



PROPOSTA DE PROJETO PARA O CURSO DE
AGENTES AUTÔNOMOS COM REDES GENERATIVAS
DO I2A2



ChatFiscal: Agente Inteligente para Extração e Análise de Documentos Fiscais no Contexto Brasileiro

(2025.1)



PROPOSTA DE PROJETO PARA O CURSO DE
AGENTES AUTÔNOMOS COM REDES GENERATIVAS
DO I2A2



1. Identificação do projeto discente

Título do projeto: ChatFiscal: Agente Inteligente para Extração e Análise de Documentos Fiscais no Contexto Brasileiro

Professor responsável: Celso Azevedo

Curso(s): Agentes Autônomos com Redes Generativas

Nome do grupo: Alquimistas Digitais

Integrantes	E-mail
Amanda Praça	amandapraca@gmail.com
Geovane Santos Silva	3gssilva@gmail.com
Igor Araujo de Mattos	yigor88mattos@gmail.com
Izaque Silva	izaque.silva@marjanfarma.com.br
Luana Rodrigues	luanadaconc@hotmail.com
Márcia Pereira de Jesus	marcia.pj@gmail.com
Paula Monteiro	paulamonteiro15@gmail.com
Tadeu de Bonis	tadeu.bonis@gmail.com

2. Resumo de indicadores

Total de alunos no projeto	08
Número total de atividades/eventos/ações realizados	16
Houve parceria com profissionais fora da instituição? Se sim, com quantos?	0
Houve parceria com empresas ou entidades privadas? Se sim, com quantos?	0
Houve parceria com setor público? Se sim, com quantos?	0
Houve parceria com 3º setor? Se sim, com quantos?	0
Houve produção de material didático? Se sim, quantos?	0



PROPOSTA DE PROJETO PARA O CURSO DE
AGENTES AUTÔNOMOS COM REDES GENERATIVAS
DO I2A2



DADOS DO PROJETO DISCENTE

3. NOME DO PROJETO:

ChatFiscal: Agente Inteligente para Extração e Análise de Documentos Fiscais no Contexto Brasileiro

4. ALUNOS ENVOLVIDOS:

Amanda Praça – Coordenação geral, criação do escopo do desafio, criação de tabelas e gráficos para o relatório e validação do material.

Geovane Santos Silva – Comparação técnica e revisão final.

Igor Araujo de Mattos – Levantamento de desafios e elaboração do relatório.

Izaque Silva – Avaliação de ferramentas e análise dos temas.

Luana Rodrigues – Estudo de ferramentas e criação do Pitch Deck.

Márcia Pereira de Jesus – Mapeamento técnico e apoio na apresentação.

Paula Monteiro – Formatação e validação do material.

Tadeu de Bonis – Análise comparativa e revisão de slides.

5. OBJETIVO GERAL

O presente projeto visa desenvolver um **agente autônomo inteligente** capaz de realizar a **extração e análise automatizada de dados fiscais** em documentos físicos e eletrônicos (XMLs e PDFs de NFe, CTe, NFC-e, MDF-e), com aplicação de **OCR** e **NLP**, otimizando o processo de escrituração fiscal e contábil, reduzindo erros e melhorando o tempo de fechamento tributário.

O projeto visa desenvolver um agente autônomo inteligente para a extração e análise automatizada de dados fiscais, utilizando uma abordagem robusta baseada em **LangChain**, **LlamaIndex** e **Pydantic**. Essas tecnologias possibilitam a ingestão eficiente de documentos fiscais (PDFs, XMLs), aplicação de OCR para extração de texto e validação estrutural dos



PROPOSTA DE PROJETO PARA O CURSO DE
AGENTES AUTÔNOMOS COM REDES GENERATIVAS
DO I2A2



dados conforme a legislação brasileira. O LlamalIndex facilita a extração de informações complexas de layouts tributários, enquanto o Pydantic garante a padronização dos dados, reduzindo erros e assegurando conformidade fiscal. A solução será integrada aos principais ERPs do mercado, como **Domínio, Alterdata e Protheus**, permitindo um fluxo automatizado de escrituração contábil e geração de insights tributários. Além disso, a arquitetura modular permite adaptação às mudanças legais e aprimoramento contínuo do sistema.

