SISTEMA AGÊNTICO MULTIMODAL PARA CAPTURA INTELIGENTE DE DADOS DE DOCUMENTOS FISCAIS FÍSICOS BRASILEIROS

Nome do Grupo: CRM-IA

Curso: Agentes Autônomos com Redes Generativas do I2A2

Nome	**E-mail**	**Telefone**
José Roberto	aragao@velip.com.br	+5519992069730
Alexandre	alexandrevartuli@gmail.com	+5535997171553
Fernanda	Fernandaave@gmail.com	+17786887129
Rafaelle	rafaellesouzaq@gmail.com	+5538992552170
Kellen	kellenmonteiroferreira@gmail.com	+5512988235882
Brenda	Brendaapm@gmail.com	+5527997486063
José Carlos	carloscouto@outlook.com	+5511997554264

Resumo

Este projeto propõe o desenvolvimento de um sistema agêntico de Inteligência Articial (IA) focado na primeira etapa do processamento de documentos fiscais físicos brasileiros (como DANFE, DACTE e representações imagéticas de NFS-e). Utilizando Modelos de Linguagem Multimodais (MLLMs) e o framework LlamaIndex em Python, o sistema visa realizar a captura de dados desses documentos, implementar mecanismos de verificação para mitigar alucinações e, por m, inserir os dados validados em uma base de dados estruturada. O projeto explorará as capacidades dos MLLMs para lidar com a variabilidade de leiautes e qualidades de imagem, um desafio comum em documentos físicos digitalizados.

1. DESCRIÇÃO DO TEMA ESCOLHIDO

O projeto consiste em desenvolver um sistema de IA para automatizar a captura de dados de documentos fiscais brasileiros que são recebidos em formato físico

(e posteriormente digitalizados) ou como imagens (ex: PDFs de imagem de DANFE, DACTE, NFS-e). O núcleo do sistema utilizará Modelos de Linguagem Multimodais (MLLMs) para "ler" e interpretar esses documentos visuais, extraindo campos de informação chave.

Uma componente crítica do projeto será a implementação de um sistema de verificação para identicar e auxiliar na correção de "alucinações" – informações incorretas ou fabricadas que os MLLMs podem gerar.1 Este sistema de verificação incluirá validações baseadas em regras simples e um conceito de Human-in-the-Loop (HITL) simplificado para os casos mais desafiadores.

O objetivo final era popular um banco de dados com as informações extraídas e validadas, tornando-as prontas para utilizações diversas, como integração com sistemas ERP, análises fiscais, ou auditorias. O foco é estritamente na primeira etapa do ciclo de vida do documento fiscal dentro de uma organização: a entrada e validação inicial dos dados a partir de uma fonte visual.

2. PÚBLICO ALVO

A solução proposta destina-se a:

- Empresas de todos os portes (pequenas, médias e grandes): Especialmente aquelas que ainda recebem um volume considerável de documentos fiscais em papel (que necessitam digitalização) ou em formatos de imagem não estruturados (PDFs de imagem, JPEGs, PNGs).
- Departamentos Fiscais, Contábeis e Financeiros: Profissionais que atualmente dedicam tempo significativo à digitação manual e verificação desses documentos.
- Desenvolvedores e Equipes de TI: Que buscam soluções para automatizar o fluxo de entrada de documentos fiscais, reduzindo a carga operacional e melhorando a qualidade de dados

3. JUSTIFICATIVA

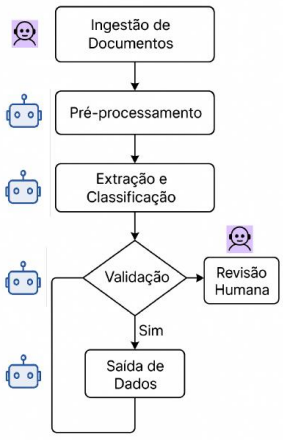
O processamento manual de documentos fiscais físicos ou digitalizados é uma tarefa intensiva em tempo, propensa a erros de digitação e onerosa para as empresas. A automação dessa etapa inicial é crucial para:

a) Redução de Custos Operacionais: Eliminar ou reduzir drasticamente a necessidade de entrada manual de dados.
b) Aumento da Eficiência: Agilizar o fluxo de informações fiscais, permitindo que os dados estejam disponíveis mais rapidamente para as etapas subsequentes do processo fiscal e contábil.
c) Minimização de Erros: Reduzir a incidência de erros humanos que podem levar a problemas de conformidade fiscal, retrabalho e potenciais sanções.
d) Melhoria da Qualidade dos Dados: Garantir que os dados inseridos nos sistemas corporativos sejam mais precisos e confiáveis desde o início.
e) Liberação de Recursos Humanos: Permitir que os colaboradores se concentrem em atividades analíticas e estratégicas de maior valor agregado.

