Dynamo AI

openai-gpt-3.5-turbo-RAG

Retrieval Augmented Generation (RAG)

Hallucination Evaluation Report

Executive Summary

This report finds some high-level issues the evaluated RAG system may exhibit, for full coverage, see the RAG Workflow Analysis section:

> For **40%** of the data points, there may be generator issues: some response may contradict information from the retrieved documents. The generator may be basing the response on its base knowledge which may not be true according to the retrieved document. The question may be trivial or may not require new knowledge from the retrieved documents.

> For **37.5%** of the data points, there may be retriever & generator issues: the top document selected by the retriever is not relevant, and the generator may be ignoring information in the set of retrieved documents.

> For **12.5%** of the data points, there may be no significant issues

The RAG hallucination evaluation relies on retrieval relevance, response faithfulness and response relevance to assess the model's performance. Overall, Dynamo AI observed an average retrieval relevance rate of 0.525 (measures the relevance of the documents retrieved from the vector database). Dynamo AI observed an average response faithfulness rate of 0.275 (measures logical consistency between the retrieved contents and the final model response). Dynamo AI observed an average response relevance rate of 0.975 (measures the overall quality of the model response with respect to the initial query). This report provides a comprehensive analysis of findings from running the Retrieval Augmented Generation (RAG) Hallucination Evaluation on **openai-gpt-3.5-turbo-RAG**, along with strategies for hallucination mitigation.

Table of Contents

RAG Hallucination Evaluation Overview

This section describes the detailed methodology used to evaluate the degree of hallucination of openai-gpt-3.5-turbo-RAG at different points of the RAG workflow.

Objective

LLMs are broadly known to have 'hallucinations,' which we can define as the generation of inaccurate, nonsensical, or false content. Detecting hallucinations among model outputs is highly important - models that tend to hallucinate more are less reliable and factual. Specifically for Retrieval Augmented Generation (RAG) workflows, which rely on an external database to add context to an LLM, hallucinations can occur due to errors in either document retrieval or response generation. In-depth hallucination analysis related to each aspect of the RAG workflow can enable your team to pinpoint the cause of hallucinations and apply the right model improvements. Dynamo AI's RAG hallucination detection module specifically evaluates the tendency of openai-gpt-3.5-turbo-RAG to have intrinsic hallucinations or inconsistencies related to retrieval relevance, response faithfulness, and response relevance.
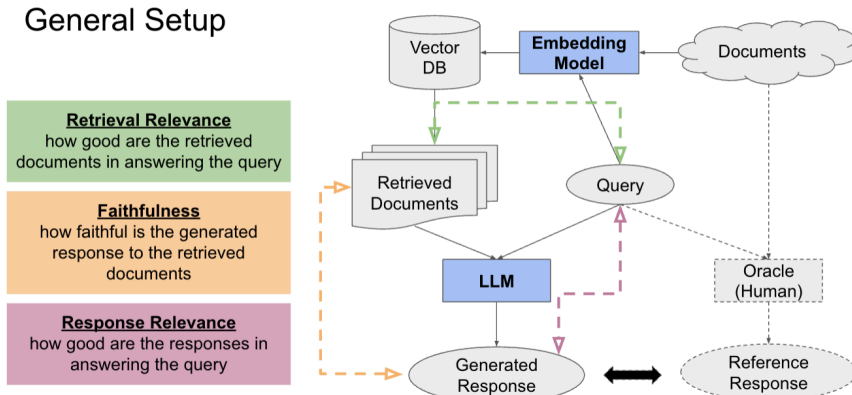
The evaluation was run 12 times, across 3 temperature(s), 2 sequence length(s), and 2 top k value(s), where temperature controls the "randomness" of the generator model, sequence length controls the number of output tokens of the generator model, and top k controls the number of documents retrieved from the vector database.

RAG Workflow

In general, there are three components in RAG: the query, the retrieved documents from a vector database, and the generated response from a language model with the query and the retrieved documents as context. Dynamo AI's RAG hallucination evaluation tool measures hallucination at three different components of the RAG workflow. Dynamo AI first measures retrieval relevance: this evaluates the relevancy of the retrieved documents to the initial query – this evaluates the quality of the chosen embedding model used for the vector database. Dynamo AI then measures response faithfulness: this evaluates the faithfulness of the generated response to the retrieved documents – this evaluates the performance of the LLM as well as the user prompt. In the case of evaluating openai-gpt-3.5-turbo-RAG, the following prompt was used: **Write how the Agent should respond to the User in the conversation, solely based on the context provided. Context: {context} Conversation: {question}**. Finally, Dynamo AI evaluates response relevance – the overall quality of the RAG workflow response to the original user query.

See diagram on the following page...

General Setup

Retrieval Relevance
how good are the retrieved documents in answering the query

Faithfulness
how faithful is the generated response to the retrieved documents

Response Relevance
how good are the responses in answering the query

Retrieval Relevance

Retrieval relevance represents the relevance of the documents retrieved from the vector database using the embedding model for each query. To measure the retrieval relevance of `openai-gpt-3.5-turbo-RAG`, Dynamo AI generates an LLM Relevance Score. Various studies have shown the effectiveness of LLMs as reference-free evaluators for tasks such as content relevance.

The LLM Relevance Score is computed by prompting an LLM to evaluate the sufficiency of the content in the retrieved documents to answer a given user query. For each query, DyanmoFL computes the relevance score against the top document retrieved from the vector database. Based on this score, each (query, retrieved document) pair will be classified as either positive (relevant) or negative (not relevant). A negative classification indicates that the retrieved document may not contain the key information required to answer the given query.

Response Faithfulness

Additional Resources: SummaC: Re-Visiting NLI-based Models for Inconsistency Detection in Summarization

https://arxiv.org/pdf/2111.09525.pdf

Response Faithfulness represents the faithfulness of model generated responses to the retrieved documents. To measure the response faithfulness of `openai-gpt-3.5-turbo-RAG`, Dynamo AI uses: Logical Consistency and Factuality.

To measure the faithfulness of `openai-gpt-3.5-turbo-RAG`, Dynamo AI uses a natural language inference (NLI) model, which labels a (document, summary) pair as either being entailing, contradicting, or neutral, with corresponding scores. Entailing indicates that the summary logically implies the content in the document while contradicting indicates otherwise, and neutral indicates that no logical relationship can be drawn between.

Dynamo AI specifically runs the NLI model over each combination of sentence pairs from a document and the generated response. This enables Dynamo AI to then retain the highest entailment score for a given generated response sentence. Finally, Dynamo AI takes the mean of all the maximum entailment scores for each sentence in the response to provide an aggregate score for the full response generated by `openai-gpt-3.5-turbo-RAG`. Based on this score, each (set of retrieved documents, response) will be classified as either positive (no issue) or negative (response potentially missing/contradicting some key information in the retrieved documents).

