Ancient Wisdom, Modern Tools: Exploring Retrieval-Augmented LLMs for Ancient Indian Philosophy

Priyanka Mandikal
Department of Computer Science, UT Austin
mandikal@utexas.edu

Abstract

LLMs have revolutionized the landscape of information retrieval and knowledge dissemination. However, their application in specialized areas is often hindered by limitations such as factual inaccuracies and hallucinations, especially in long-tail knowledge distributions. In this work, we explore the potential of retrieval-augmented generation (RAG) models in performing long-form question answering (LFQA) on a specially curated niche and custom knowledge domain. We present VedantaNY-10M, a dataset curated from extensive public discourses on the ancient Indian philosophy of Advaita Vedanta. We develop and benchmark a RAG model against a standard, non-RAG LLM, focusing on transcription, retrieval, and generation performance. A human evaluation involving computational linguists and domain experts, shows that the RAG model significantly outperforms the standard model in producing factual, comprehensive responses having fewer hallucinations. In addition, we find that a keyword-based hybrid retriever that focuses on unique low-frequency words further improves results. Our study provides insights into meaningfully integrating modern large language models with ancient knowledge systems.

1 Introduction

Generic LLMs have proven to be highly effective for broad knowledge domains. However, they often encounter challenges in niche and less popular areas, suffering from issues such as factual inaccuracies and hallucinations for long-tail knowledge distributions (Kandpal et al., 2023; Mallen et al., 2023). Moreover, the inability to verify responses against authentic sources is particularly problematic in these long-tail domains, where LLMs can generate highly inaccurate answers with unwarranted confidence (Kandpal et al., 2023; Menick et al., 2022).

In response to these limitations, there has been a growing interest in retrieval-augmented generation (RAG) models (Guu et al., 2020; Karpukhin et al., 2020; Lewis et al., 2020b; Izacard et al., 2022; Ram et al., 2023). These models, which integrate external datastores to retrieve relevant knowledge and incorporate it into LLMs, have demonstrated higher factual accuracy and reduced hallucinations compared to conventional LLMs (Shuster et al., 2021; Borgeaud et al., 2022; Menick et al., 2022). Additionally, updating these external datastores with new information is more efficient and cost-effective than retraining LLMs.

Recent studies on the societal impact of LLMs (Malhotra, 2021; Yiu et al., 2023) have highlighted the increasing significance of LLMs as cultural technologies akin to libraries for search and retrieval. Analogous to earlier technologies like writing, print, libraries and internet search, the power of LLMs can be harnessed meaningfully to preserve and disseminate human knowledge (Sommerschield et al., 2023). In this vein, we argue that RAG models show immense potential for supplementing study in diverse knowledge domains. Hence, there is a growing need to examine the applications of RAG models for unconventional, custom knowledge domains that are often niche and scarcely represented in pre-training data. The capability of RAG models to provide verified, authentic sources when answering questions is particularly advantageous for end-users.

In this work, we develop and evaluate a RAG-based language model specialized in the ancient Indian philosophy of Advaita Vedanta (Upanishads, >3000 B.C.E.; Bhagavad Gita, 3000 B.C.E.; Shankaracharya, 450 B.C.E.). To ensure that the LLM has previously not been exposed to the source material, we construct VedantaNY-10M, a custom philosophy datastore comprising transcripts of over 750 hours of public discourses on YouTube from Vedanta Society of New York. We evaluate standard non-RAG and RAG models on this domain, and find that RAG models perform significantly

Proceedings of the 1st Workshop on Machine Learning for Ancient Languages (ML4AL 2024), pages 224–250
August 15, 2024

better. However, they suffer from a number of issues, including irrelevant retrievals, sub-optimal retrieval passage length, retrieval-induced hallucinations, and reliance on outside knowledge. In early attempts to mitigate some of these issues, we find that traditional sparse retrievers have a unique advantage over dense retrievers in niche domains having specific terminology—Sanskrit terms in our case. Hence, we propose a keyword-based hybrid retriever that effectively combines sparse and dense embeddings to upsample low-frequency or domain-specific terms. In addition, a simple keyword-based retrieval refinement serves to shorten or lengthen retrievals to further refine context.

We conduct an extensive evaluation comprising both automatic metrics and human evaluation with 5 computational linguists and 3 domain experts, assessing the models along three dimensions: transcription, retrieval, and generation. Our findings are twofold. First, the standard RAG model significantly outperforms the generic non-RAG model along all axes, offering more factual, comprehensive, and specific responses while minimizing hallucinations. User preference for the RAG model over the generic counterpart is evident, with a preference rate of 81%. Second, the keyword-based hybrid RAG model further outperforms the standard deep-embedding based RAG model in both automatic and human evaluation metrics. Our study also includes detailed long-form responses from the evaluators, with domain experts specifically indicating the likelihood of using such LLMs to supplement their day-to-day study. Our work offers a step toward building and evaluating real-world RAG models for niche and esoteric ancient knowledge domains, highlighting the opportunities and challenges arising thereof.

2 Related Work

Language models for ancient texts Sommerschield et al. (2023) recently conducted a thorough survey of machine learning techniques applied to the study and restoration of ancient texts. Spanning digitization (Narang et al., 2019; Moustafa et al., 2022), restoration, (Assael et al., 2022), attribution (Bogacz and Mara, 2020; Paparigopoulou et al., 2022) and representation learning (Bamman and Burns, 2020), a wide range of use cases have benefitted from the application of machine learning techniques to study ancient texts. Recently, Lugli et al. (2022) released a digital corpus of romanized Buddhist Sanskrit texts, training and evaluating embedding models such as BERT and GPT-2 on them. However, the use of LLMs as a question-answering tool to enhance understanding of ancient esoteric knowledge systems has not yet been systematically studied. To the best of our knowledge, ours is the first work that studies the effects of RAG-based models in the niche knowledge domain of ancient Indian philosophy.

Retrieval-Augmented LMs. In current LLM research, retrieval augmented generation models (RAGs) are gaining popularity (Izacard et al., 2022; Ram et al., 2023; Khandelwal et al., 2020; Borgeaud et al., 2022; Menick et al., 2022). A key area of development in RAGs has been their architecture. Early approaches involved finetuning the language model on open-domain question-answering before deployment. MLM approaches such as REALM (Guu et al., 2020) introduced a two-stage process combining retrieval and reading, while DPR (Karpukhin et al., 2020) focused on pipeline training for question answering. RAG (Lewis et al., 2020b) used a generative approach with no explicit language modeling. ATLAS (Izacard et al., 2022) combined RAG with retrieval-based pre-training, employing an encoder-decoder architecture. Very recently, in-context RALM (Ram et al., 2023) showed that retrieved passages can be used to augment the input to the LLM in-context without any fine-tuning like prior work. In this work, we adopt the in-context retrieval augmented methodology similar to (Ram et al., 2023), where neither the retriever nor the generator is fine-tuned. This also enables us to use any combination of retrieval and generation models that best suits our application.

Applications of RAGs. The applications of RAGs are diverse and evolving. ATLAS (Izacard et al., 2022) and GopherCite (Menick et al., 2022) have shown how fine-tuning and reinforcement learning from human feedback can enhance RAGs' ability to generate verifiable answers from reliable sources. GopherCite notably focused on producing answers with verifiable quotes without modifying the retrieval model. Prompting techniques have also seen innovation. kNNPrompt (Shi et al., 2022) extended kNN-LM for zero or few-shot classification tasks, and retrieval in-context approaches (Ram et al., 2023; Shi et al., 2023) have proven effective in utilizing retrieval at the input stage. Retrieval-LMs have been shown to be particularly valuable for handling long-tail or less frequent entities (Kandpal et al., 2023; Mallen et al., 2023), updating knowledge (Izacard et al., 2022), improving parameter efficiency (Izacard et al., 2022; Mallen et al., 2023), and enhancing verifiability (Bohnet et al., 2022), making them increasingly relevant in a wide range of applications. In our work, we examine the application of RAGs for long-tail knowledge, conducting an extensive study on a niche knowledge domain of ancient Indian philosophy.

Evaluation of LFQA The field of long-form question answering (LFQA) is an emerging area of active research (Krishna et al., 2021; Nakano et al., 2021; Xu et al., 2023). Recently, Xu et al. (2023) conducted a thorough examination of various LFQA metrics, encompassing both human and automatic evaluation methods, and found that existing automatic metrics don't always align with human preferences. Based on their suggestions, we place special emphasis on conducting an extensive human evaluation utilizing the expertise of experienced computational linguists and domain experts.

3 The VedantaNY-10M Dataset

We first describe our niche domain dataset creation process. The custom dataset for our study needs to satisfy the following requirements: (**1**) **Niche:** Must be a specialized niche knowledge domain within the LLM's long-tail distribution. (**2**) **Novel**: The LLM must not have previously encountered the source material. (**3**) **Authentic:** The dataset should be authentic and representative of the knowledge domain. (**4**) **Domain experts:** should be available to evaluate the model's effectiveness and utility.

Knowledge domain. To satisfy the first requirement, we choose our domain to be the niche knowledge system of Advaita Vedanta, a 2500-year-old Indian school of philosophy (Shankaracharya, 450 B.C.E.) based on the Upanishads (>3000 B.C.E.), Bhagavad Gita (3000 B.C.E.) and Brahmasutras (3000 B.C.E.)[1]. It is a contemplative knowledge tradition that employs a host of diverse tools and techniques including analytical reasoning, logic, linguistic paradoxes, metaphors and analogies to enable the seeker to enquire into their real nature. Although a niche domain, this knowledge system has been continuously studied and rigorously developed over millenia, offering a rich and structured niche for the purposes of our study. Being a living tradition, it offers the additional advantage of providing experienced domain experts to evaluate the language models in this work.

Composition of the dataset. Considering the outlined criteria, we introduce VedantaNY-10M, a curated philosophy dataset of public discourses. To maintain authenticity while ensuring that the LLM hasn't previously been exposed to the source material, we curate our dataset from a collection of YouTube videos on Advaita Vedanta, sourced from the Vedanta Society of New York. It contains 10M tokens and encompasses over 750 hours of philosophical discourses by Swami Sarvapriyananda, a learned monk of the Ramakrishna Order. These discourses provide a rich and comprehensive exposition of the principles of Advaita Vedanta, making them an invaluable resource for our research.

Languages and scripts. The dataset primarily features content in English, accounting for approximately 97% of the total material. Sanskrit, the classical language of Indian philosophical literature, constitutes around 3% of the dataset. The Sanskrit terms are transliterated into the Roman script. To accommodate the linguistic diversity and the specific needs of the study, the dataset includes words in both English and Sanskrit, without substituting the Sanskrit terms with any English translations. Translating ancient Sanskrit technical terms having considerably nuanced definitions into English is a non-trivial problem (Malhotra and Babaji, 2020). Hence, our dual-language approach ensures that the Sanskrit terms and concepts are accurately represented and accessible, thereby enhancing the authenticity of our research material. For a sample of the Sanskrit terms present in the corpus, please refer to Appendix Tab. 2.

4 In-context RAG for niche domains

We now discuss the methodology adopted to build an in-context retrieval augmented chatbot from the custom dataset described above.

We first define a generic chatbot C_g that does not use retrieval as follows: $C_g : q \rightarrow a_g$ where q

[1]Currently there exists no consensus on accurately dating these ancient scriptures. The Upanishads (which are a part of the Vedas) have been passed on orally for millennia and are traditionally not given a historic date. However, they seem to have been compiled and systematically organized sometime around 3000 B.C.E. by Vyasa. Likewise, the time period of Adi Shankaracharya also varies and he is usually placed between 450 B.C.E to 700 C.E.

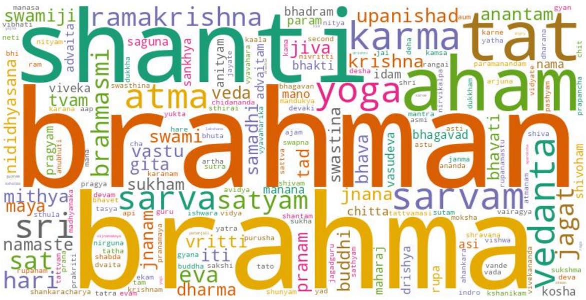

Figure 1: **Sanskrit terms in VedantaNY-10M.** Frequently occurring Sanskrit terms in the corpus.

is the user query and a_g is the answer generated by the chatbot. Now, let D_t represent the textual data corpus from our knowledge domain and R be the retriever. Our goal is to build a retrieval-augmented generation chatbot $C_r : q \times R(D_t, q) \rightarrow a_r$ that will generate answer a_r for the query by retrieving relevant context from D_t using R. An overview of our approach is illustrated in Fig. 2. We first build D_t from 765 hours of public discourses on Advaita Vedanta introduced in Sec. 3. When deployed, the system processes q by first using retriever R to identify the top-k most relevant passages P from D_t using a similarity metric. Subsequently, a large language model (LLM) is prompted with both the query and the retrieved passages in-context, following Ram et al. (2023), to generate a contextually relevant response.

We now describe each of the components in detail. We follow a four-stage process as follows:

Transcription. We first need to create a dense textual corpus targeted at our niche domain. Since our dataset consists of YouTube videos, we first employ a transcription model to transcribe the audio into text. Our video corpus D_v consists of 612 videos totaling 765 hours of content, with an average length of 1.25 hours per video. We extract audio content from D_v and transcribe it using OpenAI's Whisper large-v2 model (Radford et al., 2023). This step converts the spoken discourses into a transcribed textual corpus D_t consisting of 10M tokens in total. Since Whisper is a multilingual model, it has the capacity to support the dual-language nature of our dataset. We evaluate the transcription quality of Whisper in Sec. C.1.

Datastore creation. The transcribed text in D_t is then segmented into shorter chunks called passages P, consisting of 1500 characters each. These chunks are then processed by a deep embedder to produce deep embedding vectors z_{dense}. These embedded chunks are stored in a vector database D_z. Ultimately, we store approximately 25,000 passage embeddings $z \in D_z$, each representing a discrete chunk of the philosophical discourse in D_t.

