

Assistente Consultor Especializado baseado em LLMs para Suporte à Tomada de Decisão Gerencial - SAGE

Nome do Grupo: **Grupo 294**

Participantes do Grupo:

Nome	E-mail	Telefone
Larissa	larissagarcialp.lg@gmail.com	+5591981623492
Filipe	fanmello@gmail.com	+5531984204490
Jair	jairmilani21@gmail.com	+5561982745892
Wesley	mdisolucoestecnologia@gmail.com	+5513982296448
Akila	akylaaquino@hotmail.com	+5588981904521
Samuel	fisica141@gmail.com	+5524988790300
Isis	isis.coelho@meta.com.br	+5555996718633
Daniel	montanha78@gmail.com	+55 11 92000-7175
Eduardo	eduardobessa10@gmail.com	+55 11 99509-1781

Este projeto propõe o desenvolvimento de um Assistente Consultor Especializado denominado SAGE (Sistema de Apoio à Gestão Estratégica), alimentado por modelos de linguagem de larga escala (LLMs), com foco em fornecer apoio inteligente para gestores em informações sobre as notas fiscais, dúvidas e decisões estratégicas. A solução será voltada a áreas críticas da gestão como contabilidade, tributação, análise financeira e compliance, oferecendo respostas contextualizadas, explicações técnicas e sugestões de ações estratégicas.

Visão Geral

Desenvolver um assistente virtual especializado apoiado por Modelos de Linguagem de Grande Escala (LLMs) com Retrieval-Augmented Generation (RAG), capaz de responder de forma confiável a questões sobre notas fiscais eletrônicas (CFOP, CST, NCM, tributos estaduais e federais) e, além de, temas gerais de contabilidade (lançamentos, demonstrações, classificação de despesas). A solução será disponibilizada via Telegram, explorando sua dinâmica de bot, suporte a mensagens de voz e familiaridade dos usuários.

Público-Alvo

O público-alvo principal são gestores empresariais, tomadores de decisão, diretores financeiros, analistas contábeis e empreendedores que necessitam de apoio ágil, confiável e especializado para tomar decisões com base em dados e normas atualizadas.

Justificativa do Tema Escolhido

A complexidade crescente do ambiente tributário e contábil brasileiro — especialmente no que tange à emissão, classificação e interpretação de notas fiscais eletrônicas e às constantes mudanças na legislação — impõe desafios significativos aos gestores. Além disso, a necessidade por decisões estratégicas rápidas e bem fundamentadas exige acesso imediato a informações precisas e atualizadas.

Este projeto propõe uma solução inovadora baseada em LLMs (Large Language Models) integrados com a técnica de RAG (Retrieval-Augmented Generation), que permite recuperar informações diretamente de bases legais e normativas, garantindo maior precisão e confiabilidade nas respostas. Como diferencial, o agente será acessível via Telegram, plataforma que permite interações dinâmicas, uso de comandos de voz e fácil integração com o cotidiano dos usuários.

O assistente responderá a dúvidas específicas sobre notas fiscais, abrangendo aspectos como CFOP, CST, NCM e tributações estaduais/federais. Como valor agregado, o sistema também fornecerá orientações e esclarecimentos sobre temas gerais da contabilidade, como lançamentos contábeis, balanços e classificação de despesas.

A proposta agrega valor ao público-alvo ao promover:

- Redução drástica do tempo gasto na busca por informações confiáveis.
- Suporte qualificado à tomada de decisões, com base em dados normativos e atualizados.
- Democratização do conhecimento técnico, com acesso facilitado por linguagem natural e voz.
- Maior acessibilidade, com uma interface familiar (Telegram) e usabilidade centrada no usuário.

Objetivo

1. Fornecer respostas precisas sobre NF-e e tributos em < 3 s.
2. Disponibilizar explicações contábeis complementares sob demanda.
3. Alcançar ≥ 90 % de precisão em benchmark com especialistas.
4. Garantir UX fluida via Telegram, incluindo speech-to-text e text-to-speech.