Response Relevance

Additional Resources: NLG Evaluation using
GPT-4 with Better Human Alignment

https://arxiv.org/pdf/2303.16634.pdf

Response Relevance represents the relevance of model generated responses to the query. To measure the response relevance **openai-gpt-3.5-turbo-RAG**, Dynamo AI generates a Response Relevance Score.

Similar to the Retrieval Relevance Score, the Response Relevance Score is computed by prompting an LLM to generate a score between 0 and 1 that indicates how relevant the generated response is to the question. Based on this score, each (query, response) will be classified as either positive (relevant) or negative (not relevant). A negative classification indicates that the response may not contain the information that answers the query.
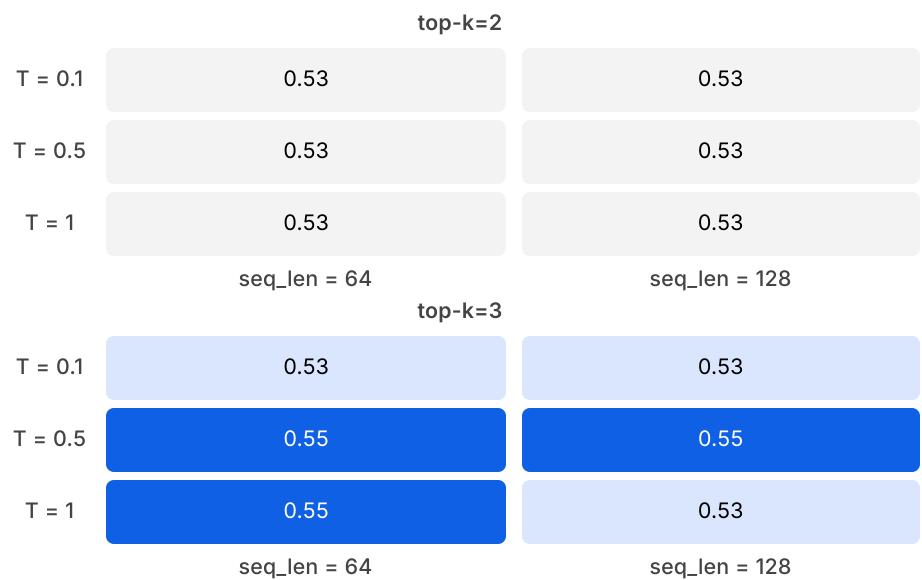
Detailed Findings

Retrieval Relevance Statistics

To understand the quality of the documents retrieved by openai-gpt-3.5-turbo-RAG, Dynamo AI measures success rate by flagging all entries as either positive or negative. The overall relevance rate is the percentage of entries with a positive classification. At a temperature of 0.1, Dynamo AI observed a relevance rate of 0.525, meanwhile, at a temperature of 1, Dynamo AI observed a relevance rate of 0.525. The chart below displays the relevance rate for each temperature evaluated.

Figure 1: Retrieval Relevance Rate by model temperature

Model temperature controls the relative randomness of outputs, we generally expect a higher temperature to lead to a decrease in average Retrieval Relevance.

top-k=2

	seq_len = 64	seq_len = 128
T = 0.1	0.53	0.53
T = 0.5	0.53	0.53
T = 1	0.53	0.53

top-k=3

	seq_len = 64	seq_len = 128
T = 0.1	0.53	0.53
T = 0.5	0.55	0.55
T = 1	0.55	0.53

Response Faithfulness Statistics

To understand the logical consistency of openai-gpt-3.5-turbo-RAG responses to the retrieved documents, Dynamo AI measures the faithfulness rate by flagging all entries as either positive or negative, based on the entailment probability computed with NLI models. The faithfulness rate is the percentage of entries with a positive classification.

At a temperature of 0.5, sequence length of 64, and top k of 2, Dynamo AI observed a faithfulness rate of 0.15, meanwhile, at a temperature of 0.5, sequence length of 128, and top k of 3, Dynamo AI observed a faithfulness rate of 0.3. The chart below displays the success rate for each temperature, sequence length, and top-k evaluated.

Figure 2: Faithfulness Rate by model temperature

Model temperature controls the relative randomness of outputs, we generally expect a higher temperature to lead to a decrease in average Response Faithfulness.
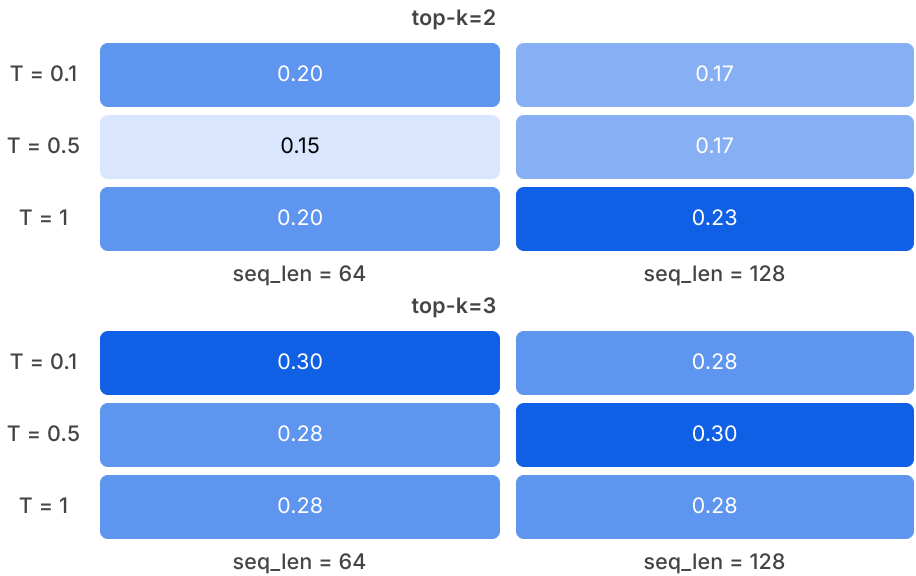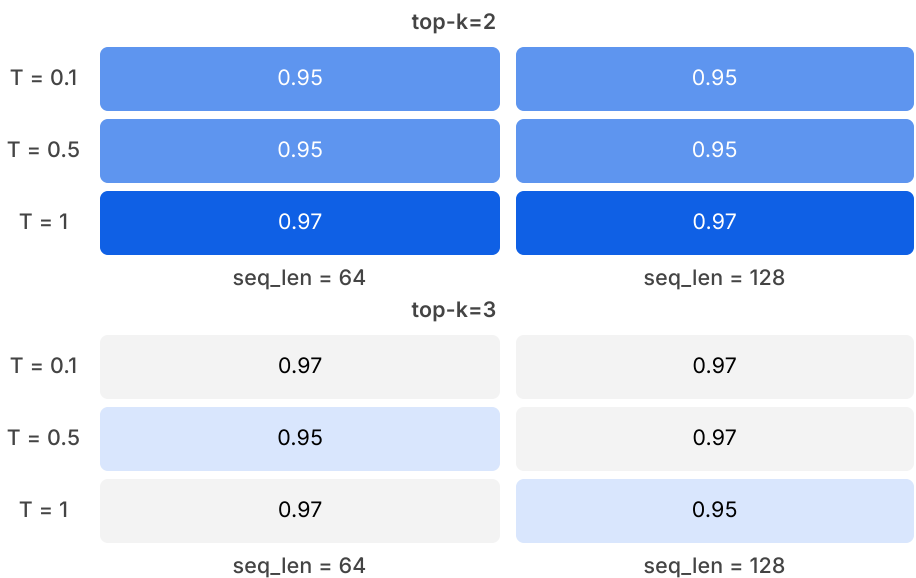
top-k=2

