Repo mẫu cho hệ thống RAG có unit test

Cấu trúc thư mục repo: `rag-pipeline-example/`

```
rag-pipeline-example/
├── rag_pipeline/
│   ├── __init__.py
│   ├── retriever.py
│   ├── generator.py
│   └── pipeline.py
├── tests/
│   ├── __init__.py
│   └── test_pipeline.py
├── requirements.txt
└── README.md
```

1. `rag_pipeline/retriever.py`

```python
class SimpleRetriever:
    def __init__(self, documents):
        self.documents = documents

    def retrieve(self, query):
        # Tìm văn bản chứa từ khoá
        return [doc for doc in self.documents if any(word.lower() in doc.lower() for word in query.split())]
```

2. `rag_pipeline/generator.py`

```python
class SimpleGenerator:
    def generate(self, query, documents):
        if not documents:
            return "I don't know."
        return f"Based on the documents, the answer to '{query}' is related to: {documents[0]}"
```

3. `rag_pipeline/pipeline.py`

```python
class RAGPipeline:
    def __init__(self, retriever, generator):
        self.retriever = retriever
        self.generator = generator


    def run(self, query):
        docs = self.retriever.retrieve(query)
        return self.generator.generate(query, docs)
```

4. `tests/test_pipeline.py`

```python
import pytest
from unittest.mock import MagicMock
from rag_pipeline.pipeline import RAGPipeline
from rag_pipeline.retriever import SimpleRetriever
from rag_pipeline.generator import SimpleGenerator

def test_pipeline_with_mock():
    retriever = MagicMock()
    generator = MagicMock()

    retriever.retrieve.return_value = ["Mock document"]
    generator.generate.return_value = "Mock answer"

    rag = RAGPipeline(retriever, generator)
    result = rag.run("What is this?")

    assert result == "Mock answer"
    retriever.retrieve.assert_called_once()
    generator.generate.assert_called_once()

def test_pipeline_with_real_components():
    retriever = SimpleRetriever([
        "Paris is the capital of France.",
        "Berlin is the capital of Germany."
    ])
    generator = SimpleGenerator()

    rag = RAGPipeline(retriever, generator)
    result = rag.run("What is the capital of France?")
```

```
    assert "Paris" in result
```

5. `requirements.txt`

```
pytest
```

Bạn có thể thêm `langchain`, `transformers`, `faiss-cpu` nếu mở rộng thực tế.

6. `README.md`

```
# RAG Pipeline Example

This is a simple Retrieval-Augmented Generation pipeline with unit tests.

## Run tests

```bash
pip install -r requirements.txt
pytest tests/ -v




---


## ▶ Cách chạy

```bash
# Tải repo
git clone <repo-url> rag-pipeline-example
cd rag-pipeline-example

# Cài đặt & chạy test
pip install -r requirements.txt
pytest tests/
```