*O que o sistema deve fazer?*

**Módulo 1: Captura e Processamento de Documentos**

* **RF001: Upload Manual de Documentos:** O sistema deve permitir que o usuário faça o upload de um ou mais documentos fiscais simultaneamente. (prefiro tbm usar e-mail, mas trará mais complexidade ao projeto)
  + Formatos suportados: PDF (texto e imagem), JPG, PNG.
  + Deve haver uma fila de processamento visível ao usuário (Ex: "Aguardando", "Processando", "Concluído", "Erro").
* **RF002: Download Automático de Fontes Oficiais:** O sistema deve ter a capacidade de se conectar a portais de Secretarias da Fazenda (SEFAZ) e prefeituras para baixar documentos fiscais. (prefeituras será o mais difícil)
  + Isso exigirá que o usuário configure suas credenciais de acesso (ex: certificado digital A1/A3, login e senha).
  + O sistema deve realizar buscas periódicas e automáticas por novos documentos.
* **RF003: Processamento de OCR:** Para documentos baseados em imagem (PDF de imagem, JPG, PNG), o sistema deve aplicar uma tecnologia de Reconhecimento Óptico de Caracteres (OCR) para converter a imagem em texto bruto.
* **RF004: Extração Inteligente de Dados (IA/LLM):** O sistema utilizará modelos de IA e LLM para:
  + Identificar o tipo de documento (NF-e, NFS-e, CT-e, etc.).
  + Analisar o texto extraído e identificar os campos-chave (CNPJs, datas, valores, etc.), mesmo que estejam em posições diferentes em layouts variados.
  + Calcular um "score de confiança" para cada campo extraído. Campos com baixa confiança devem ser sinalizados para revisão humana.
* **RF005: Interface de Validação e Correção:** O usuário deve ter uma tela onde possa visualizar o documento original lado a lado com os dados extraídos.
  + Nesta tela, ele poderá corrigir ou preencher campos que a IA não conseguiu extrair ou extraiu incorretamente.
  + As correções feitas pelo usuário devem ser usadas como feedback para o sistema (ver RF007).

**Módulo 2: Armazenamento e Aprendizado**

* **RF006: Armazenamento Estruturado:** Os dados extraídos devem ser salvos em um banco de dados de forma estruturada.
  + Cada registro de dados deve manter um vínculo com o arquivo original do documento.
* **RF007: Mecanismo de Aprendizado Contínuo (Machine Learning):** O sistema deve usar as correções manuais (RF005) para retroalimentar e refinar os modelos de extração de dados, melhorando sua precisão ao longo do tempo para novos layouts e formatos. (isso é muito importante para refinar a aprendizagem do modelo)

**Módulo 3: Consulta e Análise**

* **RF008: Interface de Consulta em Linguagem Natural:** O sistema deve oferecer um chat onde o usuário possa fazer perguntas como:
  + "Qual o total gasto com o fornecedor X no último mês?"
  + "Liste todas as notas fiscais de serviço de São Paulo em maio."
  + "Quanto pagamos de ICMS em 2024?"
* **RF009: Geração de Respostas Claras:** O LLM deve interpretar a pergunta, buscar os dados no banco de dados estruturado e apresentar a resposta de forma clara e objetiva, podendo ser em texto, tabelas ou valores numéricos.
* **RF010: Dashboard e Relatórios:** O sistema deve apresentar um painel visual (dashboard) com os principais indicadores (total faturado, principais fornecedores, impostos por período, etc.). (opcional, mas se fizer, vai ser um diferencial, podemos usar o conhecimento do grupo em python para gerar o dashboard)
* **RF011: Exportação de Dados:** O usuário deve ser capaz de exportar os resultados de suas buscas ou relatórios para formatos comuns (CSV, Excel). (opcional, mas empresas sempre adoram exportar dados em Excel)