

The PDF Document Loader

Here is an example of a file in .pdf format. This type of file is mostly used in contract, agreement, legal documents, so if you need to use it as a data source, be sure you choose the right Loader and proper Chunking strategies.

Remember

The pdf file is tricky because it can come in different formats as the following examples:

One Column

This is an example of one column pdf document. You can see in this part that it stretches the text from the left to the right as one long single piece.

Two Columns

This is an example of one column pdf document.
You can see in this part that it doesn't stretch the

text from the left to the right as one long single
piece. Instead, it breaks into the new columns

Conclusion

Using pdf files is a challenging task, you have to debug often. Also learn their limitations, and leverage them wisely, so you don't end up fixing a bug until you die!!

Note:

- *Text pdf is feasible, but Image pdf is terrible*
- *Handwritten in the pdf file can bite you in the ass*
- *Scanning documents to create a pdf file is widely used, but this is one of many enemies you have to fight with on a daily basis.*