thats it! Hope you enjoyed this deep dive into the surprising intersection of music and international policy.

Conversational, Fine-Grain, Unstructured (CFU)
Hey there! So, lets dive into this interesting bit about the so-called "Sinatra Doctrine," which is quite the shift from the old ways that were in place during the Soviet era.

Picture this: it's the late 1980s, and Mikhail Gorbachev comes into the scene with a fresh mindset. The world is changing, and the old Soviet grip on Eastern European countries is starting to loosen. Gorbachev's new man, Gerasimov, was kind of a big deal in this transition. He went on record saying that the Soviets were all about respecting the freedom of choice for all countries, which, at the time, was a pretty big deal.

It was in an interview where Gerasimov casually came up with the so-called "Sinatra Doctrine." You might be wondering why its named after Sinatra. Well, heres the fun part: it was a nod to Frank Sinatras famous song, "My Way." You know the one, right? It's all about doing things on your terms. Gerasimov said something like, "Were going with the Frank Sinatra Doctrine now. Every country gets to decide which road they want to take." This was sort of his cheeky way of saying that each country within the Soviet sphere could make its own choices.

Now, imagine someone asking Gerasimov, So, does this really mean that if a country in the Soviet bloc decides to toss out its communist regime, youd just accept that? And Gerasimov was like, Yeah, exactly. Its up to the people there to choose their political structures. This was a remarkable shift from the past when any deviation could have led to military interventions or other forms of pressure.

This whole "My Way" approach was radically different and paved the way for those countries in Eastern Europe to pursue their own paths without the fear of Soviet intervention. It was Gorbachevs way of signaling that the times were changing, and it opened the door for the fall of communism in several nations.

In essence, this doctrine wasnt some formal policy written in stone, but more a symbolic gesture indicating a real change in the political winds. It was about empowering nations to figure out their identities and futures independently. So, Gorbachev and his guys essentially said, You do you, and well do our own thing. It was a critical moment leading to the massive political shifts in Europe during that era .

So, there you have it! The Sinatra Doctrine wasn't about what the Soviet Union would do but rather about what they wouldn't dointervene in the choices of other countries. It was a big step away from the past, letting everyone sing their song My Way.

Conversational, Coarse-Grain, Structured (CCS)
The Sinatra Doctrine: A New Approach to Foreign Policy

Introduction

Hey there! Ever wondered how countries decide their own paths in the world of politics? Well, let's talk about something called the "Sinatra Doctrine." Named after Frank Sinatra's famous song "My Way," this approach is all about letting countries choose their own direction. Let's dive into what this means and who came up with it.

What is the Sinatra Doctrine?

The Idea

So, what's the big idea? The Sinatra Doctrine is all about freedom of choice. It suggests that each country has the right to decide its political path, like choosing the direction of a road trip.

The Name

Why call it the Sinatra Doctrine? It's inspired by Frank Sinatra's "I Did It My Way." Just like the song says, countries can do things their own way, without outside interference.

Who Created This Doctrine?

The Key Figure

This concept was put forward by Gerasimov. During an interview, he mentioned the idea of this "My Way" policy being recognized by the Soviets. It was a cool shift in thinking, especially for the countries in the Warsaw Pact.

A Big Change

When asked if this meant accepting the rejection of communist parties in Soviet bloc countries, Gerasimov confirmed it. He emphasized that political structures should be determined by the people living in those countries. A pretty big deal!

Final Thoughts

So, there you have it. The Sinatra Doctrine symbolizes a move towards autonomy for countries, letting them carve out their own paths. It's like a political remix of "I Did It My Way," encouraging nations to make choices that work for them.

And thats the scoop! Whether youre into politics or just enjoy a good classic song, the Sinatra Doctrine is a fascinating blend of both worlds.

Conversational, Coarse-Grain, Unstructured (CCU)
So, there was this interesting turn in Soviet foreign policy back in the day. The Soviets started chatting about how every country should have the freedom to choose its pathyes, even the other Warsaw Pact states. It was quite a shift from the usual hardline stance everyone expected from them.

Gerasimov, a Soviet official, was the one who came up with this cool analogy that caught on. He called it the "Frank Sinatra Doctrine." You know that famous Sinatra song, "I Did It My Way"? Well, that's essentially what the doctrine was about. It meant that countries could choose their own way, without interference.

Now, someone asked Gerasimov if this open-minded approach meant that Moscow would chill out if Soviet bloc countries decided to kick out their communist parties. His response was, "That's for sure." He said, quite plainly, that the political setup of a country should be in the hands of the people who actually live there.

It was a pretty big deal because it showed a more relaxed Soviet attitude towards how other countries handled their politics, especially in the Soviet sphere. It was like saying, "Hey, you do you, and we' ll do us." This approach made waves because it symbolized a new way of thinking for the Soviets during a time of change and uncertainty.