Retrieval. To perform retrieval-augmented generation, we first need to build a retrieval system $R : D_z \times q \rightarrow P$ that retrieves contextually relevant textual passages P from D_t given D_z and q. The retriever performs the following operation to retrieve relevant passages: $P = D_t[argTop\text{-}k_{z \in D_z} sim(q, z)]$, where we use cosine similarity as the similarity metric. Standard RAG models employ state-of-the-art deep embedders to encode documents and retrieve them during inference. However, these semantic embeddings can struggle to disambiguate between specific niche terminology in custom domains (Mandikal and Mooney, 2024). This can be particularly problematic in datasets having long-tail distributions such as ours. In addition, retrieved fixed-length passages are suboptimal. Short incomplete contexts can be particularly damaging for LFQA, while longer contexts can contain unnecessary information that can confuse the generation model. To mitigate these two issues, we experiment with two key changes: (1) a keyword-based hybrid retriever to focus on unique low-frequency words, and (2) a context-refiner to meaningfully shorten or expand retrieved context.

1. **Keyword-based retrieval.** To emphasize the importance of key terminology, we first employ keyword extraction and named-entity recognition techniques on the query q to extract important keywords κ. During retrieval, we advocate for a hybrid model combining both deep embeddings as well as sparse vector space embeddings. We encode the full query in the deep embedder and assign a higher importance to keyphrases in the sparse embedder. The idea is to have the sparse model retrieve domain-specific specialized terms that might otherwise be missed by the deep model. Our hybrid model uses a simple weighted combination of the query-document similarities in the sparse and dense embedding spaces. Specifically, we score a document D for query q and keywords κ using the ranking function:

$$S_{hybrid}(D, q) = \lambda \, Sim\big(z_d(D), z_d(q)\big) + (1 - \lambda) \, Sim\big(z_s(D), z_s(\kappa)\big) \tag{1}$$

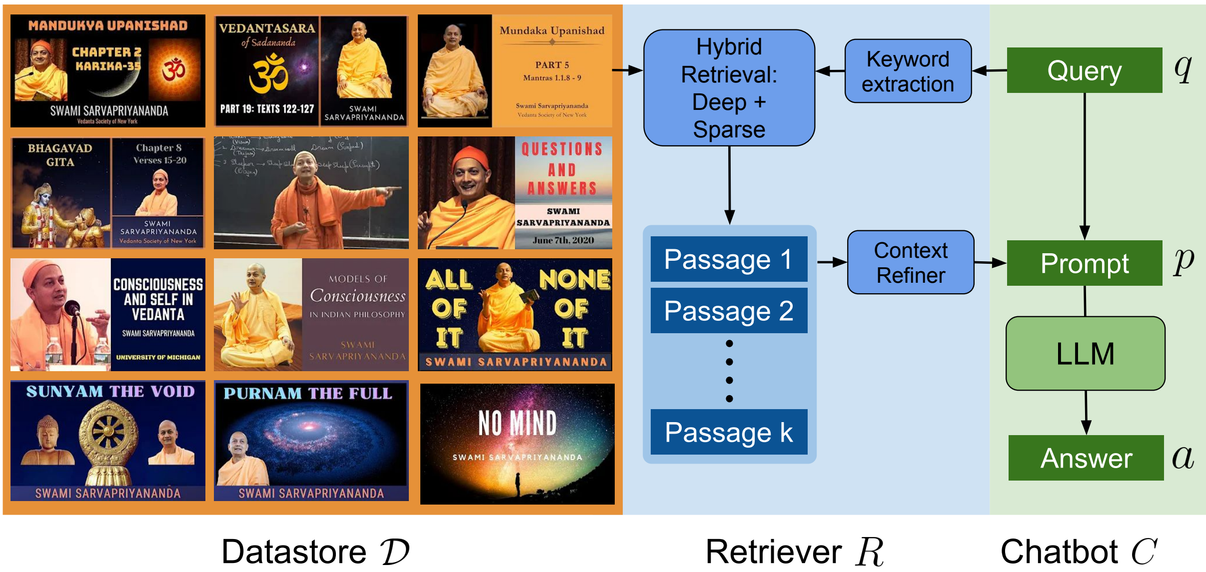

Figure 2: **Overview of the RAG model.** We build VedantaNY-10M, a datastore from 750+ hours of public discourses on the ancient Indian philosophy of Advaita Vedanta and build a Retrieval-Augmented Generation (RAG) chatbot on this knowledge domain. At deployment, given a query q, the retriever R first retrieves the top-k most relevant passages P from the datastore using a hybrid keyword-based retriever. It then refines this retrieved context using a keyword-based context reshaper to shorten or lengthen the passage. Finally, an LLM is invoked by prompting it with the query and the retrieved passages in-context. An extensive evaluation is conducted to evaluate this model with the help of computational linguists and domain experts to assess its real-world utility and identify challenges.

Where z_d and z_s denotes the dense and sparse embedding functions and Sim is cosine similarity measuring the angle between such vector embedddings. In our experiments, we set $\lambda = 0.2$. Amongst the top-n retrieved passages, we choose k passages containing the maximum number of unique keywords.

2. **Keyword-based context refinement.** Furthermore, we refine our retrieved passages by leveraging the extracted keywords using a heuristic-based refinement operation to produce $P' = Ref(P, \kappa)$. For extension, we expand the selected passage to include one preceding and one succeeding passage, and find the first and last occurrence of the extracted keywords. Next, we trim the expanded context from the first occurrence to the last. This can either expand or shorten the original passage depending on the placement of keywords. This ensures that retrieved context contains relevant information for the generation model.

Generation. For answer generation, we construct prompt p from the query q and the retrieved passages $(P'_1, P'_2, ..., P'_k) \in P$ in context. Finally, we invoke the chatbot C_r to synthesize an answer a_r from the constructed prompt. For an example of the constructed RAG bot prompt, please refer to Fig. 4. This four-stage process produces a retrieval-augmented chatbot that can generate contextually relevant responses for queries in our niche domain.

Implementation Details. For embedding and generation, we experiment with both closed and open source language models. For RAG vs non-RAG comparison, we use OpenAI's text-embedding-ada-002 model (Brown et al., 2020) as the embedder and GPT-4-turbo (OpenAI, 2023) as the LLM for both C_r and C_g. For comparing RAG model variants, we use the open source nomic-embed-text-v1 (Nussbaum et al., 2024) as our deep embedder and Mixtral-8x7B-Instruct-v0.1 (Jiang et al., 2024) as our generation model. For keyword extraction, we use an ensemble of different models including OpenKP (Xiong et al., 2019), KeyBERT (Grootendorst, 2020) and SpanMarker (Aarsen, 2020). We experimented with using language models such as ChatGPT for keyword extraction, but the results were very poor as also corroborated in Song et al. (2024). For further implementation details of the eval metrics, see Appendix Sec. A. The VedantaNY-10M dataset, code and evaluation is publicly available at `https://github.com/priyankamandikal/vedantany-10m`.

5 Evaluation

We now evaluate the model along two axes: automatic evaluation metrics and a human evaluation survey. To ensure a broad and comprehensive evaluation, we categorize the questions into five distinct types, each designed to test different aspects of the model's capabilities:

1. **Anecdotal:** Generate responses based on stories and anecdotes narrated by the speaker in the discourses.
2. **Comparative**: Analyze and compare different concepts, philosophies, or texts. This category tests the model's analytical skills and its ability to draw parallels and distinctions.
3. **Reasoning** Require logical reasoning, critical thinking, and the application of principles to new scenarios.
4. **Scriptural**: Test the model's ability to reference, interpret, and explain passages from religious or philosophical texts.
5. **Terminology**: Probe the model's understanding of specific technical terms and concepts.

For a sample set of questions across the above five categories, please refer to Appendix Tab. 4.

5.1 Automatic Evaluation

Inspired by Xu et al. (2023), we conduct an extensive automatic evaluation of the two RAG models on our evaluation set. We describe each metric type below and provide implementation details in Appendix Sec. A. Due to the lack of gold answers, we are unable to report reference-based metrics.

Answer-only metrics: We assess features like fluency and coherence by analyzing responses with specific metrics: (1) **Self-BLEU** (Zhu et al., 2018) for text diversity, where higher scores suggest less diversity, applied in open-ended text generation; (2) **GPT-2 perplexity** for textual fluency, used in prior studies on constrained generation. We also consider (3) **Word** and (4) **Sentence** counts as length-based metrics, owing to their significant influence on human preferences (Sun et al., 2019; Liu et al., 2022; Xu et al., 2023).

(Question, answer) metric: To ensure answers are *relevant* to the posed questions, we model $p(q|a)$ for ranking responses with **RankGen** (Krishna et al., 2022). This encoder, leveraging the T5-XXL architecture, is specially trained via contrastive learning to evaluate sequences generated by models based on their congruity with a given prefix, in this context, the question. A higher RankGen score indicates a stronger alignment between the question and the answer, serving as a measure of relevance.

(Answer, evidence) metric: A key challenge in LFQA is assessing answer correctness without dedicated factuality metrics, akin to summarization's faithfulness. We apply **QAFactEval** (Fabbri et al., 2022), originally for summarization, to LFQA by considering the answer as a summary and evidence documents as the source. Answers deviating from source content, through hallucinations or external knowledge, will score lower on this metric.

5.2 Human Evaluation

We have three experienced domain experts evaluate the models across the five categories. Each of these experts is closely associated with Vedanta Society of New York, and has extensively studied the philosophy in question for up to a decade on average, being well-versed with domain-specific terminology and conceptual analysis. We conduct the human survey along two dimensions: retrieval and generation. For retrieval, we evaluate relevance and completeness, and for generation we evaluate factual correctness and completeness. In addition, we ask the reviewers to provide free-form justification for their choices, which proves to be very useful in analyzing the two models.

Relevance: Defined as the relevance of the retrieved passages to the user query, this metric is scored on a scale from 1 to 5 (where 1 = Not at all relevant, 5 = Extremely relevant).

Correctness: Factual accuracy of the generated answer (1 = Factually inaccurate, 5 = No inaccuracies)

Completeness: This metric measures if the retrieved passage and generated answer comprehensively cover all parts of the query (1 = Not at all comprehensive - misses crucial points, 5 = Very comprehensive and specific).

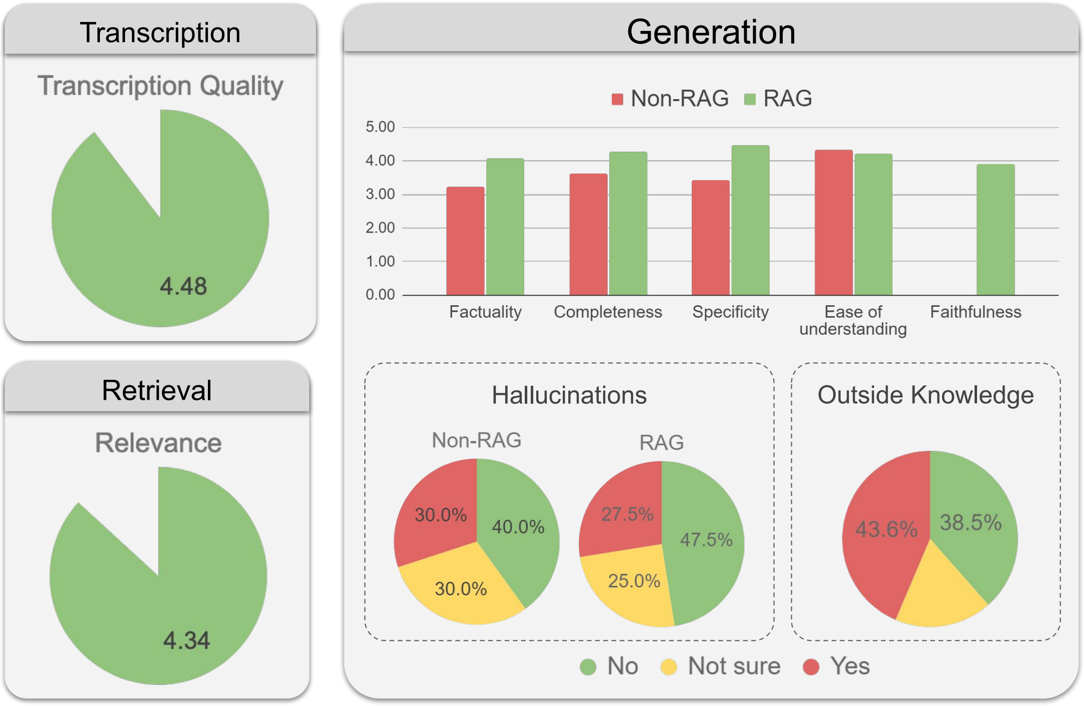

Figure 3: **Human evaluation: RAG vs non-RAG.** The RAG model outperforms the generic model across various metrics, particularly in factuality, completeness and specificity, while being marginally lower in ease of understanding.

5.3 Results: RAG vs Non-RAG

We first conduct a human evaluation survey with 5 computational linguists and 3 domain experts on RAG vs non-RAG models. In the evaluation of the generation capabilities of our models, we consider five metrics: factuality, completeness, specificity, ease of understanding, and faithfulness. The performance of the RAG model is compared against a baseline non-RAG model across these dimensions in Fig. 3. The RAG model substantially outperforms the non-RAG model across various metrics, particularly in factuality, completeness and specificity, while being marginally lower in ease of understanding. Sample responses in Figs. 6-10.

5.4 Results: Standard RAG vs Keyword-based RAG

We report results in Tab. 1. All human evaluation scores are normalized between 0 to 1. The keyword based RAG model shows strong improvement across all automatic metrics while significantly outperforming the standard model in the human evaluation. Amongst the answer-only metrics, the model tends to produce longer, more comprehensive answers (indicated by longer length), which are more coherent (lower self-bleu and perplexity). The question-answer RankGen (Krishna et al., 2022) metric evaluates the probability of the answer given the question. A higher score for the model suggests more relevant answers to the question. Most notably, the keyword model does very well on QAFactEval (Fabbri et al., 2022). It evaluates faithfulness by comparing answers from the summary (in our case, the answer) and the evidence document (retrievals). A higher score indicates greater faithfulness of the answer to retrieved passages, indicating fewer hallucinations and reliance on outside knowledge.