Tecnologia	Pontos Fortes	Pontos Fracos
 LangChain	Ampla gama de ferramentas e integrações para processamento e documentos	Pode ser excessivo para casos de uso mais simples
 LlamalIndex	Excelência na extração de informações de tabelas e layouts complexos	Menos ferramentas de pipeline em comparação com LangChain
 Pydantic	Validação e estruturação robusta de dados de saída	Curva de aprendizado para usar corretamente os modelos
 PandasAI	Análise de dados e preparação para integração em sistemas	Pode exigir extensa pré-processamento para dados não estruturados

Tecnologias envolvidas:

- **Modelos de linguagem (LLMs)** – Utilizados para interpretar e extrair informações fiscais de documentos, garantindo respostas precisas sobre obrigações tributárias.



PROPOSTA DE PROJETO PARA O CURSO DE
AGENTES AUTÔNOMOS COM REDES GENERATIVAS
DO I2A2



- **Base de conhecimento jurídica e fiscal** – Referenciada para validar a conformidade dos dados extraídos, assegurando alinhamento com normas tributárias e regulamentações brasileiras.
- **Técnicas de RAG (Retrieval-Augmented Generation)** – Aplicadas para recuperar e atualizar automaticamente leis e regulamentos fiscais, permitindo análises mais assertivas.
- **Reconhecimento fiscal automatizado** – Implementado via **Pydantic** e **PandasAI**, garantindo a correta estruturação dos dados, padronização de alíquotas e códigos fiscais (ICMS, PIS, COFINS, CFOP, CST, NCM) antes da integração com sistemas contábeis.

Tecnologia	Função no Projeto
LangChain	Ingestão e pré-processamento de documentos fiscais
LlamalIndex	Extração avançada de informações tributárias
Pydantic	Validação e estruturação de dados fiscais
PandasAI	Análise e otimização de dados contábeis
OCR (Tesseract/PaddleOCR)	Extração de texto de documentos escaneados

6. OBJETIVOS ESPECÍFICOS

- Extrair automaticamente informações fiscais de arquivos XML e PDF (DANFE), como informações do emitente e destinatário, itens da nota, impostos (ICMS, IPI, PIS, COFINS), CFOP, CST e outros códigos fiscais;
- Padronizar e validar dados conforme legislação (ICMS, PIS, COFINS, CFOP, CST, NCM);
- Reduzir o esforço e o tempo gastos na escrituração manual;
- Gerar respostas automatizadas para questões tributárias com base nos dados extraídos.



PROPOSTA DE PROJETO PARA O CURSO DE
AGENTES AUTÔNOMOS COM REDES GENERATIVAS
DO I2A2



- Desenvolver um pipeline inteligente com validações usando **Pydantic**.
- Integrar o agente aos ERPs e sistemas contábeis utilizados no mercado.
- Utilizar ferramentas **open-source** com alta escalabilidade e baixo custo.

7. PÚBLICO-ALVO

Segmento	Aplicação Direta
Pequenas e Médias Empresas	Redução de custos com automação do recebimento fiscal
Escritórios de Contabilidade	Processamento de grande volume de documentos para múltiplos clientes
Grandes Empresas com ERP	Integração de documentos com sistemas como Protheus, Domínio, Alterdata

8. JUSTIFICATIVA

A **complexidade do sistema tributário brasileiro** e o **alto volume de documentos fiscais** demandam soluções automatizadas e inteligentes para garantir eficiência e conformidade. O agente proposto resolve um problema prático e universal: **a entrada manual de dados fiscais**, que consome tempo, é suscetível a erros e impacta diretamente a área contábil.

Com a iminente reforma tributária e possíveis mudanças como o **IVA/IBS**, um sistema dinâmico e adaptável também se torna essencial.

O sistema tributário brasileiro é um dos mais complexos do mundo, exigindo que empresas dediquem **1.501 horas por ano** para cumprir obrigações fiscais, enquanto a média global é de **159 horas**. A burocracia excessiva impacta diretamente a produtividade e a gestão financeira, gerando um prejuízo estimado de **R\$1.7 trilhão por ano**, equivalente a **20% do PIB nacional**.

Além disso, escritórios de contabilidade enfrentam desafios como **mudanças constantes na legislação, alto risco de erros fiscais e dificuldades na integração de dados com ERPs**.