Escopo Funcional

Módulo	Funções-Chave
Consulta NF-e	Decodificar XML, explicar CFOP/CST, validar regras fiscais
Assistente Contábil	Guiar lançamentos, interpretar DRE, BP e DMPL
Interface Telegram	Mensagens de texto, áudio, quick-actions (/help, /nfe, /conta)
RAG	Buscar e citar legislação (Códigos, Manuais, Ato Cotepe)
Feedback & Aprendizado	Coletar avaliações de utilidade e atualizar base

Arquitetura Técnica

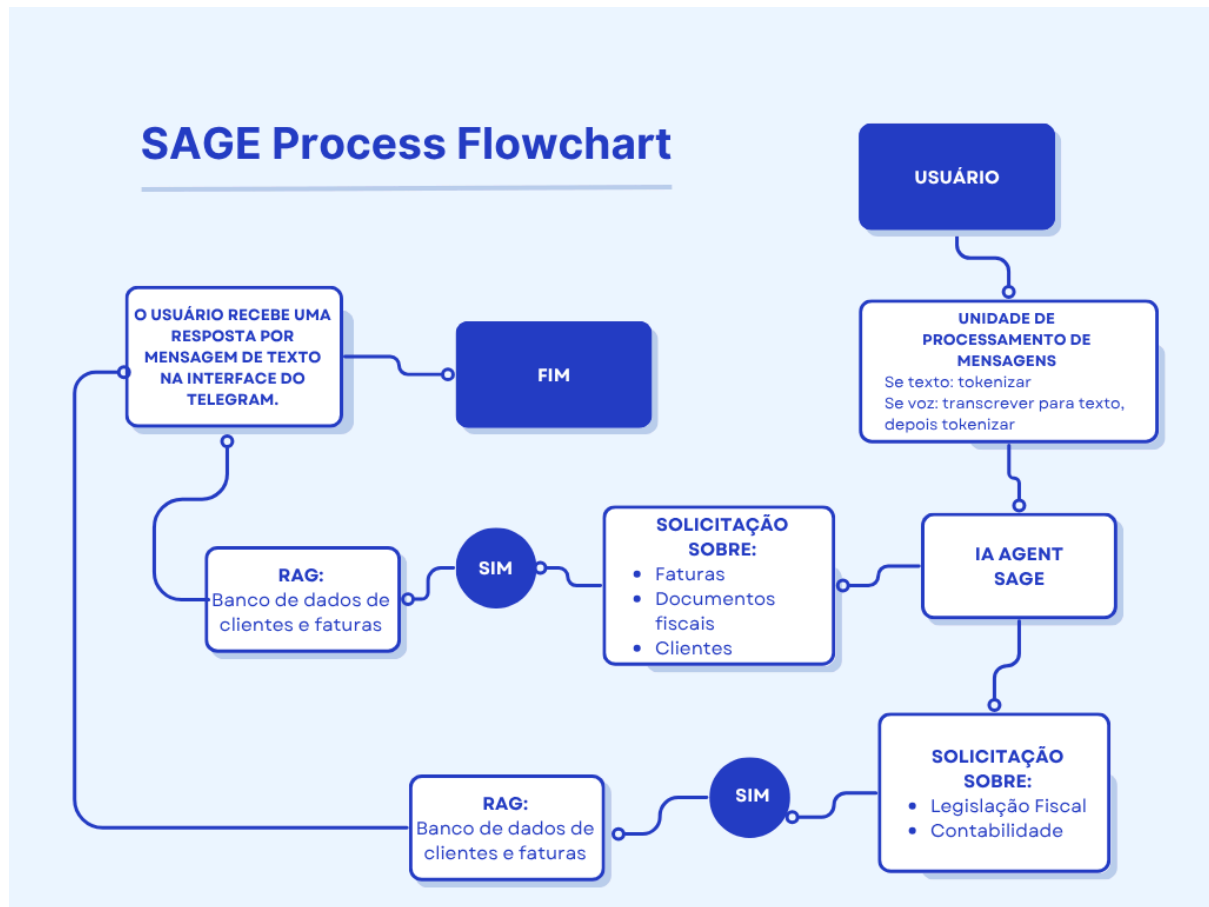


Figura 1 - Fluxo Geral

- Orquestração: n8n (workflows low-code)
- Modelos LLM: OpenAI GPT-4o (respostas em linguagem natural)
- Embeddings: Google Vertex AI (Text Embeddings API) — geração de vetores para documentos.
- Banco de Dados: PostgreSQL + pgvector pelo Supabase (armazenamento de vetores e dados transacionais)
- Hospedagem: VPS (Ubuntu) com Docker + Docker Compose