Coming to the human evaluation, from Tab. 1, a relevance rating of 0.87 for keyword-based RAG vs 0.58 for standard RAG indicates a strong alignment between the retrieved content and the users' queries for our model, demonstrating the efficacy of the retrieval process. On the other hand, the standard model sometimes fails to disambiguate unique terminology and retrieves incorrect passages (see Fig. 11). In assessing the accuracy of the generated answer, the keyword-based RAG model significantly outperforms the standard model, indicating better alignment with verifiable facts. Refer to Fig. 12 for an example of a factually inaccurate response from the generic model. The keyword model gets higher completeness scores for both the retrievals as well as generation. Sample responses

Category RAG Model	<u>Mean</u> M1/M2	Anecdotal M1/M2	Comparative M1/M2	Reasoning M1/M2	Scriptural M1/M2	Terminology M1/M2
			Automatic metrics			
Answer-only						
Self-bleu ↓	0.16/**0.13**	0.11/**0.05**	0.10/**0.06**	**0.15**/0.27	**0.13**/0.16	**0.09**/0.14
GPT2-PPL ↓	16.6/**15.3**	**16.6/16.6**	16.9/**15.7**	13.9/**11.9**	**14.2**/14.7	21.5/**17.7**
# Words ↑	196/**227**	**189/189**	174/**206**	218/**282**	225/**243**	216/**261**
# Sentences ↑	9.0/**10.1**	**8.2**/7.6	7.8/**9.4**	9.6/11.8	10.0/**10.6**	9.4/**11.0**
(Question, answer)						
RankGen ↑	0.46/**0.48**	0.42/**0.52**	0.44/**0.47**	0.41/**0.43**	0.51/**0.52**	**0.52**/0.46
(Answer, retrievals)						
QAFactEval ↑	1.36/**1.60**	1.01/**1.14**	1.53/**1.94**	1.18/**1.61**	**1.52**/1.36	1.56/**1.95**
			Human evaluation			
Retrieval						
Relevance ↑	0.59/**0.82**	0.41/**0.88**	0.79/**0.85**	0.73/**0.83**	0.48/**0.73**	0.55/**0.81**
Completeness ↑	0.52/**0.79**	0.41/**0.86**	0.72/**0.79**	0.57/**0.83**	0.37/**0.68**	0.52/**0.79**
Answer						
Correctness ↑	0.61/**0.86**	0.40/**0.89**	0.81/**0.88**	0.71/**0.85**	0.52/**0.81**	0.63/**0.89**
Completeness ↑	0.58/**0.85**	0.42/**0.92**	0.80/**0.85**	0.72/**0.81**	0.49/**0.77**	0.63/**0.91**

Table 1: **Automatic and human evaluation: standard RAG (M1) vs keyword-based RAG (M2).** We report both automatic and human evaluation metrics calculated on 25 triplets of {question, answer, retrievals} across 5 different question categories. The key-word based RAG model shows strong improvement across all automatic metrics while significantly outperforming the standard model in the human evaluation.

are shown in Figs. 11-15.

6 Challenges

The evaluation in Sec. 5 shows that the RAG model provides responses that are not only more aligned with the source material but are also more comprehensive, specific, and user-friendly compared to the responses generated by the generic language model. In this section, we discuss the challenges we encountered while building the retrieval-augmented chatbot for the niche knowledge domain of ancient Indian philosophy introduced in this work.

Transcription. Our requirement of using a niche data domain having long-tail knowledge precludes the use of source material that the LLM has previously been exposed to. To ensure this, we construct a textual corpus that is derived from automated transcripts of YouTube discourses. These transcripts can sometimes contain errors such as missing punctuations, incorrect transcriptions, and transliterations of Sanskrit terms. A sample of such errors is shown in Appendix Tab. 3. A proofreading mechanism and/or improved transcription models can help alleviate these issues to a large extent.

Spoken vs written language. Unlike traditional textual corpora that are compiled from written sources, our dataset is derived from spoken discourses. Spoken language is often more verbose and less structured than written text, with the speaker frequently jumping between concepts mid-sentence. This unstructured nature of the text can be unfamiliar for a language model trained extensively on written text, which expects a more coherent and structured input. A peculiar failure case arising from this issue is shown in Appendix Fig. 5. This can be addressed by converting the spoken text into a more structured prose format with the help of well-crafted prompts to LLMs, followed by human proofreading.

Context length. The passages retrieved in the standard model are of a fixed length and can sometimes be too short for many queries, especially for long-form answering. As an example, the retrieved passage may include a snippet from the middle of the full context. As a result, the chatbot response may be incomplete or incoherent (Fig. 10). This motivated us to employ a keyword-based context-expansion mechanism to provide a more comprehensive context. While this results in much better answer generation, the retrieved passage may contain too much information, making it difficult for the generator to reason effectively. Moreover, the increase in the number of tokens increases processing time. Future work can explore more advanced retrieval models capable of processing longer contexts and summarizing them effectively before input to the LLM.

Retrieval-induced hallucinations. There are scenarios when the RAG models can latch onto a particular word or phrase in the retrieved passage and hallucinates a response that is not only irrelevant but also factually incorrect. A sample of such a hallucination is shown in Fig. 9. This is a more challenging problem to address, however retrieval models that can extract the full context, summarize it and remove irrelevant information should be capable of mitigating this issue to a reasonable extent.

7 Conclusion

In this work, we integrate modern retrieval-augmented large language models with the ancient Indian philosophy of Advaita Vedanta. Toward this end, we present VedantaNY-10M, a large dataset curated from automatic transcriptions of extensive philosophical discourses on YouTube. Validating these models along various axes using both automatic and human evaluation provides two key insights. First, RAG models significantly outperform non-RAG models, with domain experts expressing a strong preference for using such RAG models to supplement their day-to-day study. Second, the keyword-based hybrid RAG model underscores the merits of integrating classical and contemporary deep learning techniques for retrieval in niche and specialized domains. While there is much work to be done, our study underscores the potential of integrating modern machine learning techniques to unravel ancient knowledge systems.

Limitations and Future Work

While our study demonstrates the utility of integrating retrieval-augmented LLMs with ancient knowledge systems, there are limitations and scope for future work. First, this study is conducted for a single niche domain of Advaita Vedanta as taught by a single teacher. Extending this study to include other ancient systems of philosophy such as the Vedantic schools of Vishishtadvaita, Dwaita, as well as the various Buddhist and Jain schools will be an interesting extension of this work. Second, expanding the evaluation set and involving more subjects for evaluation and will considerably strengthen the scope of the study. Third, in addition to the spoken discourses, incorporating the primary scriptural sources that the philosophical school is based on will further enhance the authenticity of the RAG model generation. Fourth, while we only experiment with RAG models in this study, fine-tuning the language models themselves on the philosophy datasets is an interesting future direction. Finally, while the language models in this work are primarily in English and in the Latin script, building native LLMs having the capacity to function in the original Sanskrit language of the scriptures using Devanagari script is essential future work.

Acknowledgments

The author would like to thank Prof. Kyle Mahowald for an insightful course on form and functionality in LLMs. The material covered in the course served as a guide to evaluate the models presented in this paper. Fangyuan Xu helped with information on the automatic metrics for LFQA evaluation. The author would also like to thank all the human evaluators who participated in the RAG bot evaluation survey and provided valuable feedback—special thanks to Dr. Anandhi who coordinated the effort amongst domain experts. Finally, the author would like to extend their gratitude to the Vedanta Society of New York and Swami Sarvapriyananda for the 750+ hours of public lectures that served as the dataset for this project.

References

Tom Aarsen. 2020. SpanMarker for Named Entity Recognition.

Yannis Assael, Thea Sommerschield, Brendan Shillingford, Mahyar Bordbar, John Pavlopoulos, Marita Chatzipanagiotou, Ion Androutsopoulos, Jonathan Prag, and Nando de Freitas. 2022. Restoring and attributing ancient texts using deep neural networks. *Nature*, 603.

David Bamman and Patrick J Burns. 2020. Latin bert: A contextual language model for classical philology. *arXiv preprint arXiv:2009.10053*.

Bhagavad Gita. 3000 B.C.E. The Bhagavad Gita.

Bartosz Bogacz and Hubert Mara. 2020. Period classification of 3d cuneiform tablets with geometric neural networks. In *2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR)*. IEEE.

Bernd Bohnet, Vinh Q Tran, Pat Verga, Roee Aharoni, Daniel Andor, Livio Baldini Soares, Jacob Eisenstein, Kuzman Ganchev, Jonathan Herzig, Kai Hui, et al. 2022. Attributed question answering: Evaluation and modeling for attributed large language models. *arXiv preprint arXiv:2212.08037*.

Sebastian Borgeaud, Arthur Mensch, Jordan Hoffmann, Trevor Cai, Eliza Rutherford, Katie Millican, George Bm Van Den Driessche, Jean-Baptiste Lespiau, Bogdan Damoc, Aidan Clark, et al. 2022. Improving language models by retrieving from trillions of tokens. In *International conference on machine learning*. PMLR.

The Brahmasutras. 3000 B.C.E. The Brahmasutras.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*.

Anthony Chen, Gabriel Stanovsky, Sameer Singh, and Matt Gardner. 2020. MOCHA: A dataset for training and evaluating generative reading comprehension metrics. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6521–6532, Online. Association for Computational Linguistics.

Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. 2020. ELECTRA: Pre-training text encoders as discriminators rather than generators. In *ICLR*.

Alexander Fabbri, Chien-Sheng Wu, Wenhao Liu, and Caiming Xiong. 2022. QAFactEval: Improved QA-based factual consistency evaluation for summarization. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2587–2601, Seattle, United States. Association for Computational Linguistics.

Maarten Grootendorst. 2020. Keybert: Minimal keyword extraction with bert.

Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat, and Mingwei Chang. 2020. Realm: Retrieval augmented language model pre-training. In *International conference on machine learning*. PMLR.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python.

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

Albert Q. Jiang, Alexandre Sablayrolles, Antoine Roux, Arthur Mensch, Blanche Savary, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Emma Bou Hanna, Florian Bressand, Gianna Lengyel, Guillaume Bour, Guillaume Lample, Lélio Renard Lavaud, Lucile Saulnier, Marie-Anne Lachaux, Pierre Stock, Sandeep Subramanian, Sophia Yang, Szymon Antoniak, Teven Le Scao, Théophile Gervet, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2024. Mixtral of experts. *Preprint*, arXiv:2401.04088.

Nikhil Kandpal, Haikang Deng, Adam Roberts, Eric Wallace, and Colin Raffel. 2023. Large language models struggle to learn long-tail knowledge. In *International Conference on Machine Learning*. PMLR.

Vladimir Karpukhin, Barlas Oğuz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. 2020. Dense passage retrieval for open-domain question answering. *EMNLP*.

Urvashi Khandelwal, Omer Levy, Dan Jurafsky, Luke Zettlemoyer, and Mike Lewis. 2020. Generalization through memorization: Nearest neighbor language models. *ICLR*.

Kalpesh Krishna, Yapei Chang, John Wieting, and Mohit Iyyer. 2022. Rankgen: Improving text generation with large ranking models. *arXiv preprint arXiv:2205.09726*.

Kalpesh Krishna, Aurko Roy, and Mohit Iyyer. 2021. Hurdles to progress in long-form question answering. Association for Computational Linguistics.

Philippe Laban, Tobias Schnabel, Paul N. Bennett, and Marti A. Hearst. 2022. Summac: Re-visiting nli-based models for inconsistency detection in summarization. *Transactions of the Association for Computational Linguistics*, 10:163–177.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020a. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020b. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*.

Yixin Liu, Alexander R. Fabbri, Pengfei Liu, Yilun Zhao, Linyong Nan, Ruilin Han, Simeng Han, Shafiq R. Joty, Chien-Sheng Wu, Caiming Xiong, and Dragomir R. Radev. 2022. Revisiting the gold standard: Grounding summarization evaluation with robust human evaluation. *ArXiv*, abs/2212.07981.

Ligeia Lugli, Matej Martinc, Andraž Pelicon, and Senja Pollak. 2022. Embeddings models for buddhist Sanskrit. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*.

Rajiv Malhotra. 2021. *Artificial Intelligence and the Future of Power: 5 Battlegrounds*. Rupa Publications.

Rajiv Malhotra and Satyanarayana Dasa Babaji. 2020. *Sanskrit Non-Translatables: The Importance of Sanskritizing English*. Amaryllis.

Alex Mallen, Akari Asai, Victor Zhong, Rajarshi Das, Hannaneh Hajishirzi, and Daniel Khashabi. 2023. When not to trust language models: Investigating effectiveness and limitations of parametric and non-parametric memories. *ACL*.

Priyanka Mandikal and Raymond Mooney. 2024. Sparse meets dense: A hybrid approach to enhance scientific document retrieval. In *The 4th CEUR Workshop on Scientific Document Understanding, AAAI*.

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.

Jacob Menick, Maja Trebacz, Vladimir Mikulik, John Aslanides, Francis Song, Martin Chadwick, Mia Glaese, Susannah Young, Lucy Campbell-Gillingham, Geoffrey Irving, et al. 2022. Teaching language models to support answers with verified quotes. *arXiv preprint arXiv:2203.11147*.

Ragaa Moustafa, Farida Hesham, Samiha Hussein, Badr Amr, Samira Refaat, Nada Shorim, and Taraggy M Ghanim. 2022. Hieroglyphs language translator using deep learning techniques (scriba). In *2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*. IEEE.

Reiichiro Nakano, Jacob Hilton, Suchir Balaji, Jeff Wu, Long Ouyang, Christina Kim, Christopher Hesse, Shantanu Jain, Vineet Kosaraju, William Saunders, et al. 2021. Webgpt: Browser-assisted question-answering with human feedback. *arXiv preprint arXiv:2112.09332*.

Sonika Narang, MK Jindal, and Munish Kumar. 2019. Devanagari ancient documents recognition using statistical feature extraction techniques. *Sādhanā*, 44.

Zach Nussbaum, John X. Morris, Brandon Duderstadt, and Andriy Mulyar. 2024. Nomic embed: Training a reproducible long context text embedder. *Preprint*, arXiv:2402.01613.

