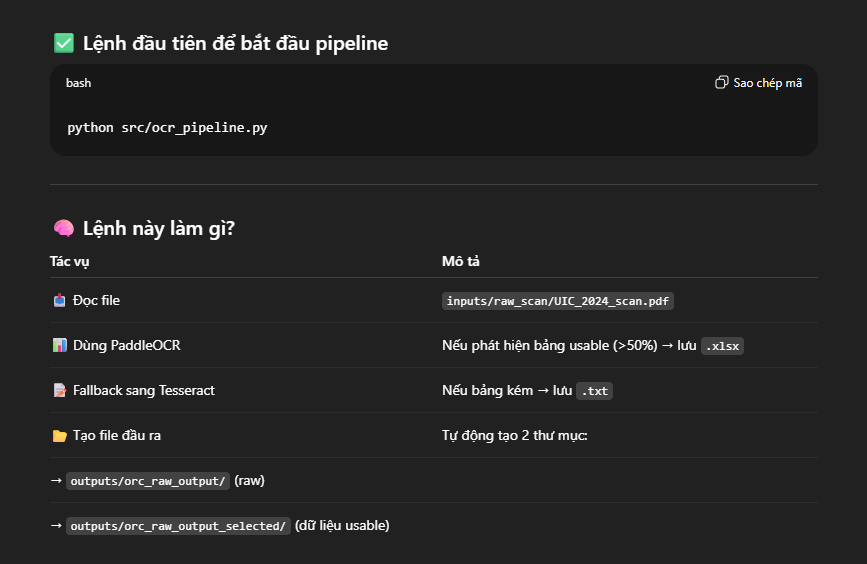
# https://drive.google.com/drive/folders/1yKkKkRuYNZtSQs7biqmuMY3NGuSO9UIO?usp=sharing

# BƯỚC 1 — XỬ LÝ OCR TỪ FILE SCAN

**src/ocr\_pipeline.py: file này để đọc data từ bản scan 🡪 lưu vào orc\_output 🡪 phân chia text / excel + làm sạch nhẹ = pandas 🡪 lưu vào file orc\_output\_selected.**

* 1. Vào môi trường ORC\_STRATEGY

python src/ocr\_pipeline.py



# BƯỚC 2 — **CHUYỂN MÔI TRƯỜNG** STRATEGY, CLEAN BƯỚC 1 TỪ RAW OUTPUT 🡪 CLEAN\_RAW\_OUTPUT

python src/clean\_ocr\_raw\_output.py

# BƯỚC 3 — CLEAN SÂU TỪ CLEAN\_RAW\_OUTPUT 🡪 CLEAN SCAN

python src/clean\_orc\_to\_input.py

# BƯỚC 4 — phân tích tai chinh của file cleaned scan va raw input

python src/ratio\_engine.py --input inputs/cleaned\_scan\_input --output inputs/analyze\_cleaned\_scan\_input

python src/ratio\_engine.py --input inputs/raw\_clean\_docs --output inputs/analyze\_raw\_clean\_docs

# BƯỚC 5 — chaỵ build\_index cho 2 loại giữ liệu.

python -m src.build\_index\_cleaned\_scan\_data

python -m src.build\_index\_raw\_data

# BƯỚC 5 — chaỵ giao diện local

streamlit run app.py

# BƯỚC 5 — chaỵ giao diện cloud

# Khi bạn chỉ sửa 1 file hoặc thêm code mới → chỉ cần:

# git add <tên\_file> # hoặc git add .

# git commit -m "update file ..."

# git push origin main

# 👉 Không cần --force. Lịch sử commit sẽ được nối tiếp, an toàn.

**Quy trình push một file mới/lần sửa đổi**

1. **Kiểm tra trạng thái repo**

git status

→ Bạn sẽ thấy requirements.txt nằm trong danh sách “modified” hoặc “untracked”.

1. **Thêm file vào stage**

git add requirements.txt

1. **Commit thay đổi**

git commit -m "add/update requirements.txt"

# Xóa file secret ra khỏi repo

git rm --cached .streamlit/secrets.toml

git rm --cached .env

git rm --cached .env.\*

1. Commit lại

git commit -m "remove secrets from repo"

git push origin main --force

1. Dungf git filter repo để xóa secret khỏi lịch sử

git filter-repo --path .streamlit/secrets.toml --path .env --path .env.\* --invert-paths --force

1. Them lại remote

git remote add origin https://github.com/huynkawa/Tomtran\_corp\_strategy.git

git remote -v # kiểm tra

1. Push lại lên github

git push origin main --force