RAG Summary System - Documentation

1. System Overview

- Local document processing pipeline (PDF/TXT/MD)
- Combines retrieval & generation (100% offline)
- Outputs: Retrieved context + generated summary

2. Requirements

Hardware

- Minimum: 8GB RAM (CPU-only)
- Recommended: 16GB RAM + GPU

Software

```
text
Copy
Download
Python 3.8+
transformers==4.36.2
torch>=2.0.0
langchain==0.0.340
sentence-transformers==2.2.2
faiss-cpu==1.7.4
PyPDF2==3.0.1
```

3. Quick Setup

```
bash
Copy
Download
python -m venv rag_env
source rag_env/bin/activate  # Linux/Mac
rag_env\Scripts\activate    # Windows
pip install -r requirements.txt
```

4. Usage

```bash
python rag_summary.py input.pdf
```

Supported Files:

- PDF (.pdf)
- Text (.txt)
- Markdown (.md)

5. Key Parameters (Customizable)

```python
# Chunking
chunk_size=512      # Tokens per chunk
chunk_overlap=100   # Overlap between chunks

# Retrieval
k=5                 # Number of chunks to retrieve

# Summary
max_length=256      # Max summary tokens
min_length=64       # Min summary tokens
```

6. Troubleshooting

- Memory Errors: Reduce `chunk_size`
- PDF Issues: Convert to text first
- Model Downloads: Check internet connection on first run

7. Limitations

- Optimal for <100 page documents
- English-focused
- Complex PDF formatting may be lost

```
offline_rag_summary("global_warming.pdf")
```

```
 Loading document...
✂ Chunking into semantic blocks...
 Building FAISS vector index...
 Retrieving top relevant chunks...
 Generating summary using BART (free)...

 Retrieved Context
--------------------------------------------------------------------------------
Review, 100(2), 52 -59. fully understand the implications of global warming or consider it a
significant problem for the future. However, global  warming is already happening, and some of its
devastating consequences are already being felt. It significantly impacts  biodiversity and disrupts
ecological balance. Due to the dangerous effects of global warming, many strategies need to  be
established. The report discusses global warming, outlines its causes and risks, and proposes
solutions to this urgent driving it. The rapid increase in greenhouse gases is problematic, as it
affects the environment faster than many living  organisms can a dapt. This changing and
increasingly complex world presents significant challenges for all forms of life.   The report
discussed here only scratches the surface of the complex field of science and engineering related to
global  warming. It is a serious threat  that demands effective measures to address it. Global
warming poses significant problems [34] Aitken, D. W. (2003). Transitioning to a renewable energy
future. ISES White Paper.    [35] Canadell, J. G., & Raupach, M. R. (2008). Managing forests for
climate change mitigation. science, 320(5882), 1456 - 1457.  https:
//doi.org/10.1126/science.1155458   [36] Robinson, S. (2010). How to reduce carbon emissions:
capture and store it. Time. com.   [37] International Energy Agency. (2009). Technology roadmap:
Carbon capture and storage. OECD Publishing.
--------------------------------------------------------------------------------

 Final Summary
--------------------------------------------------------------------------------
The report discusses global warming, outlines its causes and risks, and proposes solutions to this
urgent driving it. The rapid increase in greenhouse gases is problematic, as it affects the
environment faster than many living organisms. This changing and increasingly complex world presents
significant challenges for all forms of life. It is a serious threat that demands effective measures
to address it.
--------------------------------------------------------------------------------
```

```
offline_rag_summary("future_of_ai.md")
```

```
 Loading document...
✂ Chunking into semantic blocks...
 Building FAISS vector index...
 Retrieving top relevant chunks...
 Generating summary using BART (free)...

 Retrieved Context
--------------------------------------------------------------------------------
## Conclusion  AI will undoubtedly shape the future of our society. However, it is essential to
address its challenges proactively and ensure its development benefits all of humanity. ## Key
Developments  - **Machine Learning & Deep Learning:** These have enabled computers to learn from
data and make decisions with minimal human intervention. - **Natural Language Processing:** With
advancements in NLP, machines can now understand, generate, and translate human languages more
effectively. - **Computer Vision:** AI-powered systems can interpret visual inputs, aiding in facial
recognition, medical imaging, and autonomous vehicles.  ## Ethical Considerations # The Future of
Artificial Intelligence  Artificial Intelligence (AI) has rapidly evolved over the past decade,
influencing multiple domains ranging from healthcare and finance to education and entertainment.  ##
Key Developments ## Ethical Considerations  While AI presents numerous benefits, it also raises
concerns:  - **Job Displacement:** Automation could reduce the need for human labor in some sectors.
- **Bias and Fairness:** AI systems trained on biased data may produce unfair outcomes. -
**Privacy:** AI systems that monitor user behavior pose privacy concerns.  ## Conclusion
--------------------------------------------------------------------------------

 Final Summary
--------------------------------------------------------------------------------
Artificial Intelligence (AI) has rapidly evolved over the past decade, influencing multiple domains
ranging from healthcare and finance to education and entertainment. While AI presents numerous
benefits, it also raises concerns. It is essential to address its challenges proactively and ensure
its development benefits all of humanity. The Future of Artificial Intelligence is published by
Oxford University Press.
--------------------------------------------------------------------------------
```

```
offline_rag_summary("illiteracy.txt")
```

```
 Loading document...
✂ Chunking into semantic blocks...
 Building FAISS vector index...
 Retrieving top relevant chunks...
 Generating summary using BART (free)...

 Retrieved Context
--------------------------------------------------------------------------------------------------------
interpret a table about blood pressure, age, and physical activity; or compute and compare the cost
per ounce of food items. that countries with lower levels of functional illiteracy among their adult
populations tend to be those with the highest levels of scientific literacy among the lower stratum
of young people nearing the end of their formal academic studies. This correspondence suggests that
the capacity of schools to ensure students attain the functional literacy required to comprehend the
basic texts and documents associated with competent citizenship contributes to a society's level of
civic literacy.[3] The National Center for Education Statistics provides more detail.[10] Literacy
is broken down into three parameters: prose, document, and quantitative literacy. Each parameter has
four levels: below basic, basic, intermediate, and proficient. For prose literacy, for example, a
below basic level of literacy means that a person can look at a short piece of text to get a small
piece of uncomplicated information, while a person who is below basic in quantitative literacy would
be able to do simple addition. while a person who is below basic in quantitative literacy would be
able to do simple addition. In the US, 14% of the adult population is at the "below basic" level for
prose literacy; 12% are at the "below basic" level for document literacy, and 22% are at that level
for quantitative literacy. Only 13% of the population is proficient in each of these three
areas—able to compare viewpoints in two editorials; interpret a table about blood pressure, age, and
physical activity; or compute and compare the cost
--------------------------------------------------------------------------------------------------------

 Final Summary
--------------------------------------------------------------------------------------------------------
 Literacy is broken down into three parameters: prose, document, and quantitative literacy. Each
parameter has four levels: below basic, basic, intermediate, and proficient. For prose literacy, for
example, a person can look at a short piece of text to get a small piece of uncomplicated
information. A person who is below basic in quantitative literacy would be able to do simple
addition.
--------------------------------------------------------------------------------------------------------
```