Dataset: FEVER

Authoritative, Fine-Grain, Structured (AFS)

Analysis of Kentucky's Musical Heritage

Introduction

Kentucky's musical heritage is rich and diverse, with deep roots in Appalachian folk traditions. This analysis examines the historical and cultural elements that have shaped Kentucky's music, assessing whether the state is indeed known for bluegrass music, a well-recognized genre today.

Historical Background

Early Musical Influences

- **Camp Meetings:** In 1800, the first camp meeting was held in Logan County, Kentucky. These religious gatherings often included singing, fostering a communal musical culture.
- **Cane Ridge Revival:** The 1801 meeting in Cane Ridge, Bourbon County, attracted between ten and twenty thousand people over six days, further embedding music into the community's social fabric.

20th Century Developments

- **Folk Music Collection:** In 1916, Loraine Wyman and Howard Brockway collected and published Kentucky folk music in two folios. This effort was crucial in preserving the state's Appalachian musical traditions.
- **Jug Band Music:** Louisville emerged as a center for some of the earliest jug band music, characterized by homemade, improvised instruments. This genre contributed to the broad spectrum of Kentuckys music.

.....

Conclusion

Kentuckys music is a tapestry of various influences, from religious camp meetings to folk and jug band music. The state holds a distinguished place in the development of American music, particularly due to its contribution to bluegrass. Given its historical roots and cultural evolution, the statement "Kentucky is known for bluegrass music" is well-supported and accurate.

This structured document provides a detailed exploration of Kentucky's musical traditions, establishing bluegrass as a defining genre of the state.

Authoritative, Fine-Grain, Unstructured (AFU)
In the early 19th century, Kentucky emerged as a significant site for religious and cultural gatherings, exemplified by the first camp meeting in Logan County in July 1800. This marked the beginning of a series of influential gatherings, with the 1801 Cane Ridge meeting in Bourbon County becoming a notable event. This meeting attracted between ten to twenty thousand attendees over six days, reflecting the region's burgeoning role as a center for cultural assembly and exchange.

As the 20th century approached, Kentucky's cultural landscape continued to evolve, preserving and enhancing its rich musical heritage. In 1916, collectors Loraine Wyman and Howard Brockway embarked on an ambitious project to document Kentucky's folk music traditions. Their efforts culminated in the publication of two folios, encapsulating the distinctive Appalachian influences that pervade the region's music. The emphasis on Appalachian folk music underscores Kentucky's deep-rooted connections to this genre, particularly in the eastern part of the state.

.....

Kentucky's prominent place in the development and proliferation of bluegrass music underscores its reputation as a key contributor to this genre. Bluegrass has since become an integral part of Kentuckys cultural identity, celebrated through festivals, concerts, and continued musical innovation within the state.

In conclusion, the assertion that Kentucky is known for bluegrass music is substantiated by the state's historical and cultural contributions to the genre. The evolution from its Appalachian folk roots to the creation of bluegrass illustrates Kentucky's enduring influence in American music, making the statement both accurate and reflective of Kentucky's rich musical heritage.

Authoritative, Coarse-Grain, Structured (ACS)
Music in Kentucky: An Overview

Introduction

Kentucky has a rich musical heritage that spans several centuries. This document provides an overview of key musical developments in Kentucky, focusing on historical roots, significant musical styles, and their cultural impact.

Historical Roots of Kentucky Music

Early Camp Meetings

- **July 1800**: The first camp meeting, a significant religious and social gathering, took place in Logan County, Kentucky.
- **1801 Cane Ridge Meeting**: This event in Bourbon County lasted six days, attracting ten to twenty thousand attendees.

Folk Music Collection

- **1916 Folk Music Collection**: Loraine Wyman and Howard Brockway collected and published Kentucky folk music in two folios. This work highlights the state's long-standing tradition of Appalachian folk music.

.....

Conclusion

Kentucky's musical heritage is rich and multifaceted, prominently featuring Appalachian folk and jug band music. While bluegrass music was not explicitly referenced in the initial historical context, it is widely recognized as a defining genre associated with the state. This supports the claim that Kentucky is known for bluegrass music.

References

- Historical records of camp meetings in Logan and Bourbon Counties.
- Publications by Loraine Wyman and Howard Brockway on folk music.
- Various sources acknowledging Kentucky's role in the development of bluegrass music.

Authoritative, Coarse-Grain, Unstructured (ACU)
In the early 19th century, Kentucky emerged as a significant site for religious gatherings, with the first camp meeting taking place in Logan County in July 1800. This event laid the groundwork for subsequent gatherings, such as the notable 1801 Cane Ridge meeting in Bourbon County, which was a monumental event lasting six days and attracting between ten to twenty thousand participants. These gatherings played a crucial role in sowing the seeds of Kentucky's rich musical tradition.