OpenAI. 2023. Gpt-4 technical report.

Asimina Paparigopoulou, John Pavlopoulos, and Maria Konstantinidou. 2022. Dating greek papyri images with machine learning. In *ICDAR Workshop on Computational Paleography, https://doi. org/10.21203/rs*, volume 3.

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2023. Robust speech recognition via large-scale weak supervision. In *International Conference on Machine Learning*, pages 28492–28518. PMLR.

Ori Ram, Yoav Levine, Itay Dalmedigos, Dor Muhlgay, Amnon Shashua, Kevin Leyton-Brown, and Yoav Shoham. 2023. In-context retrieval-augmented language models. *arXiv preprint arXiv:2302.00083*.

Ramakrishna Order. Belur math.

Adi Shankaracharya. 450 B.C.E. Commentary on the upanishads.

Weijia Shi, Julian Michael, Suchin Gururangan, and Luke Zettlemoyer. 2022. Nearest neighbor zero-shot inference. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*.

Weijia Shi, Sewon Min, Michihiro Yasunaga, Minjoon Seo, Rich James, Mike Lewis, Luke Zettlemoyer, and Wen-tau Yih. 2023. Replug: Retrieval-augmented black-box language models. *arXiv preprint arXiv:2301.12652*.

Kurt Shuster, Spencer Poff, Moya Chen, Douwe Kiela, and Jason Weston. 2021. Retrieval augmentation reduces hallucination in conversation. *ACL*.

Thea Sommerschield, Yannis Assael, John Pavlopoulos, Vanessa Stefanak, Andrew Senior, Chris Dyer, John Bodel, Jonathan Prag, Ion Androutsopoulos, and Nando de Freitas. 2023. Machine Learning for Ancient Languages: A Survey. *Computational Linguistics*.

Mingyang Song, Xuelian Geng, Songfang Yao, Shilong Lu, Yi Feng, and Liping Jing. 2024. Large language models as zero-shot keyphrase extractors: A preliminary empirical study. *Preprint*, arXiv:2312.15156.

Simeng Sun, Ori Shapira, Ido Dagan, and Ani Nenkova. 2019. How to compare summarizers without target length? pitfalls, solutions and re-examination of the neural summarization literature. *Proceedings of the Workshop on Methods for Optimizing and Evaluating Neural Language Generation*.

Upanishads. >3000 B.C.E. The Upanishads.

Vedanta Society of New York. Swami Sarvapriyananda's Vedanta discourses.

Lee Xiong, Chuan Hu, Chenyan Xiong, Daniel Campos, and Arnold Overwijk. 2019. Open domain web keyphrase extraction beyond language modeling. *arXiv preprint arXiv:1911.02671*.

Fangyuan Xu, Yixiao Song, Mohit Iyyer, and Eunsol Choi. 2023. A critical evaluation of evaluations for long-form question answering. In *Association of Computational Linguistics*.

Eunice Yiu, Eliza Kosoy, and Alison Gopnik. 2023. Imitation versus innovation: What children can do that large language and language-and-vision models cannot (yet)? *ArXiv*, abs/2305.07666.

Yaoming Zhu, Sidi Lu, Lei Zheng, Jiaxian Guo, Weinan Zhang, Jun Wang, and Yong Yu. 2018. Texygen: A benchmarking platform for text generation models. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 1097–1100.

A Implementation Details

A.1 Automatic Metrics

Following Xu et al. (2023), we implement a number of automatic evaluation metrics for LFQA as described below.

Length We use the Spacy package (Honnibal et al., 2020) for word tokenization.

Self-BLEU We calculate Self-BLEU by regarding one sentence as hypothesis and all others in the same answer paragraph as reference. We report self-BLEU-5 as a measure of coherence.

RankGen For a given question q and a model-generated answer a, we first transform them into fixed-size vectors $(\boldsymbol{q}, \boldsymbol{a})$ using the RankGen encoder (Krishna et al., 2022). To assess their relevance, we compute the dot product $\boldsymbol{q} \cdot \boldsymbol{a}$. We utilize the T5-XXL (11B) encoder, which has been trained using both in-book negative instances and generative negatives.

QAFactEval QAFactEval is a QA-based metric recently introduced by Fabbri et al. (2022). It has demonstrated exceptional performance across multiple factuality benchmarks for summarization (Laban et al., 2022; Maynez et al., 2020). The pipeline includes four key components: (1) Noun Phrase (NP) extraction from sentence S represented as $Ans(S)$, (2) BART-large (Lewis et al., 2020a) for question generation denoted as Q_G, (3) Electra-large (Clark et al., 2020) for question answering labeled as Q_A, and (4) learned metrics LERC (Chen et al., 2020), to measure similarity as $Sim(p_i, s_i)$. An additional answerability classification module is incorporated to assess whether a question can be answered with the information provided in document D. Following Xu et al. (2023), we report LERC, which uses the learned metrics to compare Ans_S and Ans_D(a).

A.2 Chat Prompt

For an example of the constructed RAG bot prompt, please refer to Fig. 4. In this scenario, the RAG bot C_r is presented with the top-k retrieved passages alongside the query for generating a response, whereas a generic bot C_g would only receive the query without additional context.

B Sample Sanskrit terms

Tab. 2 contains excerpts from passages containing Sanskrit terms. The Sanskrit terms are italicized

RAG Bot

You are a helpful assistant that accurately answers queries using Swami Sarvapriyananda's YouTube talks. Use the following passages to provide a detailed answer to the query: **{query}**
Passages:
{Passage 1}
{Passage 2}
...
{Passage k}

Generic Bot

You are a helpful assistant that accurately answers queries using Swami Sarvapriyananda's YouTube talks. Provide a detailed answer to the query: **{query}**

Figure 4: **Prompts for the RAG and generic chatbots.** RAG Bot receives the top-k retrieved relevant passages in the prompt along with the query, while the generic bot only receives the query.

and underlined. Notice that the passages contain detailed English explanations of these terms. To retain linguistic diversity, authenticity and comprehensiveness of the source material, we retain these Sanskrit terms as is in our passages as described in Sec. 3. Note that these are direct Whisper (Radford et al., 2023) transcriptions with no further post-processing or proofreading. Transcriptions may not always be accurate.

C Transcription

We asses the transcript quality and list out some common errors.

C.1 Transcript Evaluation

Transcription quality is scored on a scale from 1 to 5 (where 1 = Poor, 5 = Perfect). On 10 randomly sampled transcripts, evaluators assign a high average score of 4.48 suggesting that the transcription of YouTube audio into text is highly accurate and clear, indicating that our constructed custom dataset D_t is of high quality.

C.2 Transcript Errors

Tab. 3 contains a few sample transcription errors. The transcriptions are largely good for English words and sentences. However, errors often arise from incorrectly transcribing Sanskrit terms and verses. Other less common errors include missing or incorrect punctuation. Human proofreading will remove these errors to a large extent.

D Spoken vs written language

Unlike traditional textual corpora that are compiled from written sources, our dataset is derived from spoken discourses. Spoken language is often more verbose and less structured than written text, with the speaker frequently jumping between concepts mid-sentence. This unstructured nature of the text can be unfamiliar for a language model trained extensively on written text, which expects a more coherent and structured input. Fig. 6 shows a peculiar failure case arising from this issue. The speaker jumps between speaking about Aldous Huxley and Alan Watts. As a result, the RAG bot incorrectly attributes information about Huxley to Watts while answering the query. Due to the unstructured nature of the spoken text, such errors can occur. This can be addressed by converting the spoken text into a more structured prose format with the help of well-crafted prompts to LLMs, followed by human proofreading.

E Question Categories

To ensure a broad and comprehensive evaluation, we categorize the questions into five distinct types—anecdotal, teminology, scriptural, comparative and reasoning. Questions from each of these categories is provided in Tab. 4. Each category is designed to test different aspects of the model's capabilities. For details, refer to Sec. 5.

F Sample Responses: RAG vs Non-RAG

In Figs. 6-10, we present the sample queries and responses from the RAG and generic chatbots for five sample queries. The query is highlighted in yellow, RAG response is in the green box. Non-RAG response is in the red box. We use GPT-4-turbo (OpenAI, 2023) as the large language model for both of them.

G Sample Responses: Standard RAG vs Keyword-based Hybrid RAG

In Figs. 11-15, we present the sample queries and responses from the standard RAG and keyword-based hybrid RAG model for one query each from five categories. We additionally also show the human-annotated and model generated keywords. Keyword-RAG response is in the green box. Standard-RAG response is in the red box.

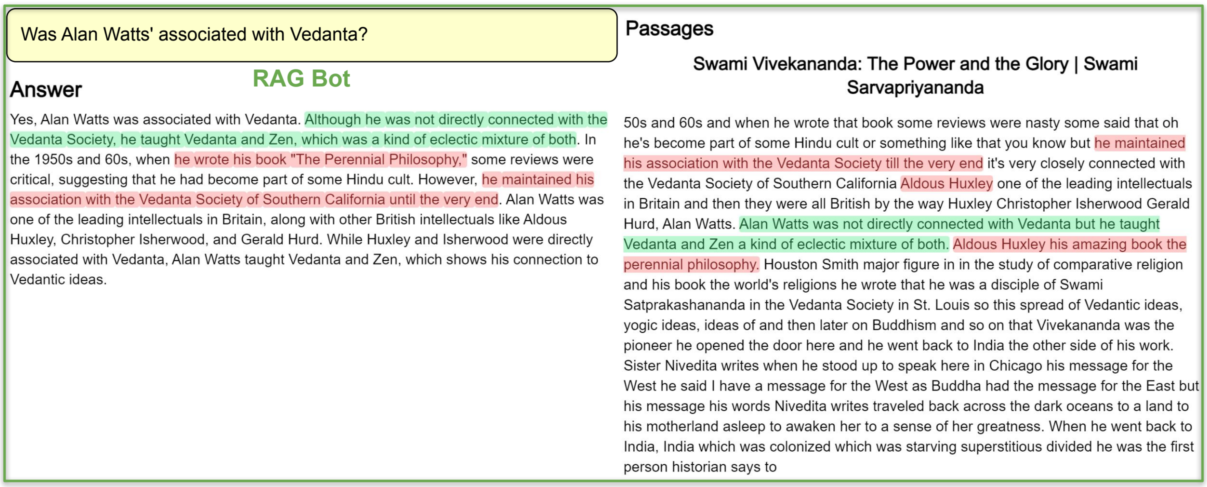

Figure 5: **Subject confusion: Aldous Huxley vs Alan Watts.** In the retrieved passage, the speaker jumps between speaking about Aldous Huxley and Alan Watts. Moreover, the transcription has poor punctuation. As a result, the RAG bot incorrectly attributes information about Huxley to Watts in the response. Due to the unstructured nature of the spoken text, such errors can occur. This can be addressed by converting the spoken text into a more structured prose format with the help of well-crafted prompts to LLMs, followed by human proofreading.

Sl. No.	Excerpts from passages	Notes
1.	*Om Bhadram Karne Bhishrinu Yamadevaha Bhadram Pashyam Akshabhirya Jatraaha Sthirai Rangai Stushta Vagam Sasthanubhi Vyase Madevahitaiyadayoh Swasthina Indro Vriddha Shravaha Swasthina Phusa Vishwa Vedaaha Swasthina Starksho Arishta Nemi Swasthino Brihas Patir Dadhatu Om Shanti Shanti Shanti*.	This is a Sanskrit chant which is directly Romanized and processed. The automatic transcriptions often contain errors in word segmentation for Sanskrit verses.
2.	*Samsara* is our present situation, the trouble that we are caught in, the mess that we are caught in. *Samsara* is this. In Sanskrit, normally when you use the word *samsara*, it really means this world of our life, you know, being born and struggling in life and afflicted by suffering and death and hopelessness and meaninglessness.	*Samsara* is a Sanskrit term. The excerpt contains an explanation of the concept in English.
3.	The problem being ignorance, solution is knowledge and the method is *Jnana Yoga*, the path of knowledge. So what is mentioned here, *Shravana* *Manana* *Nididhyasana*, hearing, reflection, meditation, that is *Jnana Yoga*. So that's at the highest level of practice, way of knowledge.	The excerpt contains an explanation of *Jnana Yoga*, the path of knowledge.
4.	In Sanskrit, *ajnana* and *adhyasa*, ignorance and superimposition. Now if you compare the four aspects of the self, the three appearances and the one reality, three appearances, waker, dreamer, deep sleeper, the one reality, *turiyam*, if you compare them with respect to ignorance and error, you will find the waker, that's us right now. We have both ignorance and error.	*Ajnana*, *adhyasa* and *turiyam* are Sanskrit terms. Notice that the passage implicitly contains rough English translations of these terms in the context of the overall discourse. For instance, *ajnana* is translated as ignorance and *adhyasa* is translated as superimposition.
5.	*Mandukya* investigates this and points out there is an underlying reality, the *Atman*, pure consciousness, which has certain characteristics. This is causality, it is beyond causality. It is neither a cause nor an effect. The *Atman* is not produced like this, nor is it a producer of this. It is beyond change. No change is there in the *Atman*, *nirvikara*. And third, it is not dual, it is non-dual, *advaitam*. This is *kadyakarana* in Sanskrit, this is *kadyakarana vilakshana Atma*. In Sanskrit this is *savikara*, this is *nirvikara Atma*. This is *dvaita*, this is *advaita Atma*. So this is *samsara* and this is *moksha*, freedom.	The excerpt contains an explanation of different Sanskrit technical terms.

Table 2: **Excerpts from passages containing Sanskrit terms.** These excerpts contain detailed English descriptions of technical terms in Sanskrit (italicized and underlined). To retain authenticity to the source material, we retain these Sanskrit terms as is in our passages. Note that these are direct Whisper (Radford et al., 2023) transcriptions with no further post-processing or proofreading, so transcriptions may not always be accurate. For more details, refer to Sec. B.