	seq_len = 64	seq_len = 128
T = 0.1	0.20	0.17
T = 0.5	0.15	0.17
T = 1	0.20	0.23

top-k=3

	seq_len = 64	seq_len = 128
T = 0.1	0.30	0.28
T = 0.5	0.28	0.30
T = 1	0.28	0.28

Response Relevance Statistics

To understand the quality of the responses generated by **openai-gpt-3.5-turbo-RAG**, Dynamo AI measures relevance rate by flagging all entries as either positive or negative based on response relevance scores. The relevance rate is the percentage of entries with a positive classification.

At a temperature of 1, sequence length of 128, and top k of 3, Dynamo AI observed a relevance rate of 0.95, meanwhile, at a temperature of 1, sequence length of 64, and top k of 2, Dynamo AI observed a relevance rate of 0.975. The chart below displays the success rate for each temperature evaluated.

Figure 3: Response Relevance Rate by model temperature

Model temperature controls the relative randomness of outputs, we generally expect a higher temperature to lead to a decrease in average Response Relevance.

top-k=2

	seq_len = 64	seq_len = 128
T = 0.1	0.95	0.95
T = 0.5	0.95	0.95
T = 1	0.97	0.97

top-k=3

	seq_len = 64	seq_len = 128
T = 0.1	0.97	0.97
T = 0.5	0.95	0.97
T = 1	0.97	0.95

Topical Analysis

To provide deeper insight into RAG hallucination, Dynamo AI conducted further analysis to determine the best and worst performing topics on each score for **openai-gpt-3.5-turbo-RAG**. Dynamo AI observed that based on the retrieval relevance rates, **openai-gpt-3.5-turbo-RAG** tended to hallucinate the most on the following topics: **student scholarship**. Based on response faithfulness rates, **openai-gpt-3.5-turbo-RAG** tended to hallucinate the most on the following topics: **student scholarship**.Based on response relevance rates, **openai-gpt-3.5-turbo-RAG** tended to hallucinate the most on the following topics: **others (undefined)**. The charts below display average success rates by topic.

The topics used for the analysis were the following:

Topic	# of queries
driver license registration	7
student scholarship	1
eligibility and benefits for the disabled	10
social security services	9
others (undefined)	13

Dynamo AI observed that based on the retrieval relevance scores, **openai-gpt-3.5-turbo-RAG** tended to hallucinate the most on the following topics: **student scholarship**. Based on response faithfulness, **openai-gpt-3.5-turbo-RAG** tended to hallucinate the most on the following topics: **student scholarship**. Based on response relevance, **openai-gpt-3.5-turbo-RAG** tended to hallucinate the most on the following topics: **others (undefined)**.

Figure 4: Retrieval Relevance Rate by model temperature and topic

Here, we see average Retrieval Relevance score per temperature by topic categories in the testing dataset

driver license registration

	seq_len = 64	seq_len = 128
T = 0.1	0.71	0.71
T = 0.5	0.71	0.71
T = 1	0.63	0.63

student scholarship

	seq_len = 64	seq_len = 128
T = 0.1	0	0
T = 0.5	0	0
T = 1	0	0

eligibility and benefits for the disabled

	seq_len = 64	seq_len = 128
T = 0.1	0.60	0.60
T = 0.5	0.60	0.60
T = 1	0.73	0.64

social security services

	seq_len = 64	seq_len = 128
T = 0.1	0.22	0.22
T = 0.5	0.22	0.22
T = 1	0.27	0.27

others (undefined)