The early 20th century saw a deliberate effort to document and preserve the folk music of Kentucky. In 1916, Loraine Wyman and Howard Brockway embarked on a collection endeavor that culminated in the publication of Kentucky folk music in two comprehensive folios. This work highlighted the rich tapestry of Appalachian folk music, which forms the core of Kentucky's musical heritage, particularly in the eastern regions of the state. Louisville also played a pivotal role in the early development of jug band music, marking its significance in the state's diverse musical landscape.

.....

In this context, it is evident that bluegrass music, as a derivative of Appalachian folk traditions, is indeed a defining aspect of Kentucky's musical identity. The state is known for its contributions to this genre, reinforcing the claim that Kentucky is known for bluegrass music.

Conversational, Fine-Grain, Structured (CFS)

```
# Exploring Kentucky's Musical Heritage: Is Kentucky
Known for Bluegrass Music?

Hey there! If you've ever wondered about Kentucky's
musical roots, you're in for a treat. Were diving deep
 into the sounds that have shaped the Bluegrass State,
 and we'll see if Kentucky truly deserves its claim to
 fame in bluegrass music. So grab a comfy seat and
lets chat!

## Early Musical Gatherings

### Camp Meetings in the 1800s

Lets start way back in the early 1800s. Did you know
the very first camp meeting was held in Logan County
in July 1800? These gatherings were all about
community and music, setting the stage for big musical
 events. And then came 1801, when the Cane Ridge
meeting in Bourbon County drew a crowd of ten to
twenty thousand people for a six-day musical and
spiritual extravaganza! Imagine the energy and the
tunes floating through the air.

......

## The Verdict: Is Kentucky Known for Bluegrass Music?

So, does Kentucky deserve to be known for bluegrass
music? Absolutely! While its musical roots are diverse
 and rich, bluegrass has a special place in the heart
of Kentucky's musical identity. Its no wonder the
state is associated with those lively banjo picks and
harmonious vocals.

And there you have it! From camp meetings to jug bands
 to Bill Monroe, Kentucky has been a musical melting
pot leading straight to the vibrant bluegrass scene we
 love today. Thanks for exploring this musical journey
 with me!
```

Conversational, Fine-Grain, Unstructured (CFU)

```
Wow, Kentucky has such a rich musical heritage! Lets
chat about it. So, picture this: the early 1800s roll
around, and folks in Kentucky are already getting
together for these lively camp meetings. The first one
 was in July 1800 in Logan County. Things got even
bigger in 1801 at Cane Ridge in Bourbon County, where
a massive meeting brought in anywhere from ten to
twenty thousand people! Imagine all those folks
gathered together, possibly sharing music and stories.

Fast forward a bit to 1916it was a big year for music
collecting. Loraine Wyman and Howard Brockway were
trekking around Kentucky, gathering up all this
incredible folk music. They ended up publishing it in
two folios, capturing the essence of Kentuckys musical
 soul during that time.

......

So, after all this, its clear Kentuckys music is a
tapestrywoven with strains of folk, bluegrass, and jug
 band rhythms. Each style influencing the next, making
 Kentucky a cornerstone of American music history.
Bluegrass, with its warm, lively tones, is just one of
 the pearls in Kentucky's musical crown.
```

Conversational, Coarse-Grain, Structured (CCS)

```
# Kentucky's Musical Roots: A Quick Chat

Hey there! Let's dive into the vibrant world of
Kentucky's music scene and see if the Bluegrass State
really lives up to its name.

## Early Musical Gatherings

### Camp Meetings
- **1800**: The first ever camp meeting went down in
Logan County. It kicked off a long tradition of
gathering for music and community vibes.

- **1801**: In Cane Ridge, Bourbon County, a massive
meeting pulled in up to 20,000 folks for a six-day
musical extravaganza. Talk about a festival!

......

## Verdict on the Claim

### Statement: Kentucky is known for bluegrass music.
- **Absolutely!** While the roots may lie in
Appalachian and folk styles, bluegrass has flourished
here, becoming a signature sound that Kentucky proudly
 claims.

So, there you have it. Kentucky is as bluegrass as you
 can get, thanks to its rich musical tapestry. Got
more questions about Kentucky's tuneful legacy? Just
ask!
```

