# BẢN PROMT CHO PHẦN p1a\_clean10\_ocr\_bctc.py

- Tôi cần trích xuất bảng scan (ảnh đính kèm là 1 ví dụ đang làm) ra .txt dạng đoạn văn mô tả bảng và meta.json. Trong văn bản thì mỗi giá trị sẽ có đầy đủ tên cột lẫn danh mục, để cho thành mỗi giá trị đầy đủ ý nghĩa ấy

- tôi gửi bạn 2 file .txt và .json được sinh ra từ code ở file đính kèm.

- Bạn hãy xem kỹ số liệu sai lêch như nào file txt sinh ra và ảnh scan gốc, đặc biệt các giá trị mã số, thuyết minh, số đầu năm, cuối năm, ngoài ra các tên danh mục và số trước danh mục ví dụ I, II, III, IV, 1,2,3,1.1, 1.2.....

- Bạn xem kỹ file code là có thể tối ưu được để cho số chuẩn xác hơn không.

- ngoài ra có thêm phương án là dùng file YAML để nó đọc trích xuất và chọn số liệu tốt hơn.

- YAML bạn đề xuất nguyên tắc nào để trong YAML để có thể tránh lỗi đọc nhầm bảng báo cáo tài chính, ví dụ dựa vào giá trị số đầu năm và cuối năm nó chỉ chênh lệch tối đa là 50%, hoặc là thông thường các danh mục tên tài sản, dự phòng bồi thường, hay nợ phải trả nó sẽ tương ứng với các mã số tiêu chuẩn chung mà mọi báo cáo tài chính đều giống nhau. Tôi gửi bạn file YAML, bạn xem chỉnh sửa và tối ưu cho tôi.

- Bạn hãy cho tôi trong code các thông tin sau để chỉ cần chạy lênh “python -m src.p1a\_clean10\_ocr\_bctc --start 8 --end 8” nó sẽ in ra trang 8, còn chạy lênh python -m src.p1a\_clean10\_ocr\_bctc thì nó sẽ in tất cả các file ở inputs.

- yaml “D:\1.TLAT\3. ChatBot\_project\1\_Insurance\_Strategy\configs\p1a\_clean10\_ocr\_bctc.yaml”

- input “D:\1.TLAT\3. ChatBot\_project\1\_Insurance\_Strategy\outputs\p1\_prelight\_ocr\_bctc”

- file output xuất ra “D:\1.TLAT\3. ChatBot\_project\1\_Insurance\_Strategy\outputs\p1a\_clean10\_ocr\_bctc”

- file code chính: D:\1.TLAT\3. ChatBot\_project\1\_Insurance\_Strategy\src\p1a\_clean10\_ocr\_bctc.py

Sau khi xuất ra bởi code p1a\_clean10\_ocr\_bctc.py thì được 1 file .txt và 1 file meta.json.

bạn hãy so sánh các data file .txt và hình ảnh tham chiếu và cho tôi kết quả so sánh đúng bao nhiêu số, sai bao nhiêu số: mã số, thuyết minh, số chỉ mục của danh mục (A, I, II, III, IV, 1,2...), số đầu năm, số cuối năm.

# BẢN PROMT ĐỂ TỐI ƯU CHO a1\_ocr\_text\_only.py