Os MLLMs oferecem um potencial significativo para superar as limitações das tecnologias de OCR tradicionais, especialmente na capacidade de lidar com variações de layout, qualidades de imagem diversas e na extração direta de informações de formatos visuais. A abordagem proposta, que inclui um sistema de verificação, visa tornar o uso de MLLMs mais robusto e confiável para esta aplicação crítica. A estruturação dos dados capturados em um banco de dados agrega valor ao facilitar consultas, análises e integrações futuras.

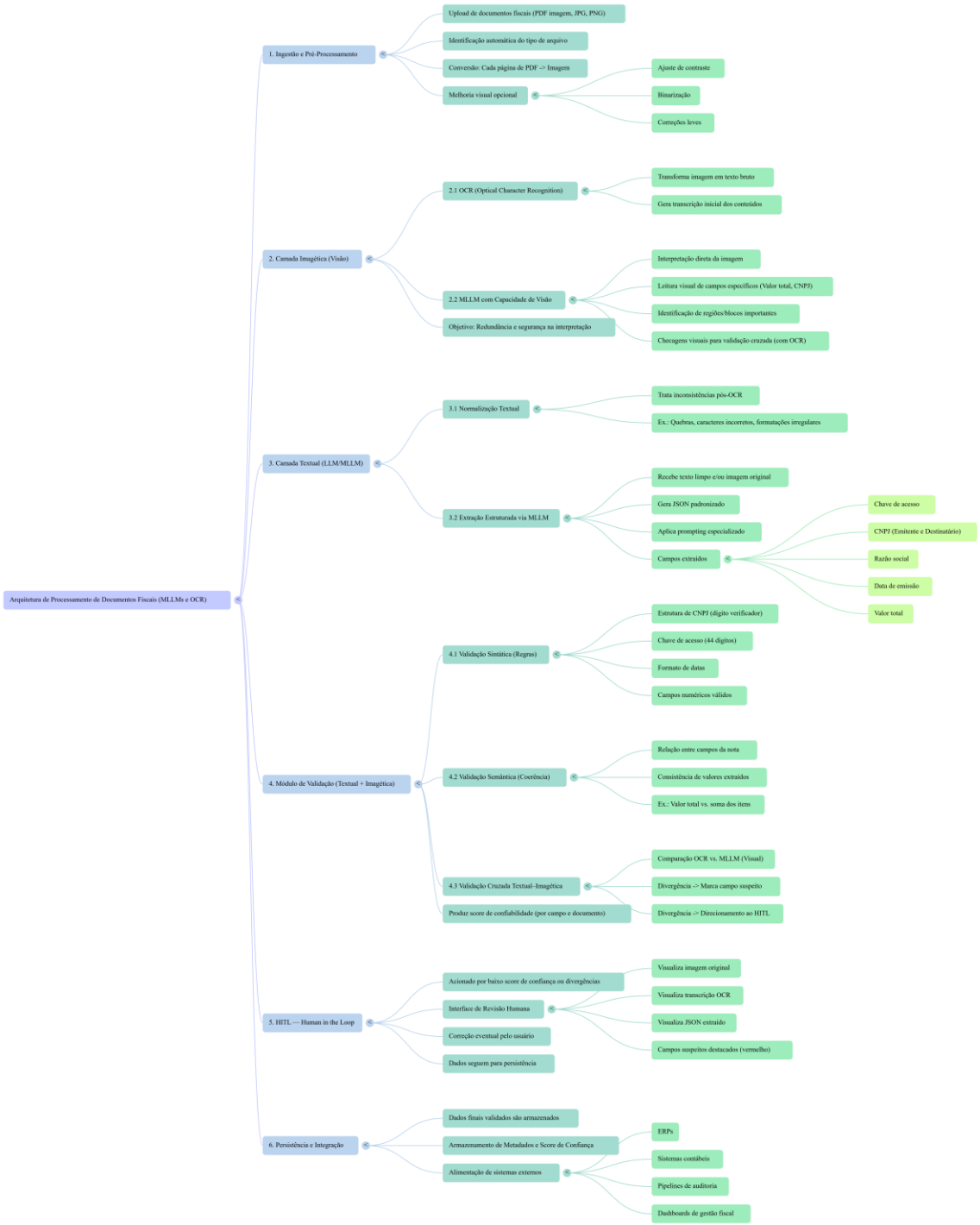
4. DETALHAMENTO DESENVOLVIMENTO

A arquitetura do sistema desenvolvido integra múltiplos componentes especializados para realizar o processamento, interpretação e validação de documentos fiscais recebidos em formato de imagem ou PDF.



