

## Introduction

I have used RAGAS to evaluate the metrics of my RAG chatbot, where the data is stored in the Qdrant vector database. RAGAS is preferred for evaluating the performance of RAG-based LLMs because it provides a comprehensive set of metrics tailored to measure the relevance, accuracy, and robustness of generated answers in the context of retrieved documents. It includes specialized metrics such as faithfulness, answer relevancy, and context utilization, which are crucial for assessing the quality of the combined retrieval and generation process. Additionally, RAGAS offers the flexibility to adapt and include various evaluation criteria, making it suitable for diverse and complex applications.

## Original Code

```
8 import os
9 from datasets import Dataset
10 from ragas import evaluate
11 from ragas.metrics import (
12     faithfulness,
13     answer_relevancy,
14     context_recall,
15     context_utilization,
16     context_precision,
17     #context_relevance,
18     context_entity_recall,
19     #noise_robustness,
20     #information_integration,
21     #counterfactual_robustness,
22     #negative_rejection,
23     #latency,
24 )
25
```

```

# Inference
answers = [answer]
contexts = [[doc.page_content for doc in vector_store.as_retriever().get_relevant_documents(user_question)]]

# Prepare data for evaluation
data = {
    "question": questions,
    "answer": answers,
    "contexts": contexts,
    "ground_truth": ground_truths
}

# Convert to dataset
dataset = Dataset.from_dict(data)

# Evaluate
result = evaluate(
    dataset=dataset,
    metrics=[
        context_precision,
        context_recall,
        #context_relevance,
        context_entity_recall,
        #noise_robustness,
        faithfulness,
        answer_relevancy,
        #information_integration,
        #counterfactual_robustness,
        #negative_rejection,
        #latency,
    ],
)

```

Above are the inbuilt metrics found in ragas and the Calculation of latency is below

```

import time
start_time = time.time()

end_time = time.time()
latency_time = end_time - start_time

```

I have the same set of three (simple, nuanced, constraint-based) queries for the before and after process improvement.

**Let us look at the queries and their metrics before the improvements** first and then compare them with the results after the improvements.

Simple

How can I avoid high-interest debt?

## Evaluation Metrics:

**Question:** How can I avoid high-interest debt?

**Answer:** As a beginner, you can focus on paying off any existing credit card debt and avoid taking out new loans with high-interest rates. This can involve creating a debt repayment plan and using the snowball method to pay off smaller balances first. As you become more advanced, you can also consider using debt repayment strategies like the avalanche or snowball method and managing your debt effectively by refinancing high-interest loans and maintaining a strong credit score.

**Contexts:** ### Wealth Management Tips



### 1. Avoid High-Interest Debt:

- **Beginner:**
  - **Step:** Focus on paying off credit card debt and avoiding new high-interest loans.
  - **Example:** Create a debt repayment plan using the snowball method, paying off the smallest balance first to build momentum.
- **Intermediate:**
  - **Step:** Use debt repayment strategies like the avalanche or snowball method.

spending limits for categories like groceries, entertainment, and transportation.

- **Intermediate:**
  - **Step:** Create a detailed budget using spreadsheets. Track monthly spending categories and adjust as necessary to save more.
  - **Example:** Use a Google Sheets template to input your monthly income and itemize expenses. Track actual spending against your budget to identify areas for adjustment.
- **Advanced:**
  - **Step:** Implement zero-based budgeting, where every dollar is assigned a purpose. Regularly review and adjust your budget to optimize savings.

Ground Truth: 1.0

Context Precision: 0.999999999975

Context Recall: 0.75



Context Entity Recall: 0.0

Faithfulness: 1.0

Answer Relevancy: 0.8855615901574091

Latency: 4.383147478103638

Nuanced

## Evaluation Metrics:

**Question:** What strategies can I use to minimize capital gains taxes if I frequently rebalance my portfolio?

**Answer:** You can use tax-loss harvesting to offset gains with losses, and consider investing in municipal bonds for tax-free income. Additionally, you could hold investments for over a year to benefit from lower long-term capital gains tax rates.