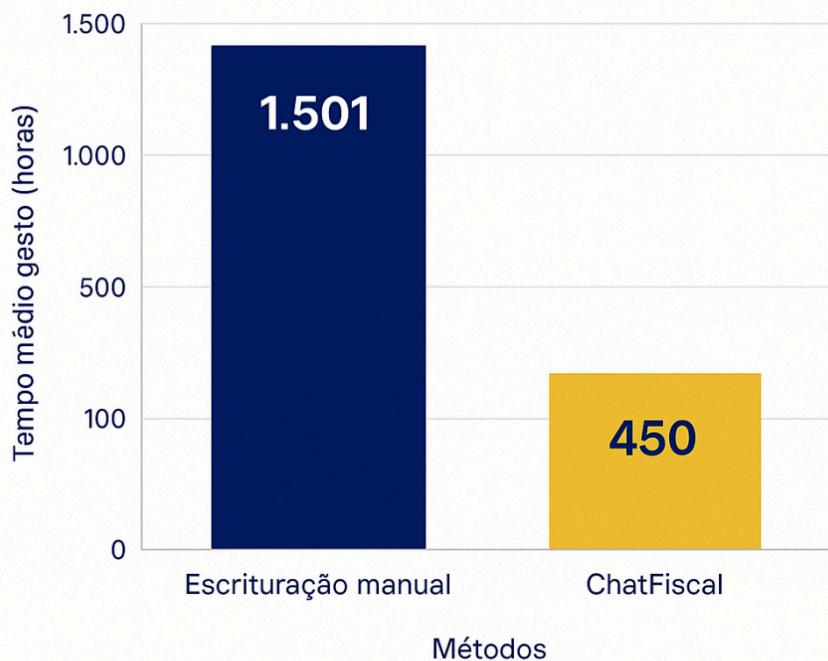


Pequenas e médias empresas, que dependem desses serviços, sofrem com **a falta de automação e a necessidade de processos manuais demorados.**

Com base em estudos de automação fiscal, estima-se que soluções inteligentes podem reduzir **até 70% do tempo gasto na escrituração fiscal**. Isso significa que processos que tradicionalmente demandam **1.501 horas anuais** poderiam ser reduzidos para cerca de **450 horas/ano** com a implementação de sistemas automatizados.

O **ChatFiscal** surge como uma solução inovadora para minimizar esses problemas, automatizando a **extração e validação de documentos fiscais**, garantindo conformidade tributária e **reduzindo significativamente o tempo de escrituração manual**.

