Here’s a step-by-step guide to convert a PDF to a JSON file using Python. I'll break it down into easy-to-follow steps:

Step 1: Install Necessary Libraries

To start, you’ll need to install libraries that allow you to work with PDFs and JSON in Python. Use the following commands:

```
pip install PyPDF2
pip install pdfminer.six
pip install json
```

- `PyPDF2` will help in reading PDF files.
- `pdfminer.six` is another option for extracting text from PDFs.
- `json` is built into Python for working with JSON data.

Step 2: Select Your PDF

Choose the PDF file from which you want to extract content. This can be done by either using the file path directly or allowing the user to input the file path.

Step 3: Create JSON File

Now, we will write a Python script that reads the PDF, extracts text, and then saves it to a JSON file.

Step 4: View and Extract Text to JSON

Once the script runs, you’ll have the content of the PDF extracted and saved as a JSON file. You can view this file and inspect the text content.

- The JSON file will have the extracted text stored under `"extracted_text"`.
- The `"pdf_path"` will store the path to the original PDF file.

Example of the JSON Output

Here’s how the content of the `output.json` will look:

```
{
    "pdf_path": "your_file.pdf",
    "extracted_text": "This is the extracted text from the PDF..."
}
```

This is the basic process for extracting text from a PDF and saving it as a JSON file. You can later modify the script to handle more complex PDFs or to format the output in specific ways (like adding metadata or structured data).