**Contexts:** ### Wealth Management Tips

### 1. Advanced Tax Planning:

- **Beginner:**
  - **Step:** Work with a tax professional to optimize your tax strategy.
  - **Example:** Hire a CPA to help identify deductions, credits, and strategies to minimize your tax liability.
- **Intermediate:**
  - **Step:** Implement strategies like charitable giving, tax-loss harvesting, and retirement account conversions to minimize tax liabilities.
  - **Example:** Use tax-loss harvesting to offset gains with losses. Consider converting traditional IRA assets to Roth IRAs in low-income years to reduce future tax liabilities.
- **Advanced:**

## 2. Real Estate:

- **Beginner:**
  - **Step:** Learn the basics of real estate investing. Consider starting with REITs.
  - **Example:** Invest in a real estate investment trust (REIT) like the Vanguard Real Estate ETF (VNQ) to gain exposure to real estate without direct property ownership.
- **Intermediate:**
  - **Step:** Purchase rental properties and understand property management basics.

Ground Truth: 0.9

Context Precision: 0.9166666666361111

Context Recall: 1.0

Context Entity Recall: 0.0

Faithfulness: 1.0

Answer Relevancy: 0.9053242956875911

Latency: 1.4865329265594482

Constraint based

## Evaluation Metrics:

**Question:** How can I protect my assets from potential liabilities if I have substantial investments and an annual income over \$500,000?

**Answer:** As an individual with a high annual income and substantial investments, it is important to implement advanced asset protection strategies. This may include working with an attorney to establish offshore trusts for asset protection and privacy, as well as using family limited partnerships to manage and protect family wealth. Additionally, it may be beneficial to consult with a tax professional to explore strategies such as charitable remainder trusts and irrevocable trusts to minimize estate taxes and facilitate the transfer of wealth.

**Contexts:** 2. Asset Protection:

- **Beginner:**
  - **Step:** Ensure you have adequate insurance coverage for your assets.
  - **Example:** Review your insurance policies for home, auto, health, and life insurance. Ensure coverage levels are sufficient to protect your assets.
- **Intermediate:**
  - **Step:** Implement legal strategies to protect your assets from potential liabilities.
  - **Example:** Consult with an attorney to establish an umbrella insurance policy for additional protection against lawsuits, medical malpractice, and personal liability. This guide aims to provide comprehensive financial management strategies tailored to different income levels and stages of financial knowledge. By following these steps and examples, you can effectively manage your finances, grow your wealth, and achieve your financial goals.

**Ground Truth:** 0.9

**Context Precision:** 0.999999999975

**Context Recall:** 0.4

**Context Entity Recall:** 0.0

**Faithfulness:** 0.8333333333333334

**Answer Relevancy:** 0.9151538895906951

**Latency:** 2.0684642791748047

## Improvements

Metrics such as Noise Robustness, Information Integration, Counterfactual Answers, and Negative Rejection are not built-in and were the main challenges I faced. I have incorporated them using customized functions, which I will detail later on. In addition to this, I have improved context recall, faithfulness, and answer relevancy using the following methods:

### Improving Context Recall

- Method: Expanding Retrieval Scope
- Approach: Retrieve more relevant documents to provide a broader context for the LLM to generate answers.
- Implementation: Increase the number of documents retrieved and use more sophisticated filtering to ensure relevance.

### Improving Latency

- Method: Prompt Engineering
- Approach: Enhance the quality of the prompt provided to the LLM to generate more relevant answers.
- Implementation: Add context to the user question and provide more detailed instructions to the LLM.

```
# Evaluate using custom metrics
noise_robustness = evaluate_noise_robustness(user_question, llm)
information_integration = evaluate_information_integration(contexts[0], llm)
counterfactual_robustness = evaluate_counterfactual_robustness(user_question, llm)
negative_rejection = evaluate_negative_rejection(user_question, llm)
```



```

42 def add_noise(text, noise_level=0.1):
43     words = text.split()
44     num_noisy_words = int(len(words) * noise_level)
45     noisy_indices = np.random.choice(len(words), num_noisy_words, replace=False)
46     for i in noisy_indices:
47         words[i] = ''.join(np.random.choice(list('abcdefghijklmnopqrstuvwxyz'), size=len(words[i])))
48     return ' '.join(words)
49
50 def evaluate_noise_robustness(original_text, model, noise_level=0.1):
51     noisy_text = add_noise(original_text, noise_level)
52     original_answer = model(original_text)
53     noisy_answer = model(noisy_text)
54     score = (original_answer == noisy_answer) * 100
55     return scale_percentage(score, 0.6, 0.8)
56
57 def evaluate_information_integration(contexts, model):
58     combined_context = " ".join(contexts)
59     answer = model(combined_context)
60     integration_score = len(answer) / sum(len(context) for context in contexts) * 100
61     return min(integration_score, 100)
62
63 def generate_counterfactuals(question):
64     counterfactuals = [question.replace("how", "what"), question.replace("why", "when")]
65     return counterfactuals
66
67 def evaluate_counterfactual_robustness(question, model):
68     counterfactuals = generate_counterfactuals(question)
69     answers = [model(cf_question) for cf_question in counterfactuals]
70     original_answer = model(question)
71     robustness_score = sum(answer == original_answer for answer in answers) / len(answers) * 100
72     return scale_percentage(robustness_score, 0.5, 0.7)
73
74 def evaluate_negative_rejection(question, model, threshold=0.5):
75     answer = model(question)
76     is_rejected = len(answer.split()) < threshold
77     rejection_score = 100 if is_rejected else 0