Sl. No.	Transcription errors	Notes
1.	That's what Sam Altman, Chachjipiti, somebody asked him.	Should be ChatGPT
2.	Last year, you studied extensively with Professor Garfield, I believe, studying *Vajamaka* and the teachings of the *Garjuna*.	Should be *Madhyamaka* and *Nagarjuna*, respectively
3.	From attachment comes desire, *raga*, I want it and if that desire is satisfied then there is no end to it, greed, *lobha*. But if it is somehow thwarted, then anger, *kama krodho vijayate*.	Should be *bhijayate*
4.	In fact, one of the terms which is used in Mandukya Upanishad, Brahman is *abhyavaharyam*.	Should be *avyavaharam*
5.	So, one of them was the *Brahmo Samad*, which was quite popular in Calcutta in those days.	Should be *Samaj*
6.	I am awareness I'm eternal consciousness Aldous Huxley Christopher Isherwood Gerald Hurd all of them were very close to Swami Prabhavananda in Southern California in Hollywood and look at the product of that Isherwood wrote that one of the most amazing biographies	The transcripts sometimes miss punctuation marks, making the passage difficult to comprehend for both humans and language models

Table 3: **Sample transcription errors.** For constructing our text corpus, we directly use the transcripts obtained from Whisper (Radford et al., 2023) with no further post-processing or proofreading. The transcriptions are largely good (with a score of 4.5/5 from human evaluators). However, errors arise from incorrectly transcribing Sanskrit terms, missing punctuations, etc. Human proofreading will remove these errors to a large extent.

Category	Description	Questions
Anecdotal	Stories and anecdotes narrated by the speaker in the discourses	• Does Swami speak about Wittgenstein's thesis defense? • Does Swami narrate any incident surrounding Shivaratri? • Does Swami speak about The Matrix movie? • Does Swami speak about Vachaspati Mishra? Does he narrate how Bhamati came to be written? • What was Christopher Isherwood's contribution to Vedanta?
Terminology	Probe the model's understanding of specific terms and concepts	• What is Adhyaropa Apavada? • What is Vikshepa Shakti? • What is the significance of the word 'Shraddha'? • What is Upadana Karana? • What constitutes Sadhana Chatushtaya?
Scriptural	Reference, interpret, and explain passages from religious or philosophical texts	• In Mandukya Upanishad, what is the significance of the word 'Om'? • In the Gospel, what parable does Sri Ramakrishna use to portray intense longing for God? • In the Mundaka Upanishad, how do we interpret the parable of the two birds? • How is Phala Vyapti and Vritti Vyapti defined in Vedantasara? • In the Gospel of Sri Ramakrishna, how do we understand the analogy of the salt doll that Thakur provides?
Comparative	Analyze and compare different concepts, philosophies, or texts	• As mentioned in the Yoga Sutras, is Samadhi necessary to attain enlightenment according to Advaita Vedanta? • Would Sri Ramakrishna's teachings be considered purely Advaitic? • In Kashmir Shaivism, Chit is both Prakasha and Vimarsha i.e. it is both self-lumious and self-reflective. Would Advaita Vedanta agree with this? • How does Sankhya differ from Advaita Vedanta? • What is the main difference between Buddhist Shunyavada and Advaita Vedanta?
Reasoning	Require logical reasoning, critical thinking, and the application of principles to new scenarios	• Can AI ever become conscious? • Is the waking state similar to a dream or absolutely indistinguishable from the dream state? • Do our senses report reality to us? • Dis-identifying myself from the body-mind seems to require a lot of effort. Did the reverse process of identification also involve as much effort? • If Brahman as Existence-Consciousness-Bliss is the eternal subject that cannot be objectified, yet eternally reveals itself to give rise to apparent objects through the principle of Maya, can we infer that Brahman is Existence-Consciousness-Bliss-Maya? That is, is Brahman inseparable from Maya?

Table 4: **Question categories.** To ensure a broad and comprehensive evaluation, we categorize the questions into five distinct types, each designed to test different aspects of the model's capabilities. For details, refer to Sec. 5.

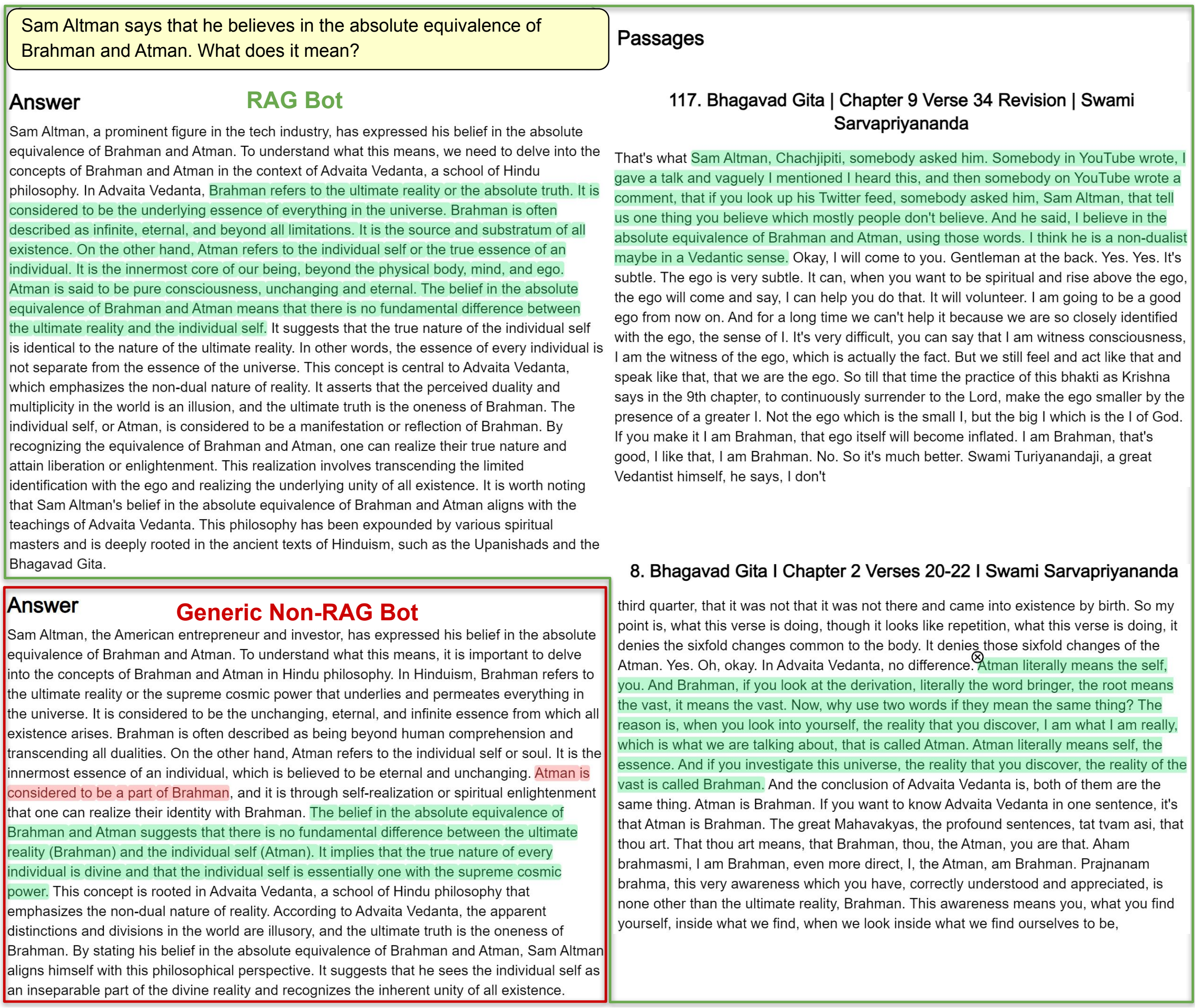

Figure 6: **RAG vs Non-RAG: Sam Altman on Atman and Brahman.** The retrieved passages are very good at capturing the context of the query. The first passage references the monk speaking about Sam Altman and his views on Atman and Brahman. The second passage contains a brief explanation of Atman and Brahman that is helpful to answer the query. The generic bot also generates a reasonable answer. However, it has a sentence stating that, "Atman is a part of Brahman". This is incorrect in the context of Advaita Vedanta as it is not the same as equivalence—this point was highlighted by the domain experts in the survey.

Does Swamiji speak about Wittgenstein's dissertation defense?

RAG Bot

Answer

Yes, Swami Sarvapriyananda does speak about Wittgenstein's dissertation defense in one of his YouTube talks. He shares an interesting story about Wittgenstein's PhD examination at Cambridge University. The examiners for Wittgenstein's thesis were Bertrand Russell and G.E. Moore, both renowned philosophers. However, instead of asking Wittgenstein questions about his thesis, Russell and Moore started chatting with each other. Eventually, Moore suggested that they should ask Wittgenstein some questions. They opened Wittgenstein's book, Tractatus, and asked him to explain a particular statement. Wittgenstein and Russell then started quarreling over it. At this point, Wittgenstein stood up, patted the examiners on their back, and said, "Give it up, boys. You'll never get it." He then walked out of the room. Surprisingly, Russell and Moore awarded him the PhD right then and there. Swami Sarvapriyananda uses this story to emphasize the difficulty of understanding certain philosophical concepts. He mentions that even great philosophers like Wittgenstein can have disagreements and struggles in their pursuit of knowledge. Swami Sarvapriyananda also highlights the importance of engaging in dialectics and considering opposing viewpoints to deepen our understanding of our own positions. While Swami Sarvapriyananda does not directly discuss Wittgenstein's dissertation defense in relation to spiritual growth or enlightenment, he does mention the value of engaging in discussions and responding to questions in the context of teaching and learning. He believes that teaching and responding to queries provide an opportunity to think about teachings from different angles and deepen one's own understanding.

Generic Non-RAG Bot

Answer

Yes, Swami Sarvapriyananda does speak about Wittgenstein's dissertation defense in one of his YouTube talks. In his talk titled "Understanding Wittgenstein's Philosophical Investigations," Swami Sarvapriyananda discusses various aspects of Wittgenstein's work, including his dissertation defense. He provides insights into the historical context of Wittgenstein's defense, the arguments he presented, and the impact it had on his later philosophical development. Swami Sarvapriyananda also delves into the significance of Wittgenstein's ideas and their relevance in contemporary philosophy. Overall, his talk offers a comprehensive analysis of Wittgenstein's dissertation defense and its implications.

Passages

The Language of Paradox in Advaita - Swami Sarvapriyanada

thesis and we will pass it. I'll get it through. So you'll get a doctorate and you don't have to bother about it. Then you can teach. We want you to teach in Cambridge University. So the PhD examination of Wittgenstein. The story is very interesting. The examiners for a PhD thesis. Imagine who are the examiners. Bertrand Russell is the examiner. It's already fixed. So and the other examiner, those who are in philosophy, know this name is G.E. Moore. G.E. Moore. Another one of Bertrand Russell's senior contemporary. One of the great names of English philosophy. So these two are the PhD examiners. Imagine the examination. And the student is Wittgenstein. They started chatting to each other. After some time, Moore said to Russell, Bertie, don't you think we should ask him some questions? So all right. They opened that book, Tractatus. They opened that book at a page and said, explain this. This statement. Wittgenstein said something. Bertrand Russell said something. They started quarreling over it. And then student Wittgenstein, he stands up. He pats the examiners on their back. The Oxford, the Cambridge dons on their back and says, give it up, boys. You'll never get it. And he walks out. And they give him the PhD there and then. You've got a PhD. Says, give it up, boys. You'll never get it. But we'll try to understand. And we'll bring the light of Shankaracharya's commentary to bear on this problem. 1,400 years ago in India, Shankaracharya, who was 16 years old, if you go by

Vedanta Retreat at Loon Lake on Who am I by Swami Sarvapriyananda Lecture 11

philosophers Of the next generation Now, one day One of the students Was saying that Professor this is what you said in the last class And Wittgenstein said No I didn't say this And the student was pretty sure that's exactly what the professor had said So that day he wrote down the exact words What Wittgenstein said Next day he said Sir you said in the last class You said this And Wittgenstein said no no no I did not say that And he said no I have written it down exactly as you said You said this And Wittgenstein said Oh but those are my words You don't understand what I mean by those words You are using the same words But we don't mean the same thing I can see that you do not understand that Very interesting What happened to him Did he denounce the Tractatus of the Praetorians And say that it was a really good answer No, when he wrote the Tractatus He said all the problems of philosophy have been solved I give up philosophy And he went away to become a school teacher And he taught mathematics To children in a village I think in Switzerland and Austria And then he realized That all that he had done in the Tractatus Was wrong So he came back again to Cambridge And he worked on a new book And he became the Philosophical Investigations So he is probably the only philosopher Who wrote two philosophical classics In his lifetime Both opposite, completely opposite Repudiating the earlier work Tractatus and Philosophical Investigations Anyway So language One has to be careful Don't

Figure 7: **RAG vs Non-RAG: Wittgenstein's thesis defense.** The retrieved passage is about the monk narrating a story about Wittgenstein's thesis defense, which is highly relevant to the query. The generated RAG bot response makes use of this passage to answer the query. However, the latter half of the response is not relevant to the query and is purely hallucinated. The generic bot response, on the other hand, is completely fabricated and references a non-existent YouTube talk.