Impacto da Automação Fiscal na Escrituração Contábil



9. METODOLOGIA



PROPOSTA DE PROJETO PARA O CURSO DE
AGENTES AUTÔNOMOS COM REDES GENERATIVAS
DO I2A2

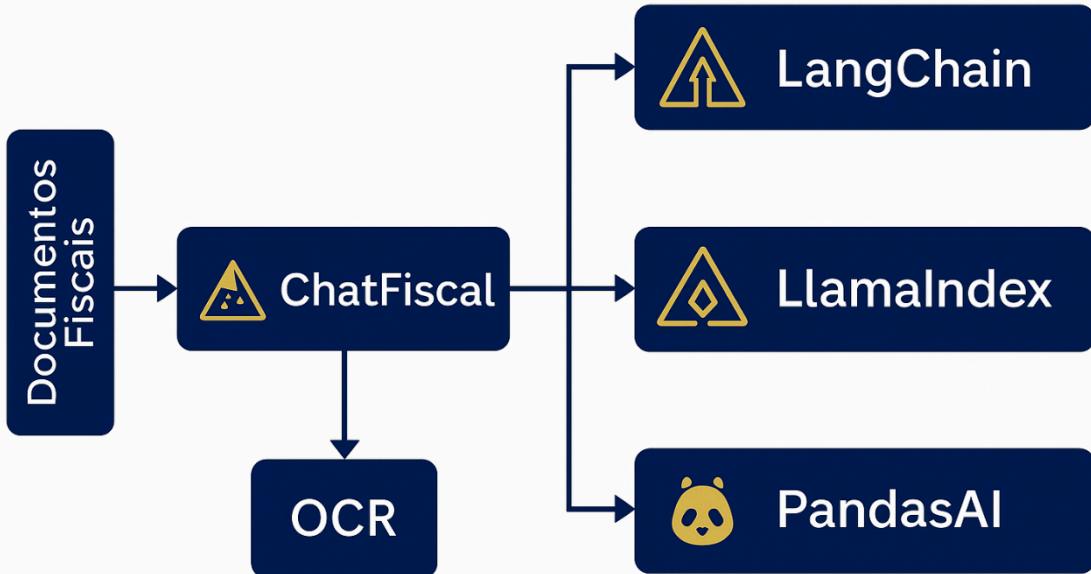


9.1 Etapas Técnicas

Etapa	Descrição
Levantamento e estudo dos documentos fiscais	Tipos, layouts, campos relevantes
Implementação do OCR	Uso de Tesseract ou PaddleOCR para DANFEs em PDF
Processamento com NLP + LangChain	Extração inteligente de entidades fiscais e normalização
Modelagem com Pydantic	Validação de estruturas (CFOP, CST, impostos)
Integração com ERP via n8n ou APIs	Conexão com Protheus, Domínio, Alterdata

Fluxo de Processamento de Dados Fiscais





9.2 Tecnologias Utilizada

- **OCR:** Tesseract/PaddleOCR
- **NLP e Agentes:** LangChain, LlamalIndex
- **Processamento Inteligente de Dados:** PandasAI
- **Validação e Schemas:** Pydantic
- **Orquestração e Integração:** n8n, LangFlow
- **Back-end:** Python
- **ERP/Contábil:** Domínio, Protheus, Alterdata

10. CRONOGRAMA QUE PODE SER USADO PARA A IMPLANTAÇÃO DA SOLUÇÃO

Fase	Duração Estimada	Entregável
Estudo inicial e levantamento técnico	Semana 1	Quadro comparativo de documentos e regras fiscais



PROPOSTA DE PROJETO PARA O CURSO DE
AGENTES AUTÔNOMOS COM REDES GENERATIVAS
DO I2A2



Estruturação do pipeline OCR/XML	Semanas 2 e 3	Extração e transformação de dados fiscais
Implementação com PandasAI	Semana 4	Perguntas diretas sobre os dados (CSV/XML)
Integração com LlamalIndex	Semana 5	Consultas avançadas e indexação
Validação e testes com Pydantic	Semana 6	Regras fiscais e checagem de estrutura
Integração ERP com n8n	Semana 7	Automação de envio/importação
Ajustes finais e documentação	Semana 8	Documentação + Manual de uso

11. RESULTADOS ESPERADOS

- Redução no tempo de digitação manual de dados fiscais;
- Validação fiscal automática de códigos e alíquotas;
- Geração de arquivos estruturados (JSON, CSV, API);
- Integração funcional com ERPs contábeis;
- Estrutura escalável e adaptável a mudanças legais futuras.

12. SUSTENTABILIDADE DO PROJETO

- Agente projetado para se adaptar a **novos layouts e legislações** tributárias;
- Possibilidade de **treinamento contínuo** do sistema para novos tipos de notas;
- Expansão para **validação fiscal e auditoria automatizada**.

13. FERRAMENTAS E TECNOLOGIAS SUGERIDAS

Categoria	Ferramenta/Framework
OCR	Tesseract / PaddleOCR
IA/NLP	LangChain, LlamalIndex, PandasAI
Validação	Pydantic
Orquestração	LangFlow, n8n