Conversational, Coarse-Grain, Unstructured (CCU)

```
Hey there! So, let's chat about music in Kentucky. It'
s pretty interesting how this state has such a rich
musical history. Back in the early 1800s, the whole
camp meeting craze started up in Logan County. Imagine
 these massive gatherings with people getting together
 for a kind of spiritual revival. Then, in 1801, Cane
Ridge in Bourbon County got crazy popular. This
meeting went on for six days and drew in thousands of
folks. Hard to believe, right?

Fast forward to 1916, and youve got Loraine Wyman and
Howard Brockway roaming around Kentucky, collecting
folk music. They ended up publishing two folios of
these traditional tunes. Its like they were capturing
the soul of Kentucky's music scene. Louisville was
already making waves with its early jug band music by
then.

Now, when people think of Kentucky music, Appalachian
folk music is front and center, especially in the
eastern parts. It's all about that cultural vibe and
storytelling through music. And of course, bluegrass
music is what a lot of people associate with Kentucky.
 Its become iconic here, influenced heavily by that
folk tradition.

So, yeah, Kentucky's music scene is as diverse and
lively as it gets, but bluegrass? It's pretty much the
 heart and soul. Hope this paints the picture!
```

F.2 Examples of Different Types of Self-generated Documents

Dataset: TQA

Authoritative, Fine-Grain, Structured (AFS)

```
# The VS-300: An Overview and Historical Context

## Introduction

The VS-300, formally designated as the Sikorsky VS
-300, is recognized as a seminal helicopter design.
This document provides an authoritative and detailed
exploration of the VS-300's nature, its historical
significance, and its role in the development of
vertical lift aircraft.

## Historical Context

### Development and Inception

- **Date**: The VS-300 was first flown in 1939,
marking a critical milestone in aviation history.
- **Designer**: Developed by Igor Sikorsky, a pioneer
in both fixed-wing and rotary-wing aviation.
- **Manufacturing Company**: Produced by the Sikorsky
Aircraft Corporation based in Bridgeport, Connecticut.

### Predecessors and Influences

The VS-300 represented a leap forward in helicopter
technology, following earlier attempts at vertical
lift vehicles such as autogiros and experimental
helicopters. Sikorsky's extensive experience in
designing fixed-wing aircraft, coupled with his
innovative thinking, propelled the development of the
VS-300.

......

## Conclusion

The VS-300, as a single-rotor helicopter with a tail
rotor, stands as a pioneering achievement in aerospace
 engineering. Developed by Igor Sikorsky and
introduced in 1940, the VS-300 not only demonstrated
the viability of vertical lift flight but also set the
 stage for advancements in helicopter design and
utility in both commercial and military sectors. Its
legacy continues to influence the vertical lift
aircraft industry, solidifying its place in aviation
history.
```

Authoritative, Fine-Grain, Unstructured (AFU)

```
The VS-300 is a helicopter, specifically a prototype
that marked a pivotal moment in the evolution of
rotorcraft technology during the early 1940s.
Developed by Sikorsky Aircraft Corporation, this
innovative aircraft would serve as a foundation for
the future of vertical take-off and landing (VTOL)
vehicles. The VS-300, with its distinctive single main
 rotor and rear-mounted tail rotor, represents one of
the earliest successful models of a practical
helicopter, which is credited with revolutionizing
aviation and transforming both military and civilian
sectors.

Igor Sikorsky, the visionary behind this
groundbreaking machine, sought to develop a vertical-
lift aircraft capable of sustained flight. The design
of the VS-300 incorporated several innovative features
 that would become standard in subsequent helicopter
designs. Among these features were the counter-
rotating rotors to counteract the torque effect and
the use of a tail rotor for directional control. The
aircraft utilized a 75-horsepower engine, which was
relatively modest by today's standards, but sufficient
 to achieve the technical milestones it was designed
to overcome.

......

In retrospect, the VS-300s significance lies not only
in its role as an early exemplar of helicopter
technology but also in its broader implications for
the field of aviation. It represents the realization
of Sikorsky's vision for practical vertical lift
aircraft, a vision that has since transformed
countless aspects of modern life, from emergency
services to aviation itself. The VS-300's legacy is
evident in the thousands of helicopters that have
followed in its wake, each contributing to the ongoing
 advancement of aviation technology.
```

Authoritative, Coarse-Grain, Structured (ACS)

```
# Overview of the VS-300

## Introduction

The VS-300 was a significant aircraft in the
development of vertical takeoff and landing (VTOL)
technology. It was an experimental helicopter designed
 and developed by Sikorsky Aircraft Corporation in the
 late 1930s.

## Key Characteristics

- **Type**: Experimental Helicopter
- **Manufacturer**: Sikorsky Aircraft Corporation
- **Period**: Late 1930s
- **Purpose**: To explore the feasibility of VTOL
aircraft

## Historical Context

- **Innovation**: The VS-300 was one of the earliest
successful helicopters, marking a critical advancement
 in aviation history.
- **Design**: It paved the way for future helicopter
designs and VTOL technologies.
- **Impact**: The VS-300's successful flight tests
demonstrated the viability of controlled helicopter
flight, influencing subsequent aviation developments.

## Conclusion

The VS-300 holds a pivotal place in the history of
aviation as an experimental helicopter that advanced
the technology of vertical takeoff and landing
capabilities. Its contributions were instrumental in
the evolution of modern helicopter designs.
```