Legenda: Arquitetura preliminar

Essa arquitetura combina OCR, Modelos Multimodais (MLLMs) e mecanismos de validação textual e imagética, garantindo que os dados extraídos sejam precisos, confiáveis e adequados para inserção em sistemas corporativos. A seguir, apresenta-se a descrição detalhada da arquitetura:



1. Ingestão e Pré-Processamento

O processo inicia-se com o upload de documentos fiscais em formatos como PDF de imagem, JPG ou PNG. O sistema identifica automaticamente o tipo de arquivo e aplica o pré-processamento necessário. No caso de PDFs, cada página é convertida em imagem. Etapas opcionais de melhoria visual — como ajuste de contraste, binarização ou correções leves — podem ser utilizadas para otimizar a leitura posterior.

2. Camada Imagética (Visão)

Nesta etapa, o sistema emprega dois mecanismos complementares:

2.1 OCR (Optical Character Recognition)

Responsável por transformar a imagem do documento em texto bruto. É aplicado sobre cada página ou recorte relevante, gerando uma transcrição inicial dos conteúdos visuais.

2.2 MLLM com Capacidade de Visão

Modelos multimodais (como GPT-4 Vision, Gemini Vision ou equivalentes) são empregados para interpretar o documento diretamente na imagem. Isto inclui:

- leitura visual de campos específicos (ex.: valor total, CNPJ);
- identificação de regiões ou blocos importantes do layout da nota;
- checagens visuais para validação cruzada do texto derivado do OCR.
- A combinação dessas duas abordagens garante redundância e maior segurança na interpretação dos dados.

3. Camada Textual (LLM/MLLM)

Com base no texto extraído via OCR e nas informações complementares obtidas pela análise visual, a arquitetura utiliza um módulo de normalização e extração estruturada:

3.1 Normalização Textual

Etapa que trata inconsistências comuns pós-OCR, como quebras, caracteres incorretos, separações erradas de palavras e formatações irregulares.

3.2 Extração Estruturada via MLLM

O modelo multimodal recebe o texto limpo e/ou a imagem original e é instruído a gerar um JSON padronizado contendo os principais campos da nota fiscal, como:

- Chave de acesso;
- Cnpj do emitente e destinatário;
- Razão social;
- Data de emissão;
- Valor total.

Esse módulo aplica prompting especializado para garantir consistência e completude dos campos extraídos.

4. Módulo de Validação (Textual + Imagética)

A validação ocorre em três níveis:

4.1 Validação Sintática

Checagens por regras, como:

- Estrutura de CNPJ com dígito verificador;
- Chave de acesso com 44 dígitos;
- Formato de datas;
- Campos numéricos válidos.

4.2 Validação Semântica

Avalia coerências internas, como:

- Relação entre campos da nota;
- Consistência entre os valores extraídos;
- Verificações como "valor total aproximar-se da soma de itens".

4.3 Validação Cruzada Textual–Imagética

Comparação direta entre o valor lido pelo OCR e o valor lido visualmente pelo MLLM na imagem da nota. Quando há divergência, o campo é marcado como suspeito e direcionado ao processo HITL. Essa camada também produz um score de confiabilidade para cada campo e para o documento como um todo.

5. HITL: Human in the Loop

Nos casos em que o score de confiança é baixo ou quando há divergências significativas, o documento é enviado para uma interface de revisão humana. Nessa interface, o usuário visualiza:

- A imagem original;
- A transcrição OCR;
- O JSON extraído;
- Campos suspeitos destacados em vermelho.
- Após a revisão e eventual correção, os dados seguem para persistência.

6. Persistência e Integração

Os dados finais validados são armazenados em banco de dados, juntamente com metadados, status de validação e score de confiança. Essa base pode posteriormente alimentar: ERPs, sistemas contábeis, pipelines de auditoria, dashboards de gestão fiscal.