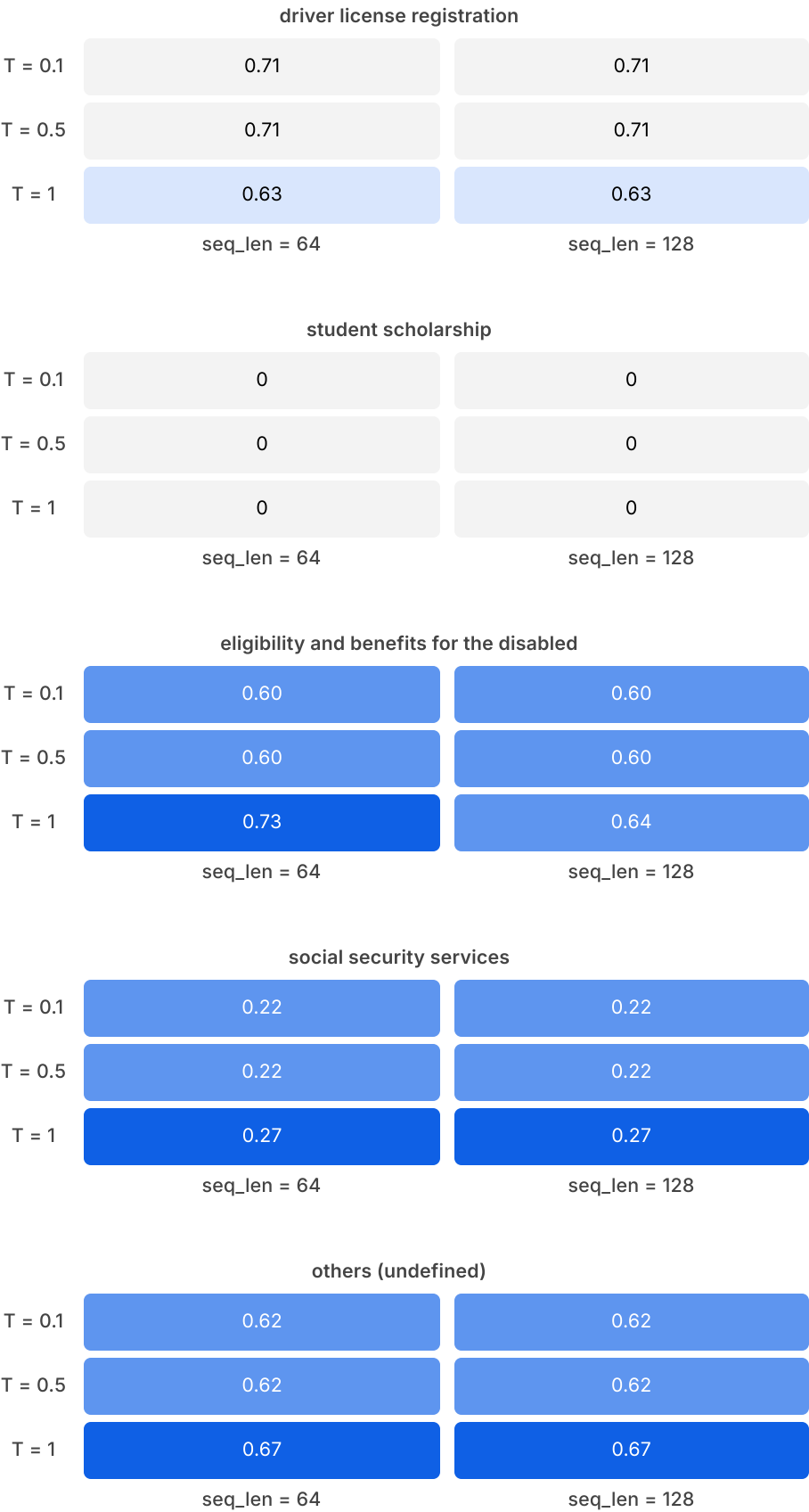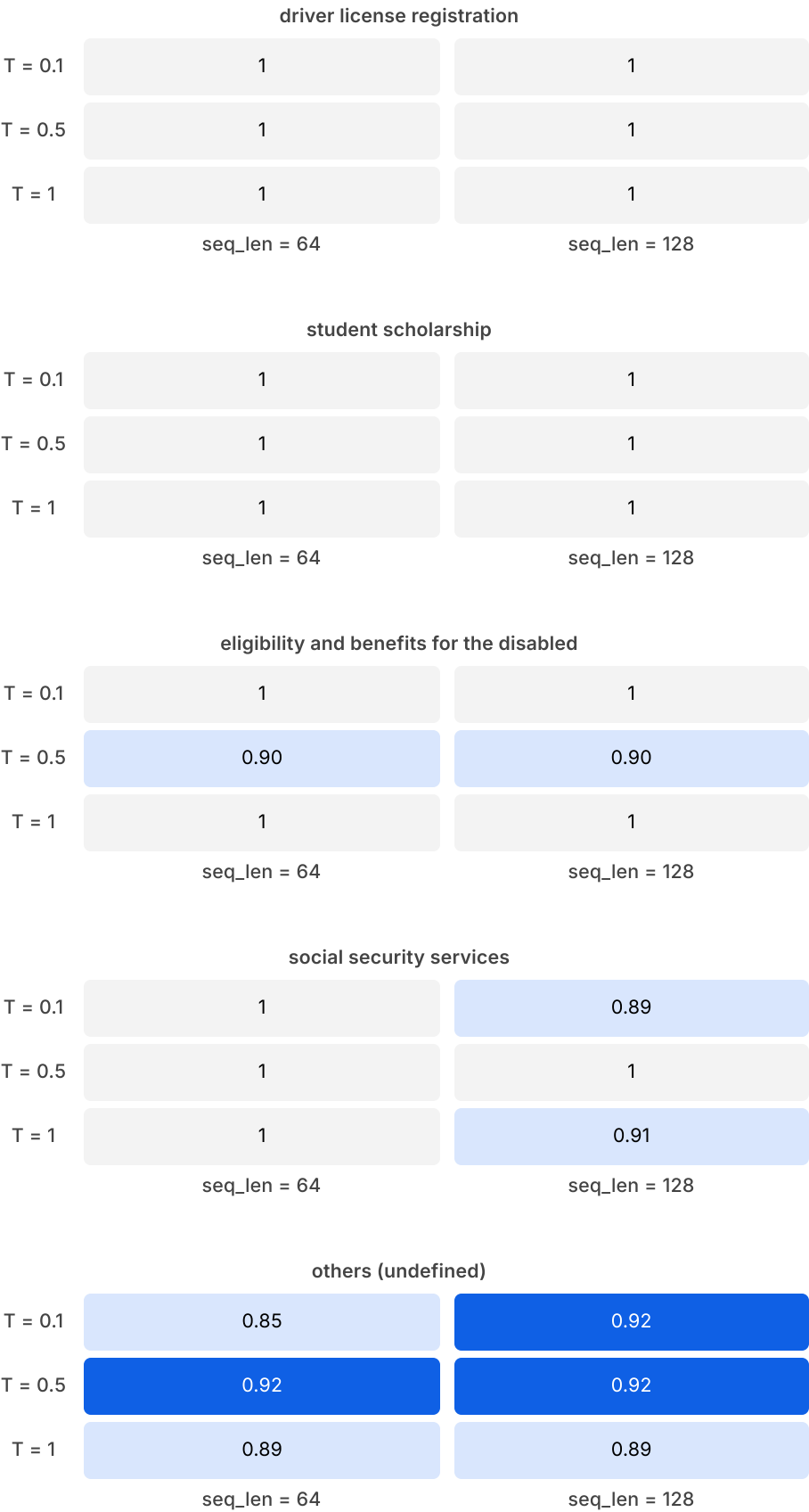|	seq_len = 64	seq_len = 128
T = 0.1	0.62	0.62
T = 0.5	0.62	0.62
T = 1	0.67	0.67

