In [ ]:

```python
import fitz  # PyMuPDF, imported as fitz for backward compatibility reasons
file_path = "./sarc.pdf"

mat = fitz.Matrix(300 / 72, 300 / 72)  # sets zoom factor for 300 dpi
doc = fitz.open(file_path)  # open document
# for all pages

num_of_pages = doc.page_count


for page in doc:
    pix = page.get_pixmap(matrix=mat, alpha=False)  # render page to an image
    pix.save(f"page_{page.number}.jpg")  # store image as a PNG
```

In [ ]:

```python
import cv2
import numpy as np
import pytesseract


for i in range(0, num_of_pages):
    img = cv2.imread(f'./page_{i}.jpg')
    lower_yellow = np.array([0, 200, 230]) # yellow
    upper_yellow = np.array([0, 234, 255]) # yellow
    mask = cv2.inRange(img, lower_yellow, upper_yellow)
    result = cv2.bitwise_and(img, img, mask=mask)
    text = pytesseract.image_to_string(result)
    with open('extracted_text.txt', 'a') as file:
        file.write(text)
```