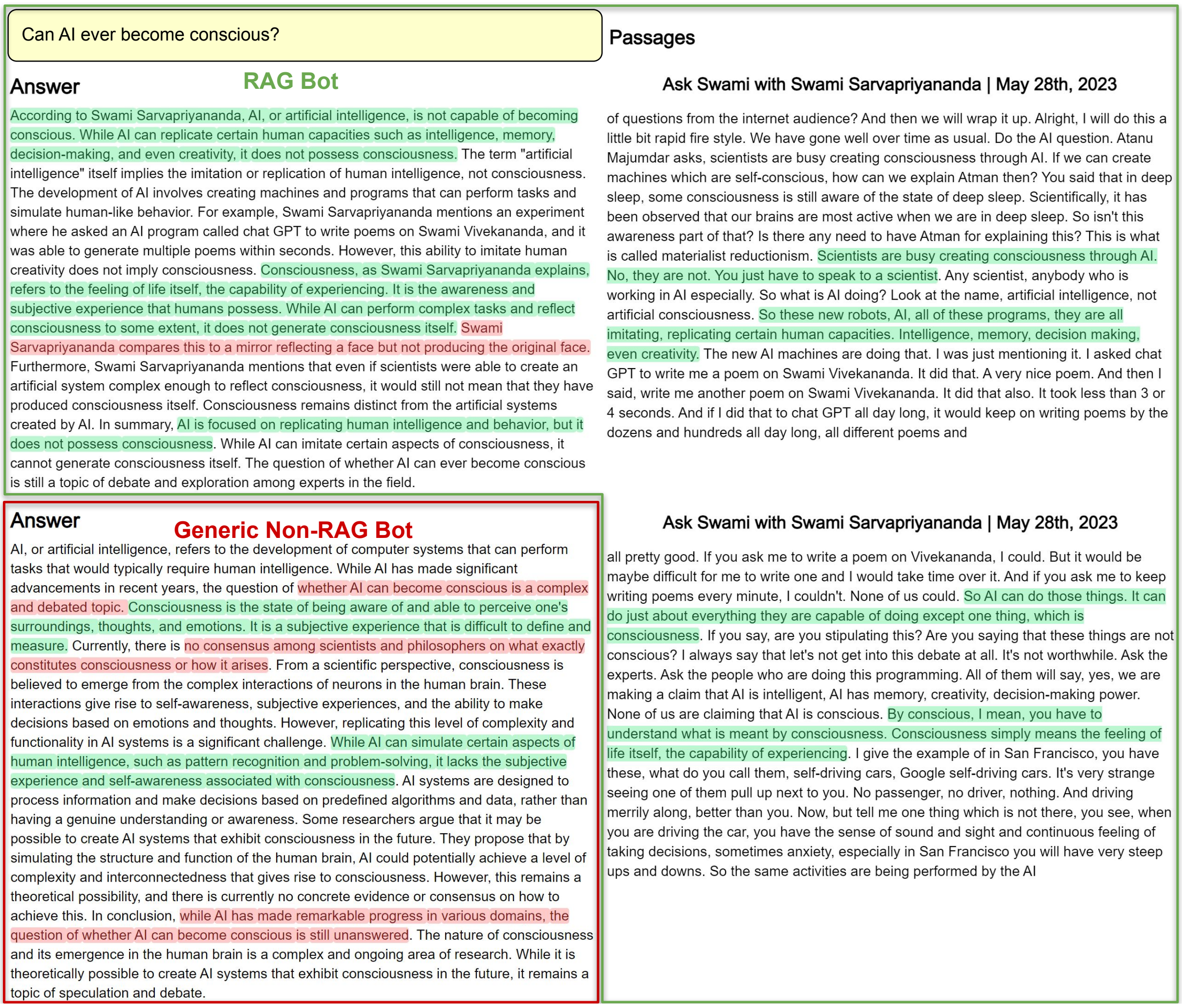

Figure 8: **RAG vs Non-RAG: AI and Consciousness.** The retrieved passages are again highly relevant to the query and the generated answer is coherent. However, it does contain a statement about "a mirror reflecting a face" which is not present in the passages. This was pointed out by the experts as a relevant statement possibly drawn from outside knowledge. The generic bot's response is highly general and does not represent the monk's views on the topic.

Do our senses report reality to us?

Answer — **RAG Bot**

According to Swami Sarvapriyananda, our senses do not report reality to us. This is a conclusion supported by modern cognitive science and evolutionary science. Our sensory organs, such as our eyes and ears, provide us with information about the world, but they do not necessarily give us an accurate representation of reality. To illustrate this point, Swami Sarvapriyananda gives the example of a virtual reality movie. In a virtual reality movie, we may become completely immersed in the experience and forget our true identity. Similarly, our senses immerse us in the experience of the world, but that does not mean that what we perceive is the ultimate reality. Advaita Vedanta, the philosophy that Swami Sarvapriyananda follows, helps us understand this perspective. It teaches us that our true nature is the experiencing consciousness, which is separate from the entities and experiences we perceive through our senses. If we mistakenly take these entities to be separate and real, we become attached to them and experience suffering. Shankaracharya, a prominent philosopher in Advaita Vedanta, explains that our ignorance of our true nature leads us to identify ourselves with our mind and body. This identification causes us to experience the external world and subjective feelings, both positive and negative. However, these problems and experiences are not inherent in the ultimate reality of consciousness. In the ultimate reality, the world is seen as an appearance, not as a separate entity. Just as a mirage does not wet a grain of sand or a rope mistaken for a snake does not become poisonous, the appearances in consciousness do not affect the reality of consciousness itself. Therefore, according to Swami Sarvapriyananda, our senses do not report reality to us. They provide us with a subjective experience of the world, but it is through the study of Vedanta and the realization of our true nature as consciousness that we can gain clarity and understand the illusory nature of our sensory perceptions.

Answer — **Generic Non-RAG Bot**

Yes, I am an AI assistant trained on a vast amount of data, including Swami Sarvapriyananda's YouTube talks. Now, to answer your query, "Do our senses report reality to us?" According to Swami Sarvapriyananda, our senses do not directly report reality to us. Instead, they provide us with a filtered and interpreted version of reality. Our senses act as a medium through which we perceive the world, but they are not infallible or objective. Our senses are limited and can only perceive a fraction of the electromagnetic spectrum. For example, we can only see a small range of colors, hear a limited range of frequencies, and so on. Additionally, our senses can be influenced by various factors such as our mental state, biases, and cultural conditioning. Furthermore, our brain plays a crucial role in processing the information received from our senses. It interprets and constructs our perception of reality based on the sensory inputs it receives. This means that our perception is not a direct representation of the external world but rather a subjective experience created by our brain. Swami Sarvapriyananda emphasizes that our senses are not to be dismissed as completely unreliable, but rather they should be understood as tools that provide us with a certain level of information about the world. However, it is important to recognize their limitations and the potential for distortion in our perception. In conclusion, our senses do not report reality to us in an unfiltered and objective manner. They provide us with a subjective interpretation of reality, influenced by various factors. Understanding this can help us develop a more nuanced and humble approach to our perception of the world.

Passages

Ask Swami with Swami Sarvapriyananda | Nov 19th, 2017

at that time. So he says you can just walk across. All right. Now it's time for our monthly Ask Swami. So there are questions. We have a wide internet audience all over the world, and they send in a lot of questions, only a few of which we can answer. They are selected. And there are also questions from the live audience. So we'll start off with some questions from the internet audience. Yes. This first question is from. Are the questions audible? Yes. This first question is from Ajoy. Modern cognitive scientists are veering more and more towards the idea that consciousness is fundamental, and there is a deeper reality in the world than our senses tell us. A honeybee has a strong nasal sensory system that takes it to the source of honey. Similarly, a bat has sound receptacles that allow it to sense danger and also find its prey. They also do not see the world as we see it. Therefore, is it not a legitimate question to ask, how correct are we to presume that what we see is indeed the reality? How can we say our sensory organs tell us the truth? Can you give us a Vedantic perspective to it and how Advaita Vedanta sees such developments? You're right. It's very interesting. As Ajoy has written, that's true. Our senses, do they report the truth to us? Do they report reality to us? And the conclusion of modern cognitive science and indeed even evolutionary science is that they do not. The senses do not report reality to us. What do you mean by that? So are they misreporting? Not

Introduction to Vedanta - Swami Sarvapriyananda - Aparokshanubhuti - Part 17 – November 29, 2016

yesterday, somebody forwarded to me a very interesting article published in the well-known magazine, The Atlantic. This is a magazine, The Atlantic. A cognitive scientist, right here, our next-door neighbor in UC Irvine, his name is Hoffman. Now, there's an interview with this scientist, and what he's saying is, what our senses reveal to us, the world, is absolutely illusory. It's an illusion. The reality is not revealed by our senses. Now, why should this be difficult to understand? Because, till now, the idea was that our senses have been generated by evolution. So, all animals, including our senses, our eyes, ears, and all our sense organs, are products of evolution. And they are there to give us an evolutionary advantage. And the thinking was, till now, till very recently, the thinking was, if the senses did not report reality, then the animal wouldn't survive. If the senses reported something else, and the world was something else, then the animal wouldn't fit in the world and wouldn't survive long enough. And so, it wouldn't transmit its genes across generations. Therefore, if we have survived for thousands and thousands of years, if our senses are the product of thousands of years of evolution, then our senses do fit the environment, so they repeat, they report to us, the reality of the environment. That was the thinking. And he says, not at all. What the senses have been evolved to do is to make us fit for survival. And we go, yes, that's what we are saying. And being

Figure 9: **RAG vs Non-RAG: Senses and reality.** In this scenario, the retrieved passages provide only partial context needed to answer the query, with additional irrelevant information. For example, one passage mentions cognitive scientist Donald Hoffman, inadvertently influencing the LLM to generate a response linking the Swami's views to Hoffman's concept of the world as a virtual reality. This is an instance where the LLM diverges from the query, using incomplete information from the passage. To mitigate this, implementing retrieval models capable of processing longer contexts and summarizing them effectively before input to the LLM could be beneficial. Despite this, the rest of the response aligns reasonably well with Advaita Vedanta philosophy, similar to the generic bot's response, which is also congruent with the philosophical context.

Dreams feel absolutely real while dreaming. What is the guarantee that I am not fast asleep in my bed and dreaming right now?

Answer — **RAG Bot**

According to Swami Sarvapriyananda, the question of whether we are dreaming or awake at any given moment is a profound one. In his talks, he refers to the insights of Gaudapada, an ancient Advaita Vedanta philosopher, who explored the nature of reality and the relationship between the dream state and the waking state. Gaudapada's insight is that there is no definitive way to differentiate between the dream state and the waking state. He argues that the clear distinction we make between dreaming and waking in our understanding is actually fuzzy. He reduces the status of our waking state experience to that of our dream state experience. One objection that often arises is the feeling of reality in dreams. Dreams can feel absolutely real while we are experiencing them. We may interact with people, objects, and environments that seem just as solid and permanent as in the waking state. This raises the question: What is the guarantee that we are not currently dreaming and that our waking experience is not just another dream? Gaudapada's response to this objection is that the feeling of reality in dreams does not necessarily indicate the truth or permanence of the objects and experiences within the dream. In a dream, we may perceive objects and people as real, but upon waking up, we realize that they were mere appearances, not possessing any substantial reality. Similarly, Gaudapada argues that the objects and experiences we perceive in the waking state are also appearances, just like in a dream. Our beloved friends, family, and surroundings may seem real and permanent, but they are as real as the "I" or the self that perceives them. In other words, the reality of the waking state is dependent on the perceiving consciousness, just as the reality of the dream state is dependent on the dreaming mind. Gaudapada's perspective is balanced and nuanced. He acknowledges that our waking experience feels real to us, just as the dream experience feels real to the dreamer. However, he also emphasizes that the reality of both states is ultimately dependent on consciousness itself. It is important to note that Gaudapada's insights are not meant to dismiss the waking state as illusory or insignificant. Rather, they invite us to question the nature of reality and our assumptions about it. The ultimate aim of Advaita Vedanta is to realize the true nature of consciousness and the underlying unity of all existence. In conclusion, according to Gaudapada's teachings, there is no guarantee that we are not currently dreaming. The feeling of reality in dreams does not provide a definitive proof of the waking state's permanence or truth. Both the dream state and the waking state are appearances within consciousness, and their reality is dependent on the perceiving consciousness itself. The exploration of these insights can lead us to a deeper understanding of the nature of reality and our own existence.

Answer — **Generic Non-RAG Bot**

Dreams can indeed feel incredibly real while we are experiencing them. However, it can be difficult to determine with absolute certainty whether we are currently dreaming or awake. One way to approach this question is by considering the concept of reality testing. Reality testing involves examining the characteristics of our current experience to determine whether it aligns with what we typically perceive as reality. There are several techniques that people use to reality test, such as checking the time, looking for inconsistencies, or attempting to perform actions that are impossible in dreams, like flying or walking through walls. In the case of your query, if you are unsure whether you are dreaming or awake, you can try reality testing. For example, you could check the time on a clock or a digital device. In dreams, the time often appears distorted or inconsistent. Another approach could be to look for inconsistencies in your surroundings or try to perform actions that are impossible in reality. It is important to note that while reality testing can be helpful, it is not foolproof. In some cases, dreams can be incredibly vivid and realistic, making it challenging to distinguish them from waking reality. Additionally, some individuals may experience conditions like lucid dreaming, where they are aware that they are dreaming and can exert some control over the dream. Ultimately, the guarantee that you are not currently dreaming while asking this question is difficult to provide. However, by employing reality testing techniques and reflecting on the consistency of your experience, you can increase your confidence in determining whether you are awake or dreaming.

Passages

20. Mandukya Upanishad | Chapter 2 Karika 6 | Swami Sarvapriyananda

main… Well, I mean, how would that… Yes. But let's take it in a very simple way. The person in the dream and every other thing that you see in the dream are actually not there. You are with me? Are actually not there. They are all your own mind. You are actually sleeping on the bed. Just follow the facts as they… as you know them to be. You are actually on your bed and sleeping and your mind generates all of this. Now compare it with what you actually experience in a dream, that you are in some place and you have a body and you are meeting with other people. But when you wake up, all of that is falsified. Why is it falsified? Because it has no existence apart from the mind which was dreaming it. Right? So that's the example. Those things are false because they were dreamt up by a mind. They did not exist apart from you which you… it felt like. It felt like they were things apart from me. Similarly, now he wants to prove in this waking world, it seems to be a real world apart from me. It seems to be real. But just like that those things do not… dream entities do not exist apart from the dreamer's mind, these waking entities also do not exist apart from consciousness. Proof, they cannot be experienced without consciousness. They are clearly in consciousness. Yes, I'll come to you. Yes. I'll come to you. Okay, don't go there now. Don't go there. Just look at it from the point of view of the witness consciousness. From the point of view of witness consciousness, this waking

4. Mandukya Upanishad | Chapter 1 Mantra 5-6 | Swami Sarvapriyananda

from every dream to dream it's different. Answer in reply, the dream state is a waking state for the dreamer. As has already been pointed out. And one knows a state to be a waking state only when there is a feeling that the objects seen are real and as such remain the same in all waking states. The answer is, in the dream state, in a dream when you see objects, they are not felt to be temporary and about to disappear at the end of this dream. They feel felt to be absolutely real. My friend feels real. The table, chair I'm sitting with feels real. The food I eat feels as real as necessary. They don't feel temporary. Just as these ones seem pretty permanent, the dream objects also seem pretty permanent in the dream. Ninth, if the objects of the waking state are exactly like those of the dream state, then our beloved friends, kith and kin, would be no more than ideas, like those of our dream world relatives. Such an attitude is repugnant to our feelings. The reply is that our relatives seen in the waking or the dream state, again a very subtle and powerful answer, are as real as the I, as you yourself. What you consider yourself as real as the waker considers himself to be, so real does the waker's world seem to be. As real as the dreamer considers himself or herself to be, so real does the dream world seem to be. Advaita's answer here is balanced and nuanced. Why? Because you know actually what Advaita wants to say is that, yes, your relatives are appearances. They are dreams

Figure 10: **RAG vs Non-RAG: Waking and dreaming.** The passages retrieved for the query are relevant but only offer a segment from a broader context. As stated earlier, this can be mitigated by using more advanced retrieval models that summarize longer contexts before input to the LLM. The RAG bot's response, while relevant, is lengthy and complex, making it challenging to comprehend. The generic bot's response although more concise and coherent, is completely general and has no relation to the Advaita Vedanta philosophy.