Figure 5: Response Faithfulness Rate by model temperature and topic

Here, we see average Response Faithfulness score per temperature by topic categories in the testing dataset

driver license registration

	seq_len = 64	seq_len = 128
T = 0.1	0.29	0.29
T = 0.5	0.43	0.29
T = 1	0.50	0.25

student scholarship

	seq_len = 64	seq_len = 128
T = 0.1	1	1
T = 0.5	1	1
T = 1	0	1

eligibility and benefits for the disabled

	seq_len = 64	seq_len = 128
T = 0.1	0.10	0
T = 0.5	0	0
T = 1	0	0

social security services

	seq_len = 64	seq_len = 128
T = 0.1	0.22	0.22
T = 0.5	0.22	0.22
T = 1	0.45	0.55

others (undefined)

	seq_len = 64	seq_len = 128
T = 0.1	0.15	0.15
T = 0.5	0	0.15
T = 1	0.22	0.22

Figure 6: Response Relevance Rate by model temperature and topic

Here, we see average Response Relevance rate per temperature by topic categories in the testing dataset

driver license registration

	seq_len = 64	seq_len = 128
T = 0.1	1	1
T = 0.5	1	1
T = 1	1	1

student scholarship

	seq_len = 64	seq_len = 128
T = 0.1	1	1
T = 0.5	1	1
T = 1	1	1

eligibility and benefits for the disabled

	seq_len = 64	seq_len = 128
T = 0.1	1	1
T = 0.5	0.90	0.90
T = 1	1	1

social security services

	seq_len = 64	seq_len = 128
T = 0.1	1	0.89
T = 0.5	1	1
T = 1	1	0.91

others (undefined)

	seq_len = 64	seq_len = 128
T = 0.1	0.85	0.92
T = 0.5	0.92	0.92
T = 1	0.89	0.89

RAG Workflow Analysis

Based on the retrieval relevance, response faithfulness, and response relevance observed, Dynamo AI provides further insights into the potential pitfalls of the existing RAG system. The table below describes a set of potential improvement areas within a general RAG system, as well as statistics regarding how many entries fall into each case. Based on the evaluation, the system improvements that will have the highest impact fall into the following category: **generator issues: some response may contradict information from the retrieved documents. The generator may be basing the response on its base knowledge which may not be true according to the retrieved document. The question may be trivial or may not require new knowledge from the retrieved documents..** This is determined based on assessing the case with the maximum number of associated datapoints.

Figure 7: Areas of Improvement

RETR: Retrieval Relevance. FTFL: Faithfulness.
RESP: Response Relevance.

Issue Type	RETR	FTFL	RESP	# Datapoints
No issue	Pass	Pass	Pass	5
Generator issue: the response may be following the information in the retrieved documents, but may be tangential to the question.	Pass	Pass	Fail	0
Generator issue: response may contradict information from the retrieved documents. The generator may be basing the response on its base knowledge which may not be true (e.g., the question may be trivial or may not require new knowledge from the retrieved docs)	Pass	Fail	Pass	16
Generator issue: response may contradict information from the retrieved docs. This may lead to a less relevant response.	Pass	Fail	Fail	0
Retriever & Generator issue: the top document selected by the retriever is not relevant, and the generator may be ignoring information in the set of retrieved documents.	Fail	Fail	Pass	15
Retriever issue: the top document selected by the retriever may not be relevant, and the response may be based on this less relevant document.	Fail	Pass	Fail	0
Retriever issue: the top document selected by the retriever may not be relevant, but other documents retrieved may be.	Fail	Pass	Pass	3
Retriever & Generator issue.	Fail	Fail	Fail	1

Usage Considerations

Importance of Hallucination Detection & Mitigation for RAG Systems

Improving model performance and reducing hallucinations is key to improving the utility and reliability of your model. Retrieval Augmented Generation (RAG) is often used as a technique to ground models to an existing database of knowledge and reduce hallucination. However, even systems with RAG can still be prone to hallucinations, as it consists of multiple components of language modeling. To improve the performance of RAG, state-of-the-art research suggests testing different retrievers, embedding models, and generators. This may include increasing model scale (parameter count), using high-quality and diverse training datasets for vector databases, changing prompts, or adjusting model hyperparameters such as 'temperature' and response length. Research also suggests that leveraging different types of verifications during inference, such as via prompting, sampling, or uncertainty estimation, can improve the quality of the generated response.

Limitations

While retrieval relevance, response relevance, and response faithfulness scores can provide a great starting point for evaluating different components of RAG workflow, it is important to note that all these scores depend on other language models. The evaluation results therefore may depend on the quality of the models used for evaluation. Additional qualitative inspection is recommended.

About Dynamo AI

Dynamo AI is the world's leading enterprise solution for privacy-preserving Generative AI. At Dynamo AI we believe that prioritizing privacy, compliance, and data security from day 1 while building Generative AI applications is the only way to responsibly scale AI and use it to augment human potential beyond what was thought possible. Our proprietary technology encapsulates state-of-the-art optimization techniques for training and deploying Generative AI models along with a robust privacy training and evaluation suite incorporating paradigms like Federated Learning and Differential privacy to bring high-performing end-to-end plug-and-play Generative AI to global enterprises.

Disclaimer and Terms of Service

The metrics and recommendations presented within this report are intended solely for informational purposes and should not be construed as legal advice. While we have undertaken diligent efforts to ensure the accuracy and reliability of the data and analyses contained herein, they are subject to inherent limitations and may not encompass all potential vulnerabilities or scenarios.

The tests performed and their resulting issues are only from the point of view of Dynamo AI. The risks of developing, deploying, implementing, managing and/or deprecating an AI model are specific to your organization and the circumstances in which you intend to deploy and use that model, and you should not rely on this report in isolation to reach a decision on such risks.

Any actions taken based on the information provided in this report are at your own discretion and risk. We disclaim any liability for or in connection with the consequences of any decisions or actions taken by your organization in reliance upon the information presented in this report.

Appendix 1: Benchmarking Examples

Here Dynamo AI provides reference examples from the evaluation to exemplify various metric scores.

Exhibit 1: Example Reference text & model output pairs for benchmarking

Examples from the evaluation dataset with model outputs are displayed to exemplify metric scoring.

Temperature: 1

Query: User says " Yes I am currently not receiving benefits" Agent says " you don t get Social Security or SSI benefits yet but are planning to apply?" User says " not" Agent says " you already have an account?" User says " yes"

Generated Output:
Agent: Great, it sounds like you already have an account set up. Let's make sure you're receiving the benefits you're eligible for. Can you please provide me with your account details so I can review your information?

Retrieved document id's: 9526ba0e-e195-4de4-8435-050cd93a4994, cda42cce-e092-4fe4-b546-adb841210990

Faithfulness: 0.0333

Retrieval Relevance: 0

Response Relevance: 0.5

Temperature: 1

Query: User says " what is disability?"

Generated Output:
Agent Response: A disability is a physical, cognitive, or mental impairment that substantially limits one or more major life activities. It can affect a person's mobility, communication, sensory perception, or other essential functions.

Retrieved document id's: 37e88c37-b1eb-444b-8fc0-f8714a265c54, e9322f55-fdfd-4014-9982-eecbb6772dba

Faithfulness: 0.15699

Retrieval Relevance: 0.3

Response Relevance: 1

Temperature: 1

Query: User says " I just wanted to know if there are any resources and/or support outside VA which I could access."

