# Projeto Final

- "Mantras" do Grupo 1
  - "As atividades são realizadas com pelo menos 2 pessoas"
  - "Quem não quiser fazer código, fará documentação"
  - "Mantenha contato com a sua dupla"
  - Reuniões
    - Google Meet
      - Link
        - https://meet.google.com/landing?pli=1
    - Acompanhamento
      - Semanais
        - Temas
          - Discussão e revisão do trabalho das duplas
          - Dificuldades encontradas
          - Quaisquer outros assuntos
  - "Quem terminar a sua atividade, pode ajudar o parceiro de grupo"
  - "Continue estudando, pois a sua vez de atuar chegará. Veja o cronograma"
- Contexto
  - Cronograma
  - Relatório
- 1 Desafio 3 - Primeira Apresentação
  - Fase 1
    - Responsáveis
      - Walison
      - Valdinei
    - Levantamento de Requisitos
      - - Análise detalhada de tipos de documentos fiscais brasileiros
        - Escolher
          - NF-e
          - Campos da NF-e e NFS-e
      - - Definição de campos obrigatórios para extração
        - Verificar link
          - https://chatgpt.com/share/68462071-dee0-8006-88ad-a79ea9fbbb84
      - - Estudo de regulamentações fiscais aplicáveis
        - ⚠️ **Verificar a norma de regulamentação para os campos que deve ser informados.**
          - **É igual para todos os municípios ?**
            - ⚠️ Será importante para o Agente 2, responsável pela extração e aprendizado
          - **Existe diferença para a norma federal ?**
            - **https://chatgpt.com/share/68465058-7604-8006-9095-35b3772b9157**
          - Link
          - Escolher um padrão que seja adotado pela maioria dos municípios
      - - Coleta de amostras de documentos para treinamento"
        - Arquivos
          - Exemplos em PDFs e imagens
          - Serão usados para testes de uploads
        - Download de arquivos
          - Para testes de acesso por downloads
            - Verificar links
              - Qual são as formas de acesso ?
                - Ex:
                - https://portaldatransparencia.gov.br/download-de-dados/notas-fiscais/202504
                - Pode-se escolher o ano e mês pela URL
            - Servidores FTP
            - SEFAZ ?
            - Inicialmente, podemos começar por esses, depois aguardar os campos utilizados em municípios
  - Fase 2
    - Desenvolvimento OCR + NLP - Agente 2
      - - Implementação do motor de OCR otimizado
        - Também atenderá com a ajuda da LLM (Pesquisa internet) e pesquisa nos documentos
        - Responsável
          - Luciane
          - Daniel
        - Tessaract + OpenCV
          - https://chatgpt.com/share/6842f1b5-3024-8006-a188-e84183f25ef2
      - - Desenvolvimento de algoritmos de NLP para extração de campos
        - Responsável
          - Antonio
      - - Criação de templates para diferentes tipos de documentos e aprendizado
        - ⚠️ **Criar um template de prompt para NF e contendo a pergunta do usuário**
          - Técnica Few-Shots
          - Responsáveis
            - Antonio
            - Ana
  - Fase 3
    - Testes iniciais
      - ⚠️ Desenvolvedores devem explicar aos testadores como executar os códigos desenvolvidos no Colab.
      - - Validação com conjunto de documentos de teste
        - Elaborar cenários de teste
          - Documento
            - Eduarda
        - Relatar bugs
        - Relatar precisão de acerto
          - Depende
        - Depende
        - Responsáveis
          - Aurélio
          - Matheus
          - Eduarda
      - - Ajustes de precisão e correção de bugs
        - Responsáveis
          - Antonio
          - Ana
          - Luciane
          - Daniel
      - - Documentação técnica inicial
        - Acrescentar a documentação atual
          - Responsáveis
            - Lorrane
            - Eduarda
  - Fase 4
    - Atividades Finais
      - - Apresentação de métricas de precisão alcançadas
      - - Planejamento de melhorias
        - Responsáveis
          - Ana
          - Antonio
      - - Demonstração do MVP Funcional
        - ⚠️ Marcar reunião extra no dia 11 para apresentar
          - Responsável
            - Ana
        - Responsáveis
          - Ana
          - Antonio
          - Eduarda
- Descrição dos Agentes
  - Solicitação
  - Subtópico
  - Arquitetura
    - Agente 1 — Recuperação de documentos fiscais
    - Agente 2 — Utilizar OCR
    - Agente 3 — Interação com o usuário sobre os documentos
    - Frontend — Valdinei
- Depende
- Começar em paralelo
- 2 Desafio 4 - Segunda Apresentação
  - Fase 5
    - Interface - Agente 1 + Frontend
      - - Desenvolvimento de APIs para download de arquivos
        - Responsáveis
          - Daniel
          - Anna
      - - Criação de interface web
        - Responsáveis
          - Valdinei
          - Matheus
      - - Implementação de funcionalidades de interação com o usuário. Chat
        - Responsáveis
          - Antonio
          - Luciane
      - - Testes de download de dados com sites reais
        - Responsáveis
          - Aurélio
          - Lorrane

## Em que podemos trabalhar?

- **Extração de Dados:**
  - Recuperar documentos fiscais em fontes conhecidas
  - Utilizar OCR (Reconhecimento Óptico de Caracteres) em conjunto com NLP (Processamento de Linguagem Natural) para extrair dados relevantes dos documentos:
    - Informações do emitente e destinatário
    - Itens da nota (descrição, quantidade, valor)
    - Impostos (ICMS, IPI, PIS, COFINS)
    - CFOP, CST e outros códigos fiscais
  - Desafios:
    - Como tornar o agente capaz de se adaptar a diferentes layouts e formatos de documentos.
    - Como se adaptar às mudanças legais (ex. IVA)