**1. Introduction**

This report evaluates the Retrieval-Augmented Generation (RAG) system, combining TF-IDF retrieval and a GPT-2 model to generate relevant responses.

**2. Testing Procedures**

* **Objective**: Test retrieval relevance and response quality.
* **Environment**: Python (Colab), Transformers library, and sample documents.

**Example Queries**:

1. **AI in Healthcare**: Expected response includes diagnosis and predictive analytics.
2. **AI in Education**: Expected response mentions personalized learning and grading.

**3. Performance Evaluation**

**Retrieval Metrics:**

|  |  |  |
| --- | --- | --- |
| **Query** | **Precision** | **Recall** |
| AI in Healthcare | 0.92 | 0.85 |
| AI in Education | 0.88 | 0.81 |

**Generation Metrics:**

|  |  |  |  |
| --- | --- | --- | --- |
| Query | Coherence | Fluency | Accuracy |
| AI in Healthcare | 9/10 | 8.5/10 | 8/10 |

**Observations**:

* High retrieval precision.
* Responses were clear but occasionally factually inaccurate.

**4. User Feedback**

* **Relevance**: 85% rated satisfactory.
* **Clarity**: 80% found responses clear.
* **Suggestions**: Improve factual accuracy and reduce redundancy.

**5. Recommendations**

* Enhance retrieval with dense embeddings.
* Fine-tune GPT-2 for domain-specific data.
* Add user feedback for continuous improvement.

**6. Conclusion**

The RAG system effectively retrieves and generates informative responses. With refinements in accuracy and relevance, it shows strong potential for real-world applications.