Generated Output:
Agent: Absolutely, there are resources and support options available outside of the VA that you can access. Your VA care team can provide you with information on community resources, organizations, and healthcare providers that may be able to provide additional support. Would you like me to connect you with someone on your care team who can help you

Retrieved document id's: 3c68d6e5-3442-4297-b827-821d8b6ffba3, 60ffdbcd-e527-4829-a58d-36c0808fd8cd

Faithfulness: 0.20002

Retrieval Relevance: 1

Response Relevance: 0.9

Temperature: 0.5

Query: User says " Hello there. My daughter is 17 years old and I came here to find out what she needs to consider, being an underage driver."

Generated Output:
Agent: Hello! It's great that you're being proactive about your daughter's driving. In most countries, the legal age for obtaining a driver's license is 18. However, there may be exceptions or additional requirements for underage drivers, depending on the specific regulations of your country. I recommend reaching out to your local Department of Motor Vehicles or equivalent authority to get accurate information on the process and requirements for underage drivers. It's important to ensure that your daughter meets all necessary criteria before she begins driving.

Retrieved document id's: 4469bacc-3053-4710-bcc4-2be759e4043a, a1b61eb6-818c-434a-83fd-3983330efdb4

Faithfulness: 0.00877

Retrieval Relevance: 0.2

Response Relevance: 0.9

Temperature: 0.5

Query: User says " Hello i would like to get information about Program Reviews"

Generated Output:
Agent: Hello! I can definitely provide you with information about Program Reviews. What specific details are you looking for?

Retrieved document id's: 7342fd17-9275-4f34-8a8d-40dcf5f75e6e, 867e8dff-b9ac-429d-a662-abce9ad70c53

Faithfulness: 0.1141

Retrieval Relevance: 0.3

Response Relevance: 1

Temperature: 0.5

Query: User says " I want to become a driving school instructor and need to know what tests are involved in the process."

Generated Output:
Agent: "To become a driving school instructor, you will need to pass the Road Sign Test, Written Test, and the Behind-The-Wheel Instructor's Driving Test. Each test has specific requirements and you are allowed to take them multiple times. If you do not pass, you will need to submit a new application. Do you have any other questions about the process?"

Retrieved document id's: 0ad20ab2-1f29-4f93-8915-84ae2fb23c6f, 4f37dae7-5307-4430-ac95-ec82088d0dc2

Faithfulness: 0.61203

Retrieval Relevance: 1

Response Relevance: 1

Temperature: 0.1

Query: User says " Yes I am currently not receiving benefits" Agent says " you don t get Social Security or SSI benefits yet but are planning to apply?" User says " not" Agent says " you already have an account?" User says " yes"

Generated Output:
Agent: "Great! Since you already have an account, you can go ahead and apply for retirement benefits by selecting the 'Apply for Retirement Benefits' button above. If you have any questions or need assistance, feel free to ask."

Retrieved document id's: 9526ba0e-e195-4de4-8435-050cd93a4994, cda42cce-e092-4fe4-b546-adb841210990, bab4e5a6-68b0-404a-807f-40587f2fadaa

Faithfulness: 0.57432

Retrieval Relevance: 0

Response Relevance: 0.5

Temperature: 0.1

Query: User says " what is disability?"

Generated Output:
Agent: Disability is a physical, cognitive, sensory, or mental condition that may affect a person's ability to perform daily activities. It can vary greatly from person to person and

may require accommodations or support to ensure equal opportunities and access to resources. If you need assistance or more information related to your disability, please let me

Retrieved document id's: 37e88c37-b1eb-444b-8fc0-f8714a265c54, 10153808-5183-4c43-b562-470227ee01bf, e9322f55-fdfd-4014-9982-eecbb6772dba

Faithfulness: 0.25455

Retrieval Relevance: 0.3

Response Relevance: 1

Temperature: 0.1

Query: User says " I just wanted to know if there are any resources and/or support outside VA which I could access."

Generated Output:
Agent: Yes, there are resources and support available outside of the VA that you can access. Your VA care team can provide you with more information on what options are available to you. Would you like me to connect you with someone from your care team to discuss this further?

Retrieved document id's: 3c68d6e5-3442-4297-b827-821d8b6ffba3, 60ffdbcd-e527-4829-a58d-36c0808fd8cd, ed412b99-e184-4cd8-a2e1-e88889933c81

Faithfulness: 0.13827

Retrieval Relevance: 1

Response Relevance: 0.9

Temperature:

Query: User says " Yes I am currently not receiving benefits" Agent says " you don t get Social Security or SSI benefits yet but are planning to apply?" User says " not" Agent says " you already have an account?" User says " yes"

Generated Output:
Agent: Great! Since you are already receiving benefits and have an account, you can proceed to the online application by selecting the "Apply for Retirement Benefits" button above. If you have any questions or need assistance along the way, feel free to reach out for help.

Retrieved document id's: 9526ba0e-e195-4de4-8435-050cd93a4994, cda42cce-e092-4fe4-b546-adb841210990, bab4e5a6-68b0-404a-807f-40587f2fadaa

Faithfulness: 0.35474

Retrieval Relevance: 0

Response Relevance: 0

Temperature:

Query: User says " what is disability?"

Generated Output:
Agent: Disability is a physical, mental, cognitive, or developmental condition that limits a person's ability to perform certain everyday activities. It can vary greatly in severity and impact on an individual's life.

Retrieved document id's: 37e88c37-b1eb-444b-8fc0-f8714a265c54, 10153808-5183-4c43-b562-470227ee01bf, e9322f55-fdfd-4014-9982-eecbb6772dba

Faithfulness: 0.28165

Retrieval Relevance: 0.3

Response Relevance: 1

Temperature:

Query: User says " I just wanted to know if there are any resources and/or support outside VA which I could access."

Generated Output:
Agent: Yes, there are resources and support available outside of the VA that you can access. It's important to maintain a well-rounded care team, so let's discuss more about your VA care team and how we can connect you with additional support if needed.

Retrieved document id's: 3c68d6e5-3442-4297-b827-821d8b6ffba3, 60ffdbcd-e527-4829-a58d-36c0808fd8cd, ed412b99-e184-4cd8-a2e1-e88889933c81

Faithfulness: 0.20783

Retrieval Relevance: 1

Response Relevance: 1

Temperature:

Query: User says " Yes I am currently not receiving benefits" Agent says " you don t get Social Security or SSI benefits yet but are planning to apply?" User says " not" Agent says " you already have an account?" User says " yes"

Generated Output:
Agent: "Great! Since you already have an account, you can go ahead and apply for retirement benefits by selecting the 'Apply for Retirement Benefits' button above. If you have any questions or need assistance, feel free to ask."