Proposta Preliminar de Desenvolvimento

Etapas do Projeto:

Etapa	Descrição
Levantamento de Requisitos	Identificação das principais dúvidas e fluxos de decisão dos usuários.
Curadoria de Conteúdo	Relação de Notas fiscais e seleção de bases legais, normativas contábeis e dados financeiros.
Treinamento do Assistente	RAG do LLM com dados contextuais e exemplos reais.
Integração com Aplicativo de Mensagem	Integração com o Telegram permitindo solicitações por texto e áudio do usuário ao assistente.
Validação e Testes	Verificação da acurácia das respostas e da experiência do usuário através de testes monitorados por humanos..

Cronograma

Mês	Etapa	Entregáveis principais
1	Planejamento e curadoria	Requisitos, fontes e estrutura de dados
1	Embeddings e base vetorial	Vetorização de documentos e banco Superbase
2	Desenvolvimento	Workflows n8n, integração Telegram, RAG
2	Testes e validação	A/B test, validação com especialistas
3	Publicação e monitoramento	Deploy, feedbacks, métricas de desempenho

Desafios Identificados e Estratégias

Desafio	Estratégia de Enfrentamento
Respostas imprecisas	Validação de conteúdo, curadoria contínua.
Garantia de Qualidade das Informações	Uso de fontes verificadas, revisão humana contínua e filtros de confiança no output do LLM.
Maximização da Experiência do Usuário	Visão centrada no usuário, utilizando aplicações já usuais como por exemplo aplicativo de mensagens. Objetivando o uso

	de uma linguagem didática, clara e objetiva.
Dificuldade de adoção	Uso de Telegram, interface por voz e tutoriais integrados.
Atualizações legais frequentes	Monitoramento de mudanças normativas e atualização da base.
Vazamento de dados sensíveis	LGPD, criptografia e anonimização de registros.

Modelo de Interação do usuário

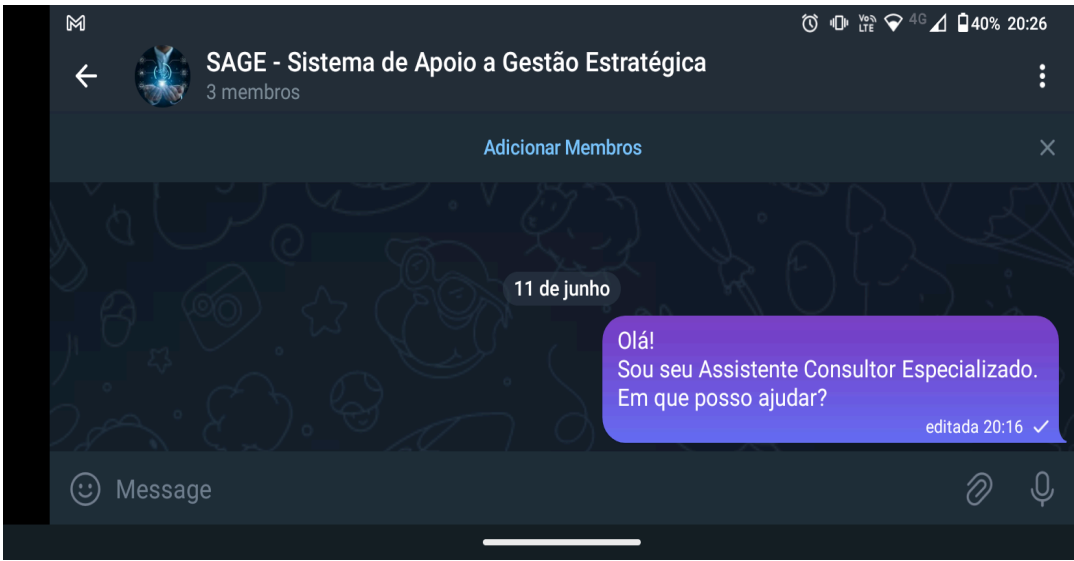


Figura 2 - Interação do usuário via Telegram

Conclusão

O projeto SAGE representa uma união estratégica entre tecnologia de ponta e necessidades práticas do setor empresarial. Utilizando LLMs, RAG, Telegram e arquitetura orientada a dados, ele propõe um salto de eficiência e confiabilidade na gestão de informações fiscais e contábeis.