**🚀 PRÓXIMO PASSO SUGERIDO: Esboço do pipeline de processamento**

Antes de escrever qualquer código ou documento extenso, o ideal é **visualizar o fluxo completo da solução**. Isso inclui **entender bem as etapas e os dados**.

**🎯 Etapa 1 – Definir o pipeline do projeto**

Monte um pequeno **esboço do fluxo**, por exemplo:

📥 PDF recebido →

🔍 OCR (extração de texto) →

🧠 NLP (identificação e classificação das informações relevantes) →

📄 Geração do relatório no modelo .odt →

📤 Exportação para o SEI (fictício)

**🧪 Etapa 2 – Escolher 1 fatura real e fazer teste manual**

Pegue **1 fatura PDF real** (já salva na pasta /data/faturas) e:

* Abra
* Leia com calma
* **Liste quais campos você precisa extrair** para o relatório:  
  Exemplo: período de competência, valor bruto, número da fatura, nome da concessionária etc.
* Depois, **abra um relatório final já pronto** (em PDF ou Word) e veja **como esses dados aparecem no texto**.

→ Esse mapeamento de fatura ➜ relatório é **o coração do projeto**.

**✍️ Etapa 3 – Criar um rascunho inicial do modelo de relatório**

* Pegue o modelo\_relatorio.odt e abra no Word ou LibreOffice
* Marque (com colchetes ou marcações) onde cada informação entrará
  + Ex: {{NOME\_CONCESSIONARIA}}, {{PERIODO\_COMPETENCIA}}
* Isso vai ser a **base para o gerador automático**
* **📁 Etapa 4 – Criar o repositório de scripts**
* Crie arquivos vazios na pasta src/:

touch extrator\_ocr.py

touch classificador\_nlp.py

touch gerador\_relatorio.py

**📆 Etapa 5 – Planejar a 1ª Sprint da licença-capacitação**

| **Data** | **Ação** |
| --- | --- |
| 08–09/08 | Leitura e mapeamento de campos da fatura → relatório |
| 10–11/08 | Criação do protótipo manual no LibreOffice com campos marcados |
| 12–14/08 | Rascunho dos scripts extrator\_ocr.py e gerador\_relatorio.py |
| 15–16/08 | Ajustes, validação, e preparação para 1ª orientação |

Se quiser, posso te montar esse cronograma em PDF ou Markdown.

**✉️ E por fim…**

Quer que eu:

* Gere um modelo visual de *pipeline* (fluxograma)?
* Monte um rascunho dos campos principais de uma fatura?
* Ajude a criar o modelo .odt com variáveis do tipo {{campo}}?