Retrieved document id's: 9526ba0e-e195-4de4-8435-050cd93a4994, cda42cce-e092-4fe4-b546-adb841210990, bab4e5a6-68b0-404a-807f-40587f2fadaa

Faithfulness: 0.57432

Retrieval Relevance: 0

Response Relevance: 0.5

Temperature:

Query: User says " what is disability?"

Generated Output:
Agent: Disability is a physical, cognitive, sensory, or mental condition that may affect a person's ability to perform daily activities. It can vary greatly from person to person and may require accommodations or support to ensure equal opportunities and access to resources. If you need assistance or more information related to your disability, please let me

Retrieved document id's: 37e88c37-b1eb-444b-8fc0-f8714a265c54, 10153808-5183-4c43-b562-470227ee01bf, e9322f55-fdfd-4014-9982-eecbb6772dba

Faithfulness: 0.25455

Retrieval Relevance: 0.3

Response Relevance: 1

Temperature:

Query: User says " I just wanted to know if there are any resources and/or support outside VA which I could access."

Generated Output:
Agent: Yes, there are resources and support available outside of the VA that you can access. Your VA care team can provide you with more information on what options are available to you. Would you like me to connect you with someone from your care team to discuss this further?

Retrieved document id's: 3c68d6e5-3442-4297-b827-821d8b6ffba3, 60ffdbcd-e527-4829-a58d-36c0808fd8cd, ed412b99-e184-4cd8-a2e1-e88889933c81

Faithfulness: 0.13827

Retrieval Relevance: 1

Response Relevance: 0.9

Temperature:

Query: User says " Yes I am currently not receiving benefits" Agent says " you don t get Social Security or SSI benefits yet but are planning to apply?" User says " not" Agent says " you already have an account?" User says " yes"

Generated Output:
Agent: Great, since you already have an account, I can help you with the application process for Social Security or SSI benefits. Let's get started. Can you please provide me with your account information?

Retrieved document id's: 9526ba0e-e195-4de4-8435-050cd93a4994, cda42cce-e092-4fe4-b546-adb841210990

Faithfulness: 0.01339

Retrieval Relevance: 0

Response Relevance: 0.6

Temperature:

Query: User says " what is disability?"

Generated Output:
Agent: Disability is a physical, mental, or emotional condition that limits a person's ability to perform certain tasks or participate in daily activities.

Retrieved document id's: 37e88c37-b1eb-444b-8fc0-f8714a265c54, e9322f55-fdfd-4014-9982-eecbb6772dba

Faithfulness: 0.3215

Retrieval Relevance: 0.3

Response Relevance: 1

Temperature:

Query: User says " I just wanted to know if there are any resources and/or support outside VA which I could access."

Generated Output:
Agent: Absolutely, there are resources and support available outside of the VA that you can access. Your VA care team can provide you with information on community resources, support groups, and other services that may be beneficial to you. Feel free to reach out to your care team for more information on how to access these resources.

Retrieved document id's: 3c68d6e5-3442-4297-b827-821d8b6ffba3, 60ffdbcd-e527-4829-a58d-36c0808fd8cd

Faithfulness: 0.14367

Retrieval Relevance: 1

Response Relevance: 0.9

Temperature:

Query: User says " Yes I am currently not receiving benefits" Agent says " you don t get Social Security or SSI benefits yet but are planning to apply?" User says " not" Agent says " you already have an account?" User says " yes"

Generated Output:
Agent: Great! Since you already have an account, you can go ahead and apply for retirement benefits by selecting the "Apply for Retirement Benefits" button above. If you have any questions or need assistance, feel free to let me know.

Retrieved document id's: 9526ba0e-e195-4de4-8435-050cd93a4994, cda42cce-e092-4fe4-b546-adb841210990, bab4e5a6-68b0-404a-807f-40587f2fadaa

Faithfulness: 0.59877

Retrieval Relevance: 0

Response Relevance: 0.5

Temperature:

Query: User says " what is disability?"

Generated Output:
Agent: Disability is a physical, mental, or cognitive condition that limits a person's ability to perform certain tasks or activities. It can vary in severity and can be temporary or permanent.

Retrieved document id's: 37e88c37-b1eb-444b-8fc0-f8714a265c54, 10153808-5183-4c43-b562-470227ee01bf, e9322f55-fdfd-4014-9982-eecbb6772dba

Faithfulness: 0.26284

Retrieval Relevance: 0.3

Response Relevance: 1

Temperature:

Query: User says " I just wanted to know if there are any resources and/or support outside VA which I could access."

Generated Output:
Agent: Yes, there are resources and support available outside of the VA that you may be able to access. Your VA care team can provide you with more information on what

options are available to you. Would you like me to connect you with a member of your care team to discuss this further?

Retrieved document id's: 3c68d6e5-3442-4297-b827-821d8b6ffba3, 60ffdbcd-e527-4829-a58d-36c0808fd8cd, ed412b99-e184-4cd8-a2e1-e88889933c81

Faithfulness: 0.07404

Retrieval Relevance: 1

Response Relevance: 0.9

Temperature:

Query: User says " Yes I am currently not receiving benefits" Agent says " you don t get Social Security or SSI benefits yet but are planning to apply?" User says " not" Agent says " you already have an account?" User says " yes"

Generated Output:
Agent: Great! Since you already have an account, you can go ahead and apply for retirement benefits by selecting the "Apply for Retirement Benefits" button above. Let me know if you need any assistance with the application process.

Retrieved document id's: 9526ba0e-e195-4de4-8435-050cd93a4994, cda42cce-e092-4fe4-b546-adb841210990, bab4e5a6-68b0-404a-807f-40587f2fadaa

Faithfulness: 0.57992

Retrieval Relevance: 0

Response Relevance: 0.5

Temperature:

Query: User says " No I have no idea." Agent says " Have you any idea as to how long the counseling process takes?" User says " No I don't know." Agent says " Do you have a verified FSA ID?" User says " Yes I have a FSA ID."

Generated Output:
Agent: Great, in that case, you can complete the counseling process in a single session, which usually takes about 20-30 minutes. Make sure you have your verified FSA ID ready before starting. Let me know if you need any assistance with the process.

Retrieved document id's: 7ca94851-1e48-4302-a44a-5dcacf58ce7b, 3a559715-4189-499e-8943-53f1b8652f12, 5866d5e6-309a-4761-ae16-c73880cd4412

Faithfulness: 0.18807

Retrieval Relevance: 0.6

Response Relevance: 0.8

Temperature:

Query: User says " I just wanted to know if there are any resources and/or support outside VA which I could access."

Generated Output:
Agent: Absolutely, there are resources and support available outside of the VA that you can access. Your VA care team can provide you with more information on these options and help connect you with the appropriate services. Would you like me to provide you with more information on your VA care team and how they can assist you in finding external resources and support?