Category: Anecdotal

Query: Does Swami narrate any incident surrounding Shivaratri?

Keywords

Human-annotated: Shivaratri

Model ensemble: Shivaratri

Keyword-based RAG

Yes, in the lecture "Raja Yoga: The Path of Meditation (Part 1)," Swami Sarvapriyananda does share an incident related to the Hindu festival of Shivaratri. He talks about his own experience of fasting on Shivaratri for the first time when he became a monk. He explains that on this day, devotees of Lord Shiva observe a strict fast, which means no food or water from sunrise to sunrise the next day.

Swami Sarvapriyananda also narrates an incident where a mathematics teacher approached a Swami and asked if he could fast on Shivaratri. The Swami responded sarcastically, pointing out that even young children are able to fast on this day, implying that the teacher should not find it difficult. This story highlights the significance of Shivaratri and the spiritual strength and dedication it is associated with.

Source:

Raja Yoga: The Path of Meditation (Part 1) | Swami Sarvapriyananda

back and meditate Then think about that So for asceticism also that Remember Common sense So Shivaratri I remember fasting I never fasted earlier The first time I fasted Was when I became a monk On Shivaratri And So You get some fruits in the afternoon But otherwise no food, no breakfast, no lunch, no dinner Whole night puja Next day in the morning you get food But in the afternoon you get some fruits So I never ate so many fruits Not because I was hungry In anticipation of being hungry Maybe I will be hungry You won't be Not really So that's what happens That's the mind I still remember this poor young man He was a teacher of mathematics And He went to a Swami very proudly And he said Tomorrow is Shivaratri I will also fast Can I also fast And the Swami I still remember He said What kind of question is that 10 year old little children are going to fast Aren't you ashamed of asking such a question Am I going to fast or not Of course you are going to fast And you should

Standard RAG

In the provided passage, Swami Sarvapriyananda shares an incident involving Swami Shivamayanji, but it does not explicitly mention Shivaratri. However, the narrative does offer insights into the experiences and teachings of the swamis.

According to the passage, Swami Shivamayanji once told a group of young monks, including Swami Sarvapriyananda, about an experience he had with his guru, Swami Gambhirananda. Swami Gambhirananda, at the time a vice president and later president of the Ramakrishna Math and Ramakrishna Mission, was describing the nature of God realization. He explained that the living form of the chosen deity appears in one's heart as a blazing or flaming form (Bengali: Amin jakhoni chok bhoji. Ishta murti. Jaljal kare bhaveshe othe). Swami Gambhirananda further emphasized the importance of realizing Brahman in all beings, indicating a deep understanding and connection with the divine in everything and everyone.

Swami Shivamayanji used this story to teach the young monks about the expectations of people who show respect and reverence to them. He wanted to remind them of the importance of their conduct and the impact they have on others.

While this narrative does not directly mention Shivaratri, it does provide a glimpse into the teachings and values of the swamis, which likely influenced their celebrations and observations of various Hindu festivals, including Shivaratri.

Source:

79 Bhagavad Gita I Chapter 6 Verse 15 I Swami Sarvapriyananda

just talking about. He says not even an effort. Anytime I close my eyes. The presence the living form of the chosen deity. Is vivid and it's a living form. It's not an. Not imagined. Not just visualized. That blazes forth. Of course he didn't explain all this. He just said in Bengali. Bhagavan lab bolte tumi ki bojo. What do you mean? What exactly do you mean by God realization? He said in Bengali. Amin jakhoni chok bhoji. Ishta murti. Jaljal kare bhaveshe othe. Bengali. It would just simply mean. It's like a flaming form or a blazing form. Of the. Ishta Devata. Appears in my heart. Next he said. But is this all? One has to realize Brahman in all beings. Bengali he said. So imagine. That was Swami Gambhiranji. Long time before he actually became the vice president. And the president. Anyway Swami Shivamayanji told us this incident. And then he looked at us. The young monks present there. And he said. Look this is what people expect. When they bow down to you. And touch your feet. And call you know like this. Or show you so much respect. This is what people expect. This is what they want. In a moment to see in a monk. So never forget that. So it's really ingrained in my memory. That incident. I can still see him looking at straight at us. And telling us. He is a wonderful monk. Swami Shivamayanji. You have so many wonderful incidents to share about. Humorous, straightforward, very simple. One little incident shows the heart of the man. He must have been in the 60s at that

Figure 11: **Keyword-RAG vs Standard-RAG: Anecdotal.** The keyword-based retriever accurately retrieves the correct snippet, while the dense retriever confuses “Shivaratri” with “Shivamayananda”, possibly owing to similar embeddings and completely misses the context. As a result, the generated standard RAG answer is meaningless and lacks substance. The keyword-RAG model identifies the two key incidents relating to Shivaratri correctly and effectively summarizes them.

Category: Terminology

Query: What is the significance of the word 'Shraddha'?

Keywords

Human-annotated: Shraddha

Model ensemble: Shraddha, significance

Keyword-based RAG

The word 'Shraddha' in Swami Sarvapriyananda's teachings, particularly in the context of learning and understanding spiritual philosophies, holds a profound significance. It is often translated as 'faith' in English, but it goes beyond a simple belief or trust. Shraddha implies a deep sense of reverence, respect, and acceptance for the teacher, the teachings, and the texts. It is a feeling that there is truth in the teachings, even if one does not fully comprehend it yet.

Swami Sarvapriyananda emphasizes that Shraddha is essential for any meaningful investigation or learning to take place. It is the foundation that allows one to approach the subject with an open mind and heart, ready to understand and learn, rather than to judge, criticize, or dismiss.

In the absence of Shraddha, one might quickly jump to conclusions, judge the teacher or the text, and create barriers to understanding. With Shraddha, one approaches the teacher and the teachings with humility, acknowledging that they have something valuable to learn, and seeking to bridge the gap between their current understanding and the teachings.

In summary, Shraddha is a deep feeling of reverence, respect, and acceptance for the teacher, the teachings, and the texts, which forms the foundation for meaningful learning and understanding in spiritual philosophies.

Source:

123. Bhagavad Gita | Chapter 10 Verse 16-22 | Swami Sarvapriyananda

devotional approach where the idea is to cultivate love and faith in God. You cannot start by questioning there. It starts with Shraddha, with absolute acceptance. The existence of God, if you start by questioning it, then the path of devotion is not for you. Luckily in Hinduism, you have a spectrum of other paths. So if you don't accept the existence of God, it doesn't appeal to you or you question it, no problem. You can still have spirituality. You can have a yogic spirituality. You can have an Advaitic, the Jnani spirituality. These paths are open to you. In the Jnani's path, you must question. Clarity is very important at every step. And it's only through your questions that the teacher can teach. If Arjuna had no questions here, the Gita would be terribly short. Most of the Upanishads, you will see, they are dialogues. And the new teaching starts only when there is a question. So that's the point. The Buddhists have this terminology. Don't listen to a leaky pot, an upside down pot and a dirty pot. An upside down pot is that which does not retain anything. You try to pour water into it, just roll up the sides. So people come, listen. What did you hear? Oh, what do I know? All very high stuff, you know. The second one is the leaky pot.

So I come, I listen and I collect something. I feel I got something. But after some time, it's gone. I got something but I'm very vague about it now. I have to look into my notes and look at the YouTube talk again. It's a leaky pot. The third is a dirty pot. Dirty pot is already there's dirt inside. So whatever you put inside will be polluted. The skeptic, the one who is not aware of the world, whatever you put inside will be polluted. The skeptical approach in the sense that I know this is wrong. Now I'm trying to find out the ways in which this is wrong. One more little thing I'll add here. Something that I've noticed. See the traditional teachers insisted on Shraddha. And we don't really talk about that too much. Vivekananda gave great importance on Shraddha. Because in our modern approach, Shraddha means faith. A kind of deep sense of what this person says and what these texts say is true. But I don't get it yet. But it's true. There's something there. Let me go ahead. Without this, no investigation is possible. You can't sign up for a course in Columbia University if you don't have Shraddha on the syllabus and the text and the teacher.

If you say it's all wrong, why would you pay thousands of dollars and go to listen to something that's all wrong? So that kind of Shraddha is very necessary. And that's the least that we talk about. We don't talk about more than that. But much more than that is actually necessary. I'm seeing many different ways in which the deep Shraddha helps. And the lack of Shraddha obstructs us. Two examples I'll give you. Two levels. One level is, I don't have much Shraddha. And Shraddha in a sense is a deep feeling of reverence for the teacher and the material. But I'm curious. And I'm excited about this philosophy. The danger of that is, right at the beginning, I don't understand what's going on. But I really like it. It sounds cool. And because it's a matter of understanding, so very soon I begin to understand. If I really put some effort into it, if I read the books, listen to the talks, think about it, I begin to understand. Then I get a grasp, my own grasp, a foothold on the subject. And immediately, if there's a lack of Shraddha, what will happen is, I'll start judging the teacher and the text. Start finding faults with, you know, your understanding is not quite right. Let me correct you on this one. You might say, what's wrong with that? No, it's not. It's wrong.

Because the approach with Shraddha would have been, I don't understand what you said. This is my understanding. Could you help me from my understanding to what you have said? It's not only the polite way of approaching, it's the way with Shraddha. Without Shraddha, it's, I read it, I got it, and this person seems, now my teacher seems to be saying something different, so he's wrong. So, it's like, there's a story of an elderly person who was passing away, and his sons gathered around him. And he said, I have many people who borrowed money from me, and his sons quickly took down the notes. Father, tell us before you pass, whom did you borrow money from? Who borrowed money from you? We need to collect that.

Standard RAG

The word 'Shraddha' has a significant meaning in the context of the Mandukya Upanishad and Hindu traditions. It is a Sanskrit word that primarily refers to the funeral rites performed for departed ancestors. This term carries both psychological and spiritual connotations.

Psychologically, Shraddha signifies the process of letting go for the people who are left behind. It helps the living come to terms with the loss and separation from their loved ones.

On the spiritual side, Shraddha is performed to ensure that the deceased embarks on a higher path (gati) in their journey after leaving the body. The belief is that the prayers, worship, and rituals conducted on behalf of the departed can add to their store of good karma. This extra positive karma acts as a booster rocket, helping the deceased reach a better world or higher level in their spiritual evolution.

In summary, 'Shraddha' is a crucial concept in Hindu traditions, symbolizing both a psychological release and a spiritual boost for the departed ancestors, enhancing their journey and ensuring a higher path for them in the afterlife.

Source:

15. Mandukya Upanishad | Chapter 1 Mantra 12 & Karika 22-23 | Swami Sarvapriyananda

that you will remember Mandukya Upanishad suddenly and wake up with a cold sweat, but it will attract you. Whatever form of religion or spirituality, genuine religion or spirituality you find in your society, in your locality, it will pull you powerfully. It could be any form of spirituality. Basically they are all paths leading to the same reality. And it says that person will start practicing from wherever he or she left off helplessly. You cannot not walk on this path. You will walk on this path. And it says like this, this person goes on until he attains liberation. I'll come to you. What is the symbolism behind Shraddha? Shraddha is a Sanskrit word meaning the funeral rites which you perform for departed ancestors. One of course is psychological reason for letting go, for the people left behind. But the other is this one, for those who are transmigrating, having left the body, my father, mother, grandparents and so on and so forth. Let them have a higher gati, a higher path. They will go according to their karma. But your prayers and your worship and the rituals that you perform on their behalf adds to their store of good karma. It's a booster rocket you strap on so that they can go a little higher, maybe to a better world. So you pray for that. So that's what traditional Hindus believe in. But every culture you will see has prayers for the dead. They even talk in terms of helping to make the transition and so on and so forth. Once explained to me that time in the other

Figure 12: **Keyword-RAG vs Standard-RAG: Terminology.** The keyword-RAG model retrieves a comprehensive exposition on the concept of Shraddha, loosely translated as conviction, in the context of qualifications for the study of Advaita Vedanta. The standard RAG although retrieves a passage containing the word, it is however not directly related to what the questioner intends. This seems to be an unfortunate case of false positive for standard RAG due to inadequate or implied meaning in the query.