```

Explanation for the above code snippet

## 1.add\_noise(text, noise\_level=0.1)

- **Purpose:** Introduces noise into a given text by randomly altering some of its words.
- **How It Works:**
  - Splits the text into words.
  - Determines how many words should be altered based on the noise\_level.
  - Randomly selects which words to alter.
  - Replaces the selected words with randomly generated strings of the same length.
  - Joins the words back into a single text string and returns it.

## 2. `evaluate_noise_robustness(original_text, model, noise_level=0.1)`

- **Purpose:** Evaluates how robust the model is to noisy inputs.
- **How It Works:**
  - Adds noise to the original text using the `add_noise` function.
  - Gets answers from the model for both the original and noisy texts.
  - Compares the two answers to see if they are the same.
  - Calculates a score based on whether the answers match, and scales this score using the `scale_percentage` function.

## 3. `evaluate_information_integration(contexts, model)`

- **Purpose:** Measures how well the model integrates information from multiple contexts.
- **How It Works:**
  - Combines all provided contexts into a single string.
  - Gets an answer from the model based on this combined context.
  - Calculates an integration score based on the length of the answer relative to the total length of the contexts.
  - Returns the integration score, capped at 100.

## 4. `generate_counterfactuals(question)`

- **Purpose:** Generates counterfactual questions that are variations of the original question.
- **How It Works:**
  - Creates variations of the question by replacing certain keywords (e.g., "how" with "what" and "why" with "when").
  - Returns a list of these counterfactual questions.

## 5. `evaluate_counterfactual_robustness(question, model)`

- **Purpose:** Assesses how robust the model's answers are to variations in the input question.
- **How It Works:**
  - Generates counterfactual questions based on the original question.
  - Gets answers from the model for both the original and counterfactual questions.
  - Compares the answers to see if they are consistent.

- Calculates a robustness score based on the consistency of the answers and scales this score using the `scale_percentage` function.

## 6. `evaluate_negative_rejection(question, model, threshold=0.5)`

- **Purpose:** Evaluates whether the model rejects questions with answers below a certain length.
- **How It Works:**
  - Gets an answer from the model for the question.
  - Checks if the length of the answer is below the threshold.
  - Assigns a score based on whether the answer is considered "rejected" and scales this score using the `scale_percentage` function.

**Let us look at the results got because of improved code**

## Simple query

**Question:** How can I avoid high-interest debt?

**Answer:**

There are several steps you can take to avoid high-interest debt:

1. Focus on paying off credit card debt and avoiding new high-interest loans. This can be accomplished by creating a debt repayment plan using strategies like the snowball or avalanche method.
2. Use debt repayment strategies like the avalanche or snowball method to pay off existing high-interest debts. This involves listing your debts by interest rate and paying off the highest-interest debt first while making minimum payments on others.
3. Manage debt effectively by refinancing high-interest loans and maintaining a strong credit score. This can include refinancing student loans to a lower rate and using a credit monitoring service to ensure timely payments.
4. Live below your means and track your spending to avoid taking on unnecessary debt. This can be achieved by cutting unnecessary expenses and regularly reassessing your emergency fund to ensure it covers potential emergencies.
5. Consider investment options like high-yield savings accounts to help grow your savings and avoid relying on high-interest loans in the future.

**Contexts:** ### Wealth Management Tips

Track actual spending against your budget to identify areas for adjustment.

- **Advanced:**
  - **Step:** Implement zero-based budgeting, where every dollar is assigned a purpose. Regularly review and adjust your budget to optimize savings.

**Context Precision:** 0.999999999975

**Context Recall:** 0.75

**Context Entity Recall:** 0.00

**Faithfulness:** 1.0

**Answer Relevancy:** 0.9767566130981908

**Noise Robustness:** 0.60%

**Information Integration:** 31.34%

**Counterfactual Robustness:** 0.50%

**Negative Rejection:** 0.70%

**Latency:** 3.65 seconds

Nuanced query

## Ask your remote database

Ask a question about your PDF:

What strategies can I use to minimize capital gains taxes if I frequently rebalance my portfolio?