Retrieved document id's: 3c68d6e5-3442-4297-b827-821d8b6ffba3, 60ffdbcd-e527-4829-a58d-36c0808fd8cd, ed412b99-e184-4cd8-a2e1-e88889933c81

Faithfulness: 0.14911

Retrieval Relevance: 1

Response Relevance: 0.9

Temperature:

Query: User says " Yes I am currently not receiving benefits" Agent says " you don t get Social Security or SSI benefits yet but are planning to apply?" User says " not" Agent says " you already have an account?" User says " yes"

Generated Output:
Agent: "Great, since you already have an account, we can help you with the application process for Social Security or SSI benefits. Let's get started on that right away."

Retrieved document id's: 9526ba0e-e195-4de4-8435-050cd93a4994, cda42cce-e092-4fe4-b546-adb841210990

Faithfulness: 0.0073

Retrieval Relevance: 0

Response Relevance: 0.3

Temperature:

Query: User says " Hello i would like to get information about Program Reviews"

Generated Output:
Agent: Hello! I can definitely help you with that. What specific information are you looking for regarding Program Reviews?

Retrieved document id's: 7342fd17-9275-4f34-8a8d-40dcf5f75e6e, 867e8dff-b9ac-429d-a662-abce9ad70c53

Faithfulness: 0.10082

Retrieval Relevance: 0.3

Response Relevance: 1

Temperature:

Query: User says " I just wanted to know if there are any resources and/or support outside VA which I could access."

Generated Output:
Agent: Absolutely, there are resources and support available outside of the VA that you can access. Would you like me to provide you with more information about your VA care team and how they can assist you in connecting with these external resources?

Retrieved document id's: 3c68d6e5-3442-4297-b827-821d8b6ffba3, 60ffdbcd-e527-4829-a58d-36c0808fd8cd

Faithfulness: 0.18992

Retrieval Relevance: 1

Response Relevance: 0.8

Temperature:

Query: User says " Yes I am currently not receiving benefits" Agent says " you don t get Social Security or SSI benefits yet but are planning to apply?" User says " not" Agent says " you already have an account?" User says " yes"

Generated Output:
Agent: Great, since you already have an account, let's go ahead and check the status of your benefits. Can you please provide me with your account information?

Retrieved document id's: 9526ba0e-e195-4de4-8435-050cd93a4994, cda42cce-e092-4fe4-b546-adb841210990

Faithfulness: 0.0102

Retrieval Relevance: 0

Response Relevance: 0.8

Temperature:

Query: User says " what is disability?"

Generated Output:
Agent: Disability is a physical or mental condition that limits a person's movements, senses, or activities. It can vary widely and may require accommodations or support to

help individuals fully participate in society.

Retrieved document id's: 37e88c37-b1eb-444b-8fc0-f8714a265c54, e9322f55-fdfd-4014-9982-eecbb6772dba

Faithfulness: 0.26133

Retrieval Relevance: 0.3

Response Relevance: 1

Temperature:

Query: User says " I just wanted to know if there are any resources and/or support outside VA which I could access."

Generated Output:
Agent: Absolutely, there are resources and support available outside of the VA that you can access. Your VA care team can provide you with information on community resources, support groups, and other services that may be beneficial to you. Feel free to ask your care team for more information on how to access these external resources.

Retrieved document id's: 3c68d6e5-3442-4297-b827-821d8b6ffba3, 60ffdbcd-e527-4829-a58d-36c0808fd8cd

Faithfulness: 0.14409

Retrieval Relevance: 1

Response Relevance: 0.9

Temperature:

Query: User says " Hello there. My daughter is 17 years old and I came here to find out what she needs to consider, being an underage driver."

Generated Output:
Agent: Hello! To legally drive in this country, your daughter will need to meet the minimum age requirement for obtaining a driver's license, which is typically 16 or 17 years old depending on the state. Additionally, she will need to pass a written knowledge test and a driving skills test. It's also important

Retrieved document id's: 4469bacc-3053-4710-bcc4-2be759e4043a, a1b61eb6-818c-434a-83fd-3983330efdb4

Faithfulness: 0.01328

Retrieval Relevance: 0.2

Response Relevance: 0.8

Temperature:

Query: User says " Hello i would like to get information about Program Reviews"

Generated Output:
Agent: Hello! I can definitely help you with that. What specific information are you looking for regarding Program Reviews?

Retrieved document id's: 7342fd17-9275-4f34-8a8d-40dcf5f75e6e, 867e8dff-b9ac-429d-a662-abce9ad70c53

Faithfulness: 0.10082

Retrieval Relevance: 0.3

Response Relevance: 1

Temperature:

Query: User says " I just wanted to know if there are any resources and/or support outside VA which I could access."

Generated Output:
Agent: Absolutely, there are resources and support available outside of the VA that you can access. Your VA care team can provide you with information on community resources, support groups, and other services that may be beneficial to you. Feel free to reach out to your care team for more information on how to access these resources.

Retrieved document id's: 3c68d6e5-3442-4297-b827-821d8b6ffba3, 60ffdbcd-e527-4829-a58d-36c0808fd8cd

Faithfulness: 0.14367

Retrieval Relevance: 1

Response Relevance: 0.9

Temperature:

Query: User says " Yes I am currently not receiving benefits" Agent says " you don t get Social Security or SSI benefits yet but are planning to apply?" User says " not" Agent says " you already have an account?" User says " yes"

Generated Output:
Agent: "Great, since you already have an account, we can help you with the application process for Social Security or SSI benefits. Let's get started on that right away."

Retrieved document id's: 9526ba0e-e195-4de4-8435-050cd93a4994, cda42cce-e092-4fe4-b546-adb841210990

Faithfulness: 0.0073

Retrieval Relevance: 0

Response Relevance: 0.3

Temperature:

Query: User says " Hello i would like to get information about Program Reviews"

Generated Output:
Agent: Hello! I can definitely help you with that. What specific information are you looking for regarding Program Reviews?

Retrieved document id's: 7342fd17-9275-4f34-8a8d-40dcf5f75e6e, 867e8dff-b9ac-429d-a662-abce9ad70c53

Faithfulness: 0.10082

Retrieval Relevance: 0.3

Response Relevance: 1

Temperature:

Query: User says " I just wanted to know if there are any resources and/or support outside VA which I could access."

Generated Output:
Agent: Absolutely, there are resources and support available outside of the VA that you can access. Would you like me to provide you with more information about your VA care team and how they can assist you in connecting with these external resources?

Retrieved document id's: 3c68d6e5-3442-4297-b827-821d8b6ffba3, 60ffdbcd-e527-4829-a58d-36c0808fd8cd

Faithfulness: 0.18992

Retrieval Relevance: 1

Response Relevance: 0.8