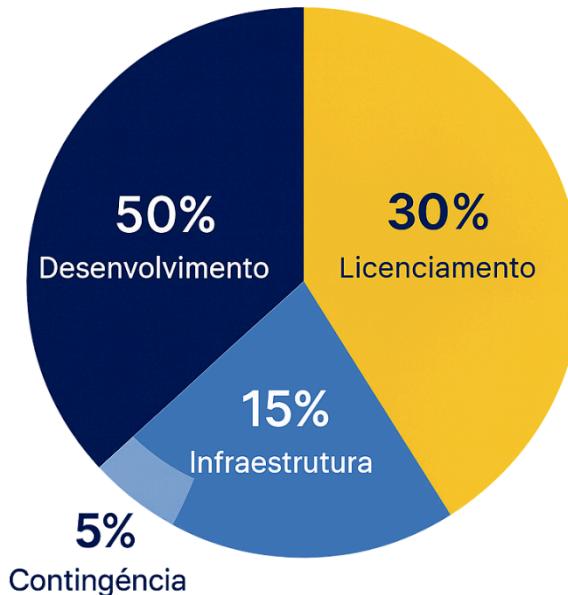


PROPOSTA DE PROJETO PARA O CURSO DE
AGENTES AUTÔNOMOS COM REDES GENERATIVAS
DO I2A2



ERP/Contabilidade	Domínio, Protheus, Alterdata
Linguagens	Python, YAML, JSON, XML
Hospedagem	Docker, FastAPI, Hugging Face Spaces
14. CUSTOS ESTIMADOS	
Item	Valor Estimado (R\$)
Infraestrutura (Servidor, OCR)	R\$ 5.000 a R\$ 7.000
Desenvolvimento Técnico	R\$ 30.000 a R\$ 35.000
Integração com ERP/API	R\$ 15.000 a R\$ 20.000
Treinamento e documentação	R\$ 3.000
Total Aproximado	R\$ 53.000 a R\$ 65.000

Projeção de Custos do Projeto





PROPOSTA DE PROJETO PARA O CURSO DE
AGENTES AUTÔNOMOS COM REDES GENERATIVAS
DO I2A2



Referências

BRASIL. **Receita Federal do Brasil.** Portal Nacional da Nota Fiscal Eletrônica – NFe. Disponível em: <https://www.nfe.fazenda.gov.br>. Acesso em: 07 jun. 2025.

BRASIL. **Lei Complementar nº 123, de 14 de dezembro de 2006.** Institui o Estatuto Nacional da Microempresa e da Empresa de Pequeno Porte. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/LCP/Lcp123.htm. Acesso em: 07 jun. 2025.

BRASIL. **Lei Complementar nº 116, de 31 de julho de 2003.** Dispõe sobre o Imposto Sobre Serviços de Qualquer Natureza. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/LCP/Lcp116.htm. Acesso em: 07 jun. 2025.

LANGCHAIN. **LangChain Documentation.** Disponível em: <https://www.langchain.com>. Acesso em: 07 jun. 2025.

PYDANTIC. **Pydantic Documentation.** Disponível em: <https://docs.pydantic.dev>. Acesso em: 07 jun. 2025.

SEFAZ-SP. **Secretaria da Fazenda do Estado de São Paulo.** Documentação e exemplos de Notas Fiscais. Disponível em: <https://portal.fazenda.sp.gov.br>. Acesso em: 07 jun. 2025.

GÉRON, Aurélien. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems.* 2. ed. Sebastopol: O'Reilly Media, 2019.

RUSSELL, Stuart; NORVIG, Peter. *Artificial Intelligence: a modern approach.* 3. ed. Upper Saddle River: Prentice Hall, 2010.

HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. **Long Short-Term Memory.** *Neural Computation*, v. 9, n. 8, p. 1735–1780, 1997. DOI: 10.1162/neco.1997.9.8.1735.



PROPOSTA DE PROJETO PARA O CURSO DE
AGENTES AUTÔNOMOS COM REDES GENERATIVAS
DO I2A2



HOEPERS, Maria Claudia. O impacto da carga tributária e da burocracia no desenvolvimento econômico: uma análise de 165 países, meio século de dados e lições para o Brasil. **Fundação Getulio Vargas**, 2022.

NASCIMENTO, Beatriz Souza do. O sistema tributário brasileiro: obrigações acessórias e o aumento da burocratização. **Universidade Federal do Rio de Janeiro**, 2023.

SOUSA, Carlos Eduardo Vivardo de. Problemas que os escritórios contábeis de micro e pequenas empresas enfrentam: estudo de multicasos na cidade de São Paulo. **Universidade de Mogi das Cruzes**, 2025.

LOGITHINK. Automação Fiscal da Logithink reduz 70% do tempo na gestão de documentos fiscais. Disponível em: Logithink. Acesso em: 09 jun. 2025.

BANCO MUNDIAL. Tempo médio gasto por empresas brasileiras com obrigações fiscais. Disponível em: Contábeis. Acesso em: 09 jun. 2025.