Authoritative, Coarse-Grain, Unstructured (ACU)

```
The VS-300 is a notable aircraft that stands out in
the history of aviation as a significant experimental
helicopter. Developed in the 1930s, the VS-300 was a
key step in advancing vertical flight technology. Its
design and testing provided crucial insights into the
mechanics of helicopter flight, contributing to the
broader understanding of rotor dynamics.

The creation of the VS-300 was overseen by Igor
Sikorsky, a visionary engineer who played a pivotal
role in the early development of helicopters. His work
 on the VS-300 helped to establish a foundation for
subsequent advancements in rotary-wing aircraft.
Through rigorous testing and iterative improvements,
the VS-300 demonstrated the feasibility of controlled
vertical takeoff and landing, which would later become
 a hallmark feature of modern helicopter designs.

Despite its historical significance, the VS-300
remained a prototype and was not produced for
widespread use. Nonetheless, its influence on the
trajectory of helicopter development is undeniable,
marking a critical point in the evolution of aerial
transportation technology.
```

Conversational, Fine-Grain, Structured (CFS)

```
# The VS-300: A Conversational Dive into History

Hey there! Let's talk about a pretty cool piece of
aviation history - the VS-300. If you've ever wondered
 what this mysterious-sounding machine was all about,
buckle up because we're about to take a deep dive into
 its story.

## The Basics: What Was the VS-300?

The VS-300, or the "Versatile 300," was the first
successful helicopter designed by Igor Sikorsky. It
might not have the fanfare of airplanes like the
Wright Flyer or the Spirit of St. Louis, but its
impact on aviation history is no less significant.

### A Crash Course in Helicopters

Before we get too deep, let's quickly cover what a
helicopter is. Unlike airplanes that rely on wings for
 lift, helicopters use spinning blades (rotors) to
achieve vertical takeoff and landing. This vertical
lift capability is what sets helicopters apart and
makes them incredibly versatile.

## Who Was Igor Sikorsky?

To really understand the VS-300, we need to dive into
the mind behind it: Igor Sikorsky. Born in Ukraine,
Sikorsky immigrated to the United States in the early
1900s and quickly made a name for himself in the
aviation industry. He was intrigued by the concept of
vertical lift aircraft from a young age and dedicated
his life to making helicopters a reality.

### The Early Years

Sikorskys fascination with helicopters started when he
 was a child in tsarist Russia. He first saw a model
helicopter design by Gustave Whitehead and was
immediately captivated. Fast forward to the 1930s, and
 Sikorsky was ready to turn his lifelong obsession
into a tangible creation.

......

## Wrapping It Up

So, there you have it  the VS-300 was Igor Sikorskys
groundbreaking helicopter that took the world by storm
. The first of its kind to prove the concept of
vertical lift flight, it paved the way for the modern
helicopters we see today. Next time you look up at a
helicopter flying overhead, you can thank the VS-300
and the vision of Igor Sikorsky for making it all
possible.
```

Conversational, Fine-Grain, Unstructured (CFU)

```
Hey there! So, let's chat about the VS-300 for a bit.
Its actually pretty cool what this thing is all about.
 The VS-300 was a type of helicopter, and it really
marked the beginning of something really revolutionary
 in the world of aviation.

Now, imagine you're in the late 1930s. Airplanes are
starting to become a common mode of transportation,
but vertical flight was still a grand mystery. That's
where the VS-300 comes in. It was designed by Igor
Sikorsky, a name you might recognize from the world of
 aviationhes a pretty big deal. Sikorsky was all about
 pushing the boundaries of what was possible, and the
VS-300 was one of his groundbreaking projects.

The VS-300 was the first practical helicopter to fly
in the United States. It was a prototype and Sikorskys
 second helicopter design. The first one, the VS-300's
 predecessor, was the V-173. But the VS-300 was the
one that really caught people's attention.

So, what made the VS-300 so special? Well, it was the
first helicopter to use a single main rotor and a
separate tail rotor for balance and control. This
design became the standard for most modern helicopters
. Before this, there had been attempts at helicopters,
 but they were generally impractical and didn't really
 work well.

......

So, when you hear about the VS-300, remember that it's
 the first practical helicopter to fly in the U.S.,
and it really laid the foundation for modern vertical
flight technology. It's a pretty crucial piece of
aviation history.
```

