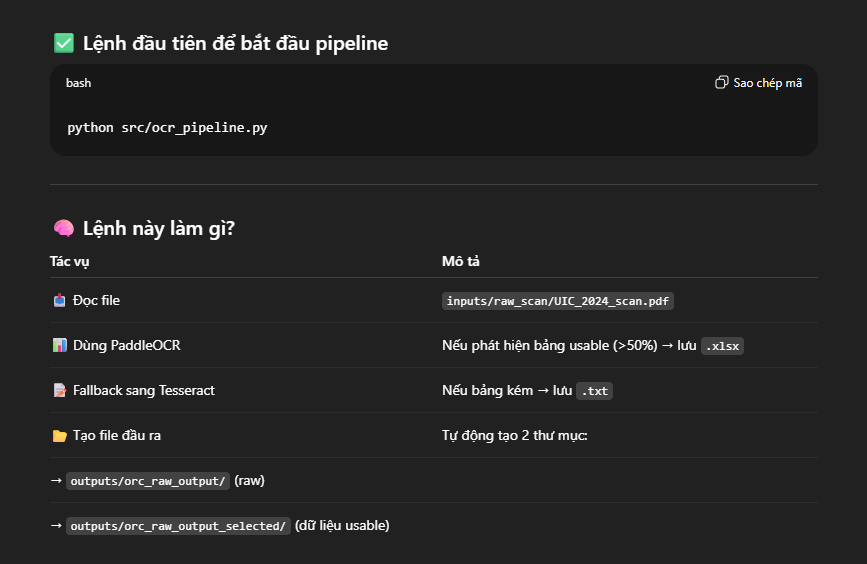
# BƯỚC 1 — XỬ LÝ OCR TỪ FILE SCAN

**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 --output inputs/analyze\_input