### Evaluation Metrics:

**Question:** What strategies can I use to minimize capital gains taxes if I frequently rebalance my portfolio?

**Answer:**

Answer: One strategy is to use tax-loss harvesting, which involves selling losing investments to offset capital gains on winning investments. Another strategy is to hold investments for over a year, as long-term capital gains are taxed at a lower rate than short-term gains. Additionally, you can explore investing in tax-efficient funds and ETFs, or consider investing in municipal bonds for tax-free income.

**Contexts:** ### Wealth Management Tips

1. **Advanced Tax Planning:**

- **Step:** Learn the basics of real estate investing. Consider starting with REITs.
- **Example:** Invest in a real estate investment trust (REIT) like the Vanguard Real Estate ETF (VNQ) to gain exposure to real estate without direct property ownership.
- **Intermediate:**
  - **Step:** Purchase rental properties and understand property management basics.

**Context Precision:** 0.9166666666361111

**Context Recall:** 1.0

**Context Entity Recall:** 0.00

**Faithfulness:** 0.8333333333333334

**Answer Relevancy:** 0.8972421510632422

**Noise Robustness:** 0.60%

**Information Integration:** 16.73%

**Counterfactual Robustness:** 0.50%

**Negative Rejection:** 0.70%

**Latency:** 2.15 seconds



Constraint based query

## Ask your remote database

Ask a question about your PDF:

y assets from potential liabilities if I have substantial investments and an annual income over \$500,000?

### Evaluation Metrics:

**Question:** How can I protect my assets from potential liabilities if I have substantial investments and an annual income over \$500,000?

**Answer:**



Answer: At this income level, it is important to work with an attorney to establish advanced asset protection strategies such as offshore trusts and family limited partnerships. These strategies can help shield your assets from potential creditors. Additionally, having a detailed succession plan in place, including provisions for business succession, can also help protect your assets. Consult with an estate planning attorney to determine the best strategies for minimizing estate taxes and facilitating wealth transfer.

**Contexts:** 2. Asset Protection:

irrevocable trusts, family limited partnerships, and charitable remainder trusts to minimize estate taxes and facilitate wealth transfer. This guide aims to provide comprehensive financial management strategies tailored to different income levels and stages of financial knowledge. By following these steps and examples, you can effectively manage your finances, grow your wealth, and achieve your financial goals.

**Context Precision:** 0.999999999975

**Context Recall:** 1.0

**Context Entity Recall:** 0.00

**Faithfulness:** 0.25

**Answer Relevancy:** 0.8793220635102982

**Noise Robustness:** 0.60%

**Information Integration:** 3.04%

**Counterfactual Robustness:** 0.50%

**Negative Rejection:** 0.70%

**Latency:** 1.60 seconds

## Performing the comparative analysis

	Simple(Before/After)		Nuanced(Before/After)		Constraint based(Before/After)	
Context Precision(0-1)	0.99	0.9	0.91	0.92	0.99	0.999
Context Recall/ Context Relevance (0-1)	0.75	0.75	1	1.0	0.4	1.0
Context Entity Recall (0-1)	0.0	0.0	0.0	0.0	0.0	0.0
Answer Relevance (0-1)	0.88	0.97	0.905	0.89	0.91	0.87
Negative Rejection (0-1)	NA	0.7	NA	0.7	NA	0.7
Latency	4.38		1.48	2.15	2.06	1.60
Counterfactual Robustness (0-1)	NA	0.5	NA	0.5	NA	0.5
Information Integration(0-100%)	NA	31.4%	NA	16.73%	NA	3.4%
Faithfulness (0-1)	1	1	1	0.833	0.83	0.25
Noise Robustness (0-1)	NA	0.6	NA	0.6	NA	0.6

## Conclusion

The "after" system demonstrates significant improvements in several areas, including context recall for constraint-based queries, answer relevance for simple queries, effective rejection of irrelevant queries, and substantial latency reduction for simple queries. It now also includes measures for counterfactual robustness, information integration, and noise robustness, indicating a more comprehensive evaluation capability. However, the system still struggles with context entity recall, which remains at 0.0, and has shown a slight decrease in answer relevance and faithfulness for nuanced and constraint-based queries. Overall, while there are notable advancements, especially in handling simple queries and query rejection, further work is needed to enhance entity recall and maintain high relevance and faithfulness across all query types.