Conversational, Coarse-Grain, Structured (CCS)

```
# The VS-300: A Quick Overview

Hey there! Let's dive into a brief look at what the VS
-300 is all about. If you're curious about its place
in history and what it represents, you've come to the
right place!

## What's a VS-300?

The VS-300 was a type of experimental helicopter. It
was a leap forward in aviation technology and served
as a crucial stepping stone for innovation in
helicopter design.

### Key Points
- **Type:** Experimental helicopter
- **Innovation:** Early foray into vertical takeoff
and landing (VTOL) capabilities
- **Significance:** Played a major role in the
development of the modern helicopter

## History in a Nutshell

The development of the VS-300 was tied to a
significant figure in aviation history: Igor Sikorsky.
 He was the mastermind behind this innovative aircraft
.

### Main Takeaways
- **Brainchild:** Igor Sikorsky
- **Purpose:** To explore the potential of VTOL
aircraft
- **Outcome:** Led to the creation of more advanced
helicopter models

## Legacy and Impact

The VS-300 might be a bit of a vintage piece today,
but it laid the groundwork for what we now know as the
 modern helicopter. So next time you see one of those
buzzing around in the sky, remember the VS-300's role
in making that possible!

### Big Picture
- **Influence:** Shaped future helicopter designs
- **Legacy:** Pioneer in VTOL technology

## Wrap Up

So there you go! The VS-300 was a significant tool in
the quest to conquer the skies from every angle. It's
just one piece of the puzzle in the incredible story
of human flight. Hope this gave you a quick glimpse
into the world of vintage aviation wonders!
```

Conversational, Coarse-Grain, Unstructured (CCU)

```
Hey there! So, you've probably heard of helicopters,
right? Well, the VS-300 was actually one of the very
first models of a whole new kind of aircraft. It's a
```

helicopter, folks! Crazy to think about, but back in the day, when folks were still figuring out flight, this little guy made one of the first successful helicopter flights.

Igor Sikorsky was the mastermind behind it. He basically took the idea of a helicopter and turned it from a dream into reality with the VS-300. It was all about proving that vertical flight was possible and that a helicopter could hover, take off vertically, and land vertically too.

The VS-300 might have been a bit clunky and not exactly what we think of helicopters today, but it was the foundation for everything that came after. From those early experiments with the VS-300, we got some of the most advanced choppers you see today, both in military and civilian roles.

So, next time you see a helicopter flying around, you can thank the VS-300 for showing the way. Its like the granddaddy of all helicopters, really!

Dataset: FEVER

Authoritative, Fine-Grain, Structured (AFS)

Analysis of the Claim: "Fox 2000 Pictures Released the Film Soul Food"

Introduction

This document aims to provide a comprehensive analysis to support or refute the claim that Fox 2000 Pictures released the film "Soul Food." The analysis will cover the production details, distribution, and historical context of the film to ensure a thorough assessment of the claim.

Historical Context and Background

"Soul Food": Overview

"Soul Food" is a 1997 American family drama film that explores the dynamics of an African American family centered around their matriarch, Memaw (played by Oprah Winfrey). The film focuses on the family's relationships and conflicts aftermath the death of Memaw, as her children and grandchildren struggle to overcome personal challenges and familial tensions.

Production and Distribution Company Details

- **Producer**: The film was produced by Oprah Winfrey's Harpo Films in association with Mandalay Pictures.
- **Distribution**: The film was distributed by Columbia Pictures, a major studio under the Sony Pictures Entertainment banner.

Film Production

Production Company

- **Harpo Films**: This production company was founded by Oprah Winfrey in 1987 to produce both television and film content. Harpo Films had a significant role in the development and production of "Soul Food."
- **Mandalay Pictures**: This company, headed by Lawrence Bender, produced and co-financed the film with Harpo Films.

Release and Reception

- **Release Date**: The film was released on January 17, 1997.
- **Box Office and Critical Reception**: "Soul Food" was a moderate box office success and received generally positive reviews from critics.

.....

Conclusion

This document has thoroughly analyzed the production, distribution, and historical context of the film "Soul Food." Through a detailed examination, it has been determined that Fox 2000 Pictures was not involved in the production or release of this film. The accurate credits for "Soul Food" belong to Harpo Films and Columbia Pictures.

Authoritative, Fine-Grain, Unstructured (AFU)

The claim that Fox 2000 Pictures released the film "Soul Food" requires detailed consideration of the film's production and distribution history. "Soul Food," a critically acclaimed drama released in 1997, centers around the dynamics of a close-knit African American family, exploring themes of love, loss, and the power of food in familial bonds. However, the studio responsible for the production and release of "Soul Food" was not Fox 2000 Pictures, a division known for producing films like "Little Women" (1994), "Intolerable Cruelty" (2003), and the "Pitch Perfect" series (2012-2017).

