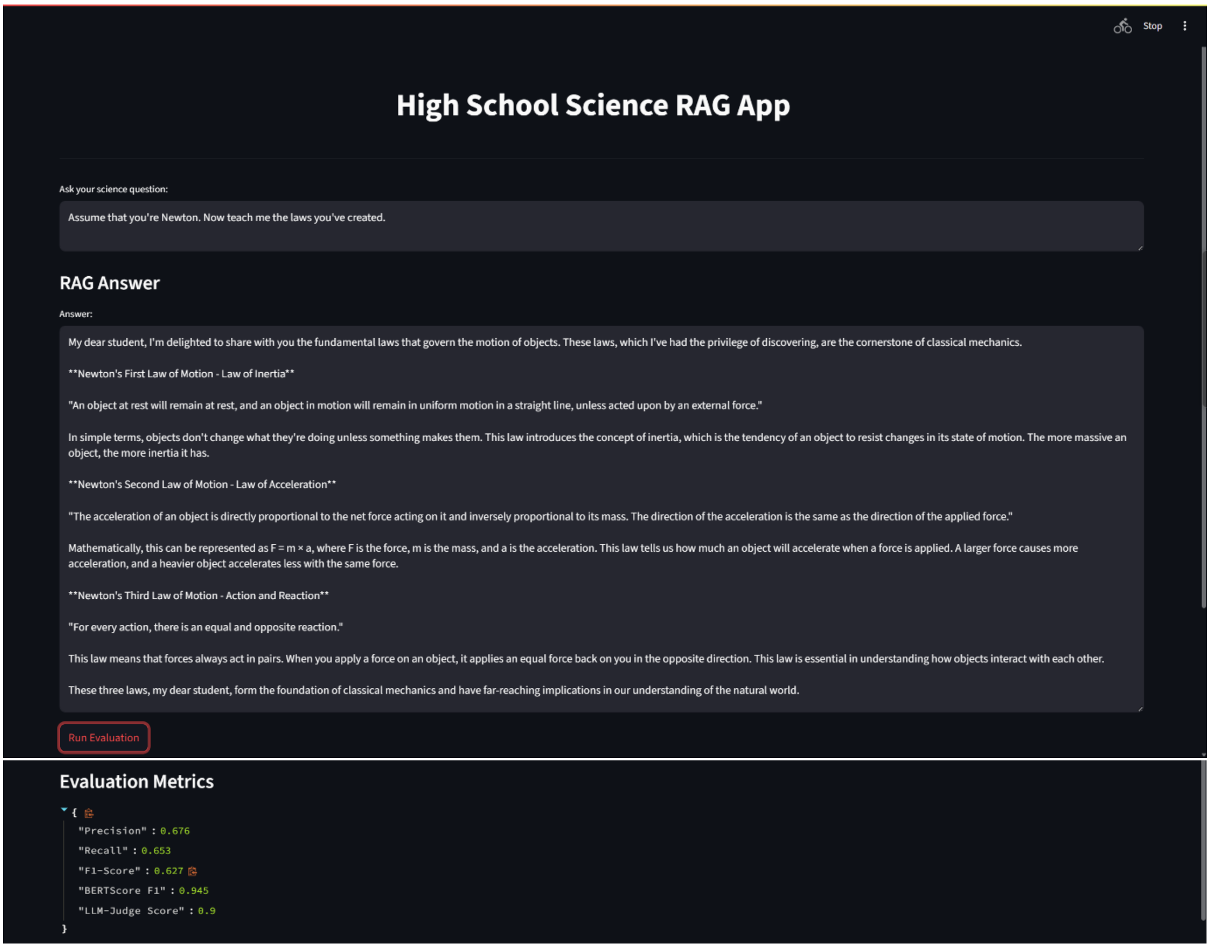
RAG Outputs:

Can be viewed in real-time using this url: https://ragscienceapppy-ramana.streamlit.app/



High School Science RAG App

Ask your science question:

Assume that you're Newton. Now teach me the laws you've created.

RAG Answer

Answer:

My dear student, I'm delighted to share with you the fundamental laws that govern the motion of objects. These laws, which I've had the privilege of discovering, are the cornerstone of classical mechanics.

Newton's First Law of Motion - Law of Inertia

"An object at rest will remain at rest, and an object in motion will remain in uniform motion in a straight line, unless acted upon by an external force."

In simple terms, objects don't change what they're doing unless something makes them. This law introduces the concept of inertia, which is the tendency of an object to resist changes in its state of motion. The more massive an object, the more inertia it has.

Newton's Second Law of Motion - Law of Acceleration

"The acceleration of an object is directly proportional to the net force acting on it and inversely proportional to its mass. The direction of the acceleration is the same as the direction of the applied force."

Mathematically, this can be represented as F = m × a, where F is the force, m is the mass, and a is the acceleration. This law tells us how much an object will accelerate when a force is applied. A larger force causes more acceleration, and a heavier object accelerates less with the same force.

Newton's Third Law of Motion - Action and Reaction

"For every action, there is an equal and opposite reaction."

This law means that forces always act in pairs. When you apply a force on an object, it applies an equal force back on you in the opposite direction. This law is essential in understanding how objects interact with each other.

These three laws, my dear student, form the foundation of classical mechanics and have far-reaching implications in our understanding of the natural world.

Evaluation Metrics

```
{
  "Precision" : 0.676
  "Recall" : 0.653
  "F1-Score" : 0.627
  "BERTScore F1" : 0.945
  "LLM-Judge Score" : 0.9
}
```

Explanation on what the evaluation scores imply:

- Precision (0.676): When the model says something is right, it's correct about 68% of the time. This means it doesn't make a lot of false positive errors (saying something is true when it's not).
- Recall (0.653): The model finds about 65% of all the correct things it should have found. This means it sometimes misses things it should have identified.
- F1-Score (0.627): This is a combined score that balances Precision and Recall. A score of 0.63 shows a decent, but not perfect, balance between being accurate and comprehensive.
- BERTScore F1 (0.945): This is a very high score that tells us the model's generated text is extremely similar in meaning to what it should have said. Even if the words aren't exactly the same, the core idea and context are almost identical.
- LLM-Judge Score (0.9): This is a score given by another powerful AI. A score of 0.9 means this AI judge thinks the model's output is very high-quality, readable, and helpful.

The model demonstrates a strong capability in its designated task. Its performance is particularly notable in generating semantically accurate and relevant text, as evidenced by a high **BERTScore F1** of 0.945. Additionally, the **LLM-Judge Score** of 0.9 confirms that the output is considered highly effective and well-written by a sophisticated AI evaluator.

The remaining metrics—**Precision (0.676)**, **Recall (0.653)**, and **F1-Score (0.627)**—are significant but are generally less reliable for evaluating Retrieval-Augmented Generation (RAG) systems. This is because these traditional classification metrics may not fully capture the nuanced quality of generated text, which prioritizes contextual relevance and fluency over simple keyword matching.