1. **Frontend / UI**
    - Upload de PDFs/imagens de DANFE/DACTE/NFS-e.
    - Tela de revisão (HITL) para corrigir campos problemáticos.
    - Visualização de histórico de documentos processados.
2. **Ingestão & Pré-processamento de Arquivos**
    - Detecta tipo de arquivo (PDF, imagem).
    - Converte PDF → imagens de página.
    - Aplica ajustes de imagem
3. **Camada "Imagética" (Visão)**
    - **OCR Engine**: extrai texto bruto da imagem.
    - **Analisador de Layout / Qualidade Visual**
        - Verifica se a imagem está legível
4. **Camada Textual (LLM / MLLM)**
    - **Normalizador Textual**: limpa o texto OCR (quebras, caracteres trocados, etc.).
    - **Extrator Estruturado (MLLM/LLM)**:
        - Prompt: "A partir desse texto de nota fiscal, devolva um JSON com chave_de_acesso, cnpj_emitente, cnpj_destinatario, data_emissao, valor_total, etc."
    - **Camada de "Raciocínio"**:
        - Decide quando refazer extração, pedir mais contexto, ou mandar para revisão humana.
5. **Módulo de Validação (Textual + Cruzada)**
    - **Validação Sintática**:
        - Formato CNPJ (regex + dígito verificador).
        - Chave de acesso com 44 dígitos.
        - Datas válidas, valores numéricos coerentes.
    - **Validação Semântica / de Consistência**:
        - CNPJ do emitente = CNPJ presente no cabeçalho
        - Valor total ≈ soma de itens (pelo menos ordem de grandeza).
    - **Validação Cruzada Textual ↔ Imagem (Imagética)**:
        - Comparar o que o OCR captou vs. o que o MLLM "leu" diretamente da imagem (via visão).

        - Ex.: MLLM com visão lê diretamente “Valor Total: 1.234,56” e você compara com o campo vindo do OCR.
    - Define um **score de confiança** para cada campo e para o documento como um todo.
6. **HITL – Human in the Loop**
    - Se score de confiança < limiar, manda o documento pra revisão.
    - Tela mostra:
        - Imagem da nota.
        - Texto OCR.
        - JSON extraído.
        - Campos destacados em vermelho (suspeitos).
7. **Persistência e Integração**
    - **Banco de Dados** (ex.: SQLite ou Postgres).
    - Tabela com:
        - Metadados do documento.
        - Campos extraídos (+ flags de confiabilidade).
    - Hooks para integração futura com:
        - ERP
        - Sistemas contábeis
        - Dashboards.
8. **Monitoramento & Telemetria**
    - Log de:
        - Taxa de sucesso sem intervenção humana.
        - Tipos de erro mais frequentes (CNPJ errado, valor divergente, etc.).
    - Isso alimenta ajustes de prompt, regras e thresholds.

Tabela de campos chave a serem extraídos:

Documento	Campo Chave	Tipo de Dado Esperado	Prioridade
NFS-e (imagem)	Número da Nota	Texto/Número	Alta
NFS-e (imagem)	Código de Verificação (se houver)	Texto/Número	Média
NFS-e (imagem)	CNPJ/CPF do Prestador	Texto (formato)	Alta
NFS-e (imagem)	Razão Social/Nome do Prestador	Texto	Média
NFS-e (imagem)	CNPJ/CPF do Tomador	Texto (formato)	Alta
NFS-e (imagem)	Razão Social/Nome do Tomador	Texto	Média
NFS-e (imagem)	Data de Emissão	Data (DD/MM/AAAA)	Alta
NFS-e (imagem)	Valor Total dos Serviços	Numérico (Moeda)	Alta
NFS-e (imagem)	Discriminação dos Serviços (resumo)	Texto	Baixa
DANFE	Chave de Acesso	Numérico (44 dígitos)	Alta
DANFE	CNPJ do Emitente	Texto (formato CNPJ)	Alta
DANFE	Razão Social do Emitente	Texto	Média

DANFE	CNPJ do Destinatário	Texto (formato CNPJ)	Alta
DANFE	Razão Social do Destinatário	Texto	Média
DANFE	Data de Emissão	Data (DD/MM/AAAA)	Alta
DANFE	Valor Total da Nota	Numérico (Moeda)	Alta
DACTE	Chave de Acesso	Numérico (44 dígitos)	Alta
DACTE	CNPJ do Emitente (Transportador)	Texto (formato CNPJ)	Alta
DACTE	CNPJ do Remetente	Texto (formato CNPJ)	Média
DACTE	CNPJ do Destinatário	Texto (formato CNPJ)	Média
DACTE	Valor Total do Documento	Numérico (Moeda)	Alta

5. REPOSITÓRIO GITHUB

https://github.com/brendaapm/IA2A_Agents

6. CONCLUSÃO

Este projeto demonstra o potencial de agentes e MLLMs para captura de dados fiscais. A experiência reforça desafios como precisão, confiabilidade e engenharia de prompt, e consolida aprendizado prático para implantação de IA em cenários reais de alta criticidade.