The actual production and release of "Soul Food" were overseen by Metro-Goldwyn-Mayer (MGM) studio. MGM, a pioneer in the film industry with a rich history dating back to 1924, has been involved in the production of numerous classic and contemporary films. In the case of "Soul Food," MGM not only produced the film but also handled its distribution, which was notably successful both critically and commercially. The film's reception highlighted the studio's capability in tackling culturally significant narratives, and it garnered praise for its authentic storytelling and strong performances.

To further contextualize the claim, it is important to examine the activities and scope of Fox 2000 Pictures during the late 1990s and early 2000s. Founded in 1993 as a division of 20th Century Fox, Fox 2000 Pictures focused on releasing limited-release independent films and select mainstream pictures. While the studio has been involved in a range of notable projects, "Soul Food" does not feature among them. Thus, any association of Fox 2000 Pictures with the production or release of "Soul Food" is inaccurate.

In summary, the claim that Fox 2000 Pictures released "Soul Food" is not supported by historical records of the film's production and distribution. MGM was the studio responsible for bringing "Soul Food" to audiences in 1997. This differentiates the actual production entity from the one inaccurately alleged in the initial claim, thus providing clear evidence to refute the statement.

Authoritative, Coarse-Grain, Structured (ACS)

Analysis of the Film "Soul Food" and Its Production Company

Introduction

This document provides a structured analysis to determine the claim regarding the film "Soul Food" and its production company, Fox 2000 Pictures. The document aims to offer a coarse-grain overview while maintaining an authoritative and objective tone.

Film Overview

"Soul Food" is a 1997 film that centers on themes of family, food, and generational conflict within an African American family. The film is known for its exploration of the cultural significance of soul food in African American traditions.

Production Companies Involved
- **Cleopatra Pictures**: This is the primary production company credited with the film "Soul Food."
- **Fox 2000 Pictures**: This company is known for producing family-oriented and drama films. However, it does not appear as a production credit for the film in question.

Release Information
- **Release Date**: "Soul Food" was released in the United States on September 12, 1997.
- **Production Credits**: The film was primarily produced by Cleopatra Pictures, with collaboration from the films director, writer, and producer duo, Kasi Lemmons and Bill Duke.

Conclusion
Based on the release information and production credits, the claim that Fox 2000 Pictures released the film "Soul Food" is refuted. Rather, "Soul Food" was produced by Cleopatra Pictures, a different production company distinct from Fox 2000 Pictures.

This structured analysis indicates that while "Soul Food" is a notable film, its production and distribution were handled by Cleopatra Pictures, with no direct production involvement from Fox 2000 Pictures.

Authoritative, Coarse-Grain, Unstructured (ACU)
The claim that Fox 2000 Pictures released the film "Soul Food" does not align with the established facts surrounding the project's production and distribution. "Soul Food" is widely recognized as a motion picture that explores themes of family, tradition, and the culinary culture of African American communities. The film premiered in 1997 and garnered attention for its portrayal of a large Southern African American family, centered around traditional soul food meals.

The production and distribution details of "Soul Food" indicate involvement from several key players in the film industry, including Trimark Pictures and New Line Cinema. These entities were instrumental in bringing the film from pre-production stages through to its theatrical release, taking on roles typically associated with the creative and financial aspects of film production and distribution.

Contrary to the claim, Fox 2000 Pictures is not historically linked to "Soul Food." Fox 2000 Pictures, an entity under the wing of 20th Century Fox, has been involved in a different slate of projects, focusing on a variety of independent and mainstream films. Their involvement in significant films such as "The Holiday" and the "Pitch Perfect" franchise clearly distinguishes their catalog from the content and era of "Soul Food."

In summary, while "Soul Food" is a critically acclaimed film that resonated with audiences, the assertion about its release by Fox 2000 Pictures cannot be substantiated. The film's production and distribution were managed by other major studios, leading to a misalignment with the claim presented.

Conversational, Fine-Grain, Structured (CFS)
Exploring the Claim: Fox 2000 Pictures and Soul Food

Hey there! Today, were going to look into the filmmakers and production companies behind the classic film "Soul Food." The claim at hand is that Fox 2000 Pictures released the film. Lets break it down and see if we can support or refute this claim.

The Film "Soul Food": Background and Context

What is "Soul Food"?
"Soul Food" is a heartfelt 1997 film that delves into the lives of the Tates, a close-knit African American family in Chicago. It follows the familys struggle to cope with the loss of their matriarch, Veronica, whose cooking united the family. The film explores themes of family dynamics, grief, and reconciliation.