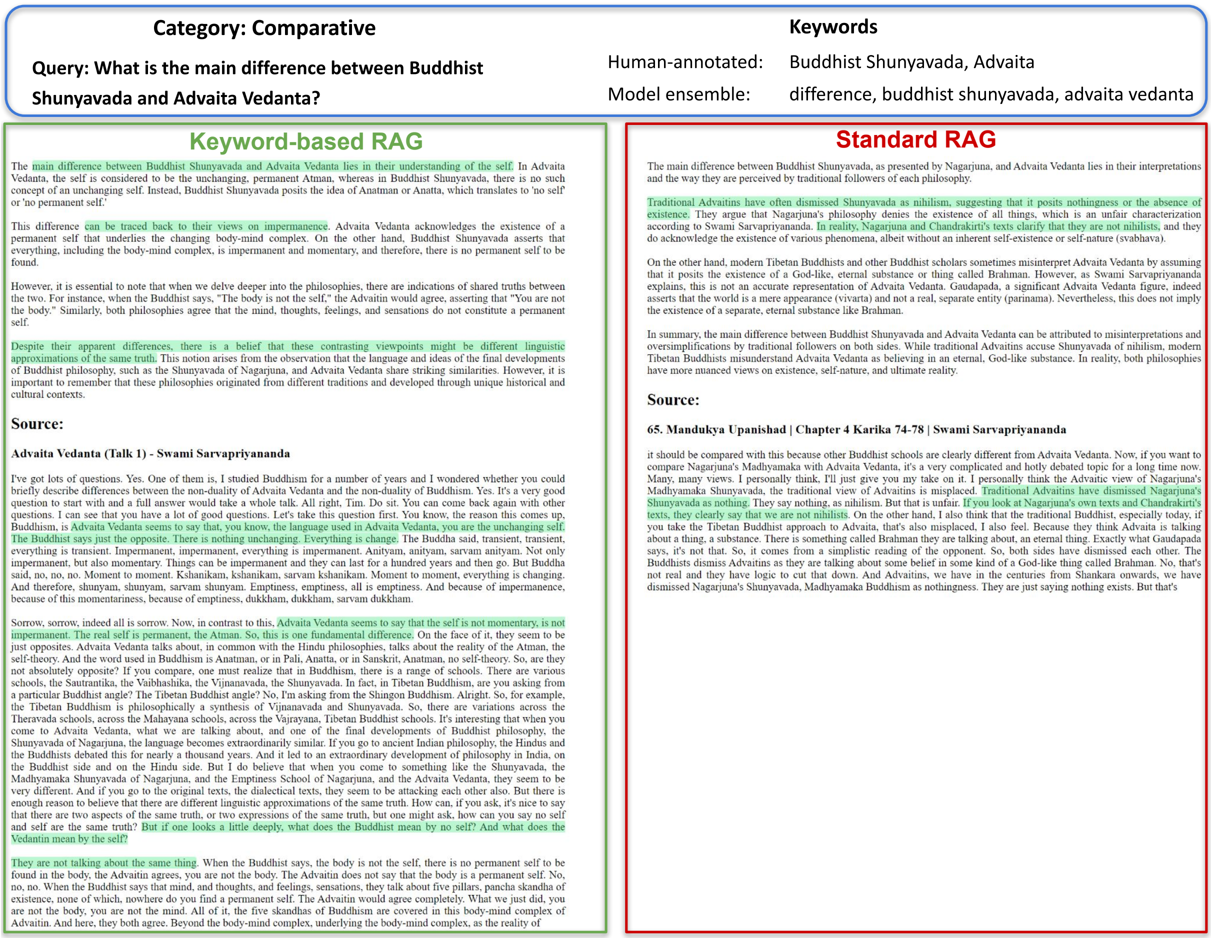

Figure 13: **Keyword-RAG vs Standard-RAG: Comparative.** This is a case where both models retrieve meaningful passages, with the keyword-based model providing greater context due to retrieval expansion. The standard model also seems to be using outside knowledge not mentioned in the passage.

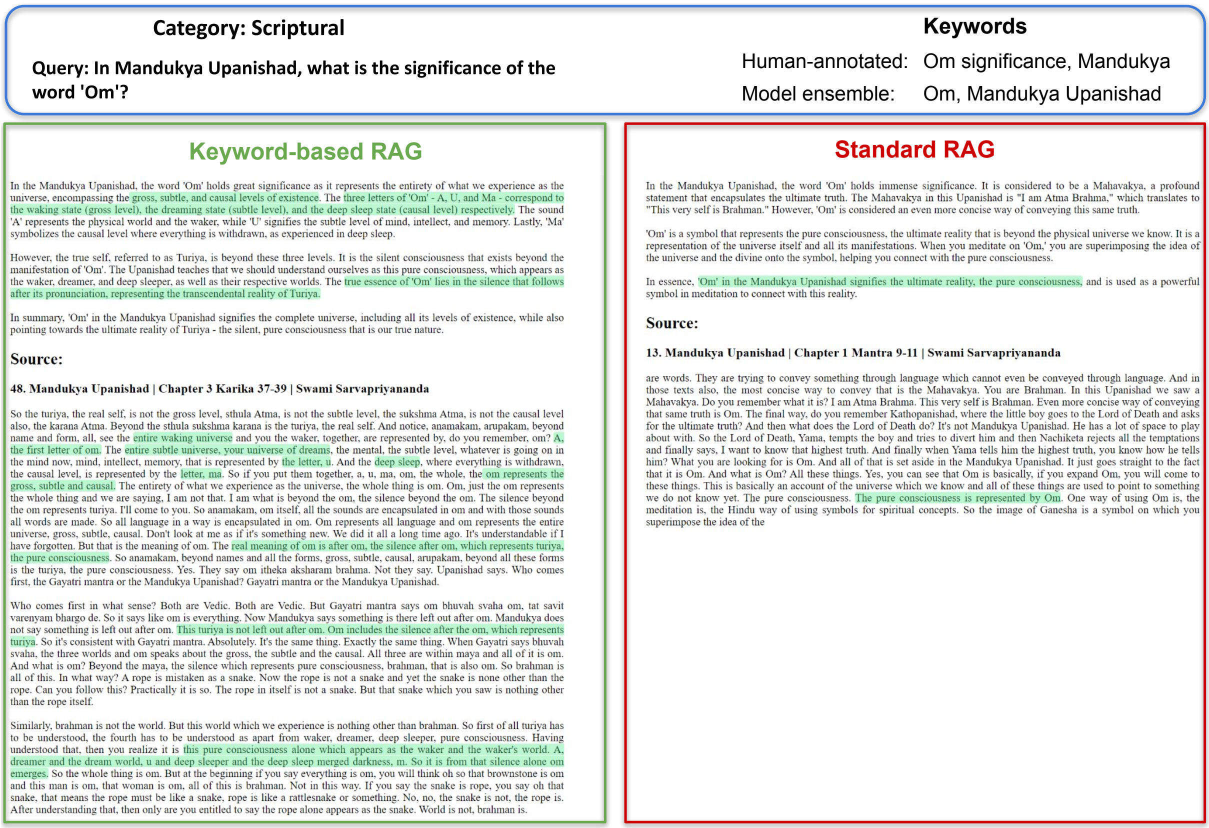

Figure 14: **Keyword-RAG vs Standard-RAG: Scriptural.** Both models retrieve from the correct scripture. However, while the retrieved passage in keyword-RAG is very comprehensive and accurately captures the context of the question, the standard RAG misses the point of the question and gives a generic answer.

Category: Reasoning

Query: Is the waking state similar to a dream or absolutely indistinguishable from the dream state?

Keywords

Human-annotated:	waking, dream, indistinguishable
Model ensemble:	indistinguishable, dream state, waking state similar

Keyword-based RAG

According to Swami Sarvapriyananda's lecture on Vedanta Retreat at Loon Lake, the waking state can be compared to a dream in the sense that both are temporary and illusory. The waking state is often referred to as the vyavaharika jiva, which is a transactional or relative empirical self that we experience ourselves to be. On the other hand, the dream state is referred to as the pratibhasika jiva, which is recognized after waking up to be an illusion or a dream.

However, there is a crucial difference between the two states. While the dream state is experienced as less real and temporary, the waking state is experienced as more real and permanent. But from the ultimate standpoint of Vedanta, both the waking and dream states are illusory and not the absolute reality.

When we wake up from a dream, the dream jiva, or the individual self in the dream, merges back into the waking jiva. Similarly, upon enlightenment, the waking jiva merges into the paramarthika jiva, which is the absolute or real self, also known as Brahman. This experience is said to be more powerful than waking up from a dream state.

In the Mandukya Upanishad, Gaudapada, Shankaracharya's teacher's teacher, states that we have only two states - dream and sleep. He considers the waking state to be a dream as well, and the reality to be beyond these states.

Therefore, while the waking state and the dream state share some similarities, they are not absolutely indistinguishable from each other. The waking state is experienced as more real and permanent than the dream state, but from the ultimate standpoint of Vedanta, both are illusory and not the absolute reality.

Source:

Vedanta Retreat at Loon Lake on Who am I by Swami Sarvapriyananda Lecture 11

your dream you in your dream is the dream jiva Jiva, individual You And Vedanta tells us we are Brahman actually So let's call that the real jiva This is Brahman Now the Sanskrit terms I'll just give you so that you can relate it to the book The book calls this the pratibhasika jiva Pratibhasika, illusion The book calls this the vyavaharika jiva Vyavaharika jiva And the dream calls this the paramarthika jiva Paramarthika, absolute Translation would be absolute jiva or real jiva Vyavaharika, transactional Transactional jiva Transactional, religious Transactional, religious Transactional, religious Transactional, religious Transactional, transactional jiva Transactional, relative, empirical what we experience ourselves to be And pratibhasika, what is recognized after waking up to be an illusion, to be a dream So the dream jiva Now he says these are the three things that we need to keep in mind And what happens is think about the experience of waking up from the dream Every day when we wake up from our dreams if you remember our dream and compare, remember the dream sometimes we do not remember the dreams but sometimes we do If you compare yourself in the dream and yourself sitting up on the bed after

waking up you will find that jiva was none other than you under dream or illusion And when you wake up that jiva merges back into you So upon waking up the dream jiva merges into the waking jiva And the point of the whole analysis is that upon enlightenment we, the waking jiva we merge into the paramarthika jiva, the real jiva That is Brahman That will be a similar experience Not exactly the same, much more powerful You will awaken into your own reality as you awaken from a dream state from a dream state into a waking state which is your own reality a greater reality In the same way we awaken from the so-called waking state into the really awake state which Buddha calls being awake In fact, in the Mandukya Upanishad Mandukya Karika Gaudapada Shankaracharya's teacher's teacher He says we have only two states dream and sleep We say we have three states, right? Waking Dreaming and sleeping Sleeping means deep sleep Of course, dream is also in sleep And we think we have these three states Deep sleep, dream and waking But Gaudapada Acharya, Shankaracharya's teacher's teacher says we have only two states Dream and sleep These two he calls dream And this he calls

sleep Only two states Only two states Dream and sleep Waking also he calls a dream And then, is there a real state when you get enlightened? He says yes, but that's not a state Reality is not a state States come and go Reality is what appears as these states So when we realize ourselves to be the Turiya The Brahman, the Paramarthika Jiva All of them are different terms for the same thing You, the reality After enlightenment Then you realize What we call the waking state is also a dream Just like the dreaming state And sleep is different from them Because in sleep we have no experience of any particular thing In dream we dream about things Waking we seem to experience with the help of sense organs But the reality of all of that is Brahman And we get enlightened, we realize that The holy mother Speaking to, I think, Arupananda Swami In Mayankathai, it's there She suddenly says Remember, Mahasarovar, he's saying In the rural setting in Bengal He's saying to this young monk Even this waking of ours is a dream And then the monk protests He says, how can this be a dream? Dreams are impermanent Every day it comes and it's gone Next time it's a new dream But the waking world is

there when you wake up And she doesn't argue She doesn't give arguments She just laughs, you know She laughs and she says in Bengali Taaholeva ba, tataaholeva Shakno bhoi toh no hai Let it be so, let it be more permanent The waking state, it is nothing other than a dream So, this analysis Now, one more point is there One example they have given for this Waking from

Standard RAG

The waking state is not absolutely indistinguishable from the dream state, as there are differences between the two. However, the dream state can be considered similar to a waking state for the dreamer. In both states, the objects and experiences felt by the individual are perceived as real and permanent within their respective contexts. The key distinction lies in the feeling of permanence and consistency of objects and experiences.

In the waking state, objects and experiences are generally consistent and remain the same from one instance to another, leading the individual to perceive them as real and permanent. On the other hand, in the dream state, the objects and experiences change from one dream to another, but they still feel real and permanent while the dream is occurring.

Swami Sarvapriyananda explains that Advaita Vedanta takes a balanced and nuanced approach to this question. While there are similarities between the waking and dream states, such as the perception of reality and permanence, there are also differences that set them apart. The perception of one's relatives in both states serves as an example of this nuanced view. In both the waking and dream states, relatives are perceived as real by the individual experiencing them. However, Advaita Vedanta does not diminish the reality of waking state relatives by comparing them to dream state relatives, which could be considered mere ideas or thoughts. Instead, it acknowledges the relative reality of both experiences, without fully equating the two states.

Source:

4. Mandukya Upanishad | Chapter 1 Mantra 5-6 | Swami Sarvapriyananda

different. Not only every night, from every dream to dream it's different. Answer in reply, the dream state is a waking state for the dreamer. As has already been pointed out. And one knows a state to be a waking state only when there is a feeling that the objects seen are real and as such remain the same in all waking states. The answer is, in the dream state, in a dream when you see objects, they are not felt to be temporary and about to disappear at the end of this dream. They feel felt to be absolutely real. My friend feels real. The table, chair I'm sitting with feels real. The food I eat feels as real as necessary. They don't feel temporary. Just as these ones seem pretty permanent, the dream objects also seem pretty permanent in the dream. Ninth, if the objects of the waking state are exactly like those of the dream state, then our beloved friends, kith and kin, would be no more than ideas, like those of our dream world relatives. Such an attitude is repugnant to our feelings. The reply is that our relatives seen in the waking or the dream state, again a very subtle and powerful answer, are as real as the I, as you yourself. What you consider yourself as real as the waker considers himself to be, so real does the waker's world seem to be. As real as the dreamer considers himself or herself to be, so real does the dream world seem to be. Advaita's answer here is balanced and nuanced. Why? Because you know actually what Advaita wants to say is that, yes, your relatives

Figure 15: **Keyword-RAG vs Standard-RAG: Reasoning.** The retrieved passage in keyword-RAG is technical and comprehensive and the generated answer effectively summarizes the main points. The standard model is also good, although the explanation is not as effective owing to the quality of retrieval.