Production and Release

Production Companies
The film was produced by several companies, including:
- **Warner Bros. Pictures:** A big player in the film industry, known for producing and distributing many well-known films.
- **Native Sun Productions:** Founded by actors LL Cool J and Lisa Ray, who was also a producer on the film.
- **Source 7 Entertainment:** A production company founded by producer Johnny Odom.

Release Information
The original release and production were handled by Warner Bros. Pictures, not Fox 2000 Pictures. Warner Bros. is a separate entity from Fox 2000, which is a subsidiary of 20th Century Studios (formerly 20th Century Fox).

.....

Final Thoughts
Its important to accurately represent the history of a film, especially when it comes to the various entities involved in its creation. "Soul Food" is a significant work that was brought to life by Warner Bros. Pictures and its fellow producers. Understanding the roles of the various companies involved helps us appreciate the collaborative effort behind the scenes.

Thanks for joining this deep dive into the production details of "Soul Food"! If you have any further questions or would like to explore other aspects of the film industry, feel free to ask!

Conversational, Fine-Grain, Unstructured (CFU)
Hey there! Let's dive into the details around Fox 2000 Pictures and the film "Soul Food." So, first off, Fox 2000 Pictures is a subsidiary of 20th Century Fox, and it primarily focuses on producing drama and family films. They've been around since 1998, and they've certainly put out some notable movies.

Now, when it comes to "Soul Food," the film was actually released in 1997. It's a drama that focuses on family dynamics and the impact of food on relationships, especially in a Black family. The plot revolves around matriarch Tashi Jones and her daughters, who reconnect after a falling-out that occurred when they were deciding who would inherit the family house. During the course of the film, they deal with various issues, including the obesity and diabetes diagnosis of one of the daughters, Denise.

So, here's the kicker: Fox 2000 Pictures wasn't around when "Soul Food" was released. The film was produced by HBO Films and wasn't until after 1998 that Fox 2000 Pictures started to release films. This means that the claim suggesting Fox 2000 Pictures released "Soul Food" is not correct.

.....

In summary, while "Soul Food" is an important and beloved film, it wasn't released by Fox 2000 Pictures. It was actually brought to life by HBO Films and distributed by Gramercy Pictures, a different entity altogether. So, the claim doesn't hold up when you look at the details.

Conversational, Coarse-Grain, Structured (CCS)

```
# Unpacking the Claim: Fox 2000 & Soul Food

Hey there! Today, we're going to take a closer look at
 whether Fox 2000 Pictures was behind the film "Soul
Food." Well jog through what the studio is known for
and where "Soul Food" really originated. Lets dive in!

## Fox 2000 Pictures: A Snapshot

### Who Are They?
Fox 2000 Pictures is a production company under the
larger 20th Century Fox umbrella (now a part of Disney
). Theyre known for producing films that straddle the
line between art-house and mainstream.

### Hit Films
Some of their notable titles include:
- **"Billy Elliot"**: An inspiring story about a boy
determined to become a dancer.
- **"The Help"**: A drama based on the novel by
Kathryn Stockett.
- **"Little Women"**: Directed by Greta Gerwig and a
modern take on Louisa May Alcotts classic.

......

### Quick Recap
- **Fox 2000**: Known for art-house and mainstream
hits but not "Soul Food."
- **"Soul Food"**: A notable family drama made by New
Line Cinema and Circle of Trust Productions.

There you have it! The claim isnt quite right, but its
 interesting to understand the different avenues of
production in the film industry. Got any other movie
connections you're curious about? Just ask!
```

Conversational, Coarse-Grain, Unstructured (CCU)

```
Hey, so let's talk about this film, "Soul Food," and
see if we can sort out who actually released it. First
 off, "Soul Food" is known for its deep exploration of
 family dynamics and the strong African American
community ties, all set amidst a mouthwatering
backdrop of delicious Southern cuisine. It came out in
 1997 and received quite a bit of attention for its
cultural relevance and the impactful performances by
its cast, including an unforgettable turn by Theresa
Randle as Mamie Simmons, the matriarch of the Johnson
family.

Now, while Fox 2000 might have some great films under
its belt, "Soul Food" isn't one of them. In fact, the
movie was produced by New Line Cinema. New Line Cinema
 had a way of picking up films that touched on deeply
personal and cultural themes, and "Soul Food"
definitely fits that description. They had a knack for
 stories connecting with audiences on a more
heartwarming level, and this film is all about love,
food, and the struggles of a family.

So in short, while Fox 2000 Pictures has their own
impressive list of films, "Soul Food" isn't one of
them. New Line Cinema is credited with this heart-
touching tale, and it's a movie that left its mark on
many viewers. Hope that clears things up!
```