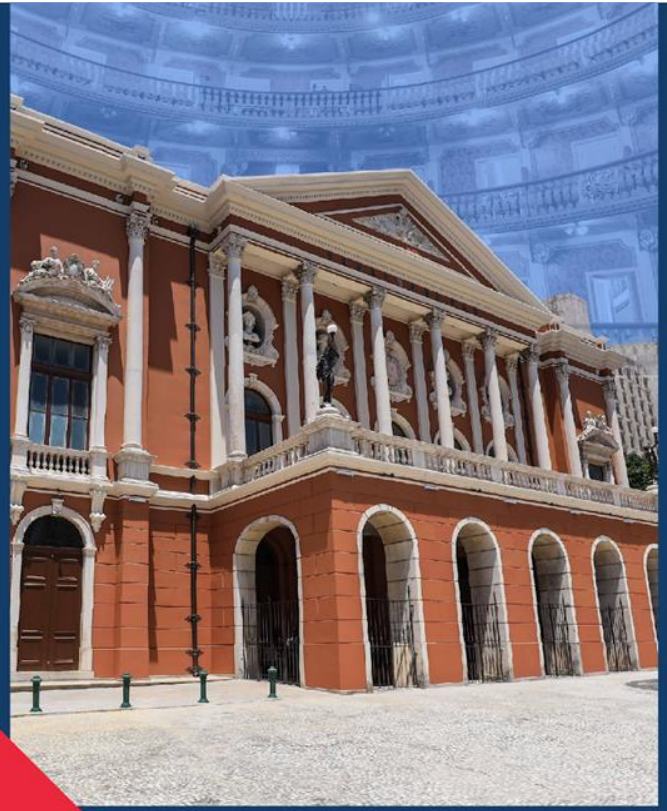


USO DE INTELIGÊNCIA ARTIFICIAL NO ATENDIMENTO TRIBUTÁRIO - SEFAZ/SC

Rossana da Cunha Flores, PhD

Engenheira de Software
@CIASC

Omar Roberto Afif Alemsan

Coordenador do Sistema de Administração Tributária de SC
@SEFAZ-SC

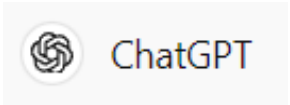
Março/2025

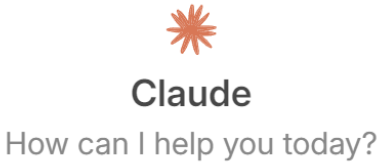
Agenda

- IA Generativa
- Contexto
- O caso de uso da CAF
- Avaliação e Métricas utilizadas
- Primeiros resultados
- Conclusão e Próximos passos

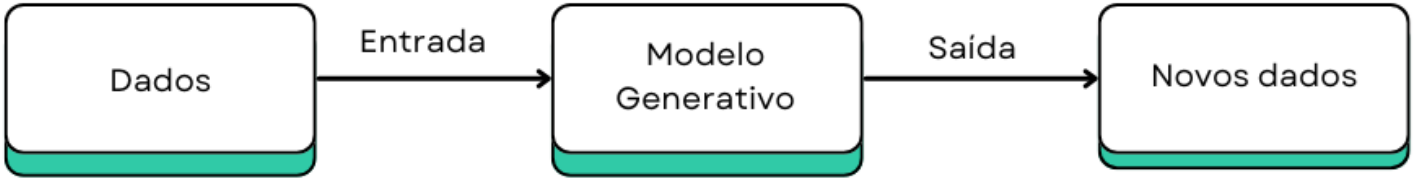
IA GENERATIVA

A Inteligência Artificial Generativa é um ramo da inteligência artificial que se concentra na criação de conteúdo novo e original, como texto, imagens ou outros meios, em resposta a comandos (TERUMI, 2024; CASELI et al., 2024). Alguns exemplos de aplicações são: ChatGPT da OpenAI, Claude da Google, entre outros.





MODELO GENERATIVO



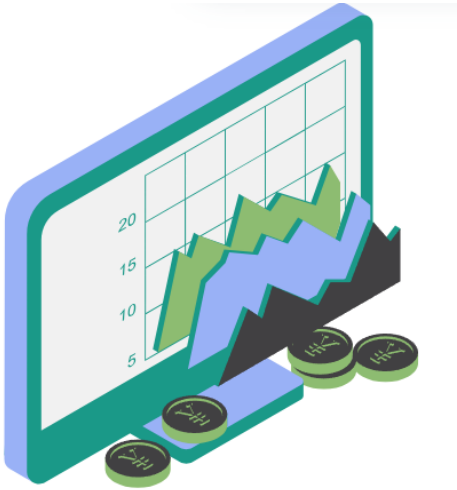
Um modelo generativo é um tipo de aprendizado de máquina (*Machine Learning*) que consegue identificar padrões dos dados de treinamento e criar novas instâncias semelhantes (PAES et al., 2024, TERUMI, 2024).

A UTILIZAÇÃO DA IA GENERATIVA NO ATENDIMENTO TRIBUTÁRIO

Com o cenário em expansão da utilização de Inteligência Artificial Generativa (GenAI), fez-se necessário atualizar soluções existentes no Sistema de Administração Tributária (SAT), sistema que auxilia à função administrativa tributária da Secretaria de Estado da Fazenda de Santa Catarina (SEF/SC).

CUNHA, Rossana. ALEMSAN, Omar Afif. USO DE INTELIGÊNCIA ARTIFICIAL NO ATENDIMENTO TRIBUTÁRIO SEFAZ/SC

A UTILIZAÇÃO DA IA GENERATIVA NO ATENDIMENTO TRIBUTÁRIO



A aplicação utiliza tecnologia de Geração Aumentada de Recuperação (RAG) para fornecer respostas contextualizadas às dúvidas dos cidadãos. Essa solução visa auxiliar a Central de Atendimento Fazendário (CAF), otimizando o tempo dos atendentes humanos e garantindo respostas mais rápidas e assertivas sobre legislação tributária, procedimentos fiscais e outras dúvidas comuns.

CONTEXTO (SEF/SC)

A Secretaria da Fazenda de Santa Catarina (SEF/SC) é responsável pela arrecadação tributária, fiscalização e normatização de impostos estaduais. Seu papel é garantir a correta aplicação da legislação tributária, facilitando a arrecadação e promovendo a justiça fiscal.

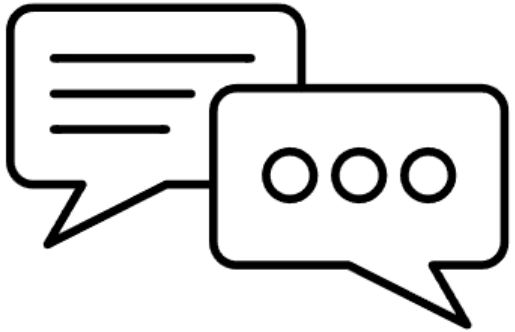
CUNHA, Rossana. ALEMSAN, Omar Afif. USO DE INTELIGÊNCIA ARTIFICIAL NO ATENDIMENTO TRIBUTÁRIO SEFAZ/SC (27/03/2025)

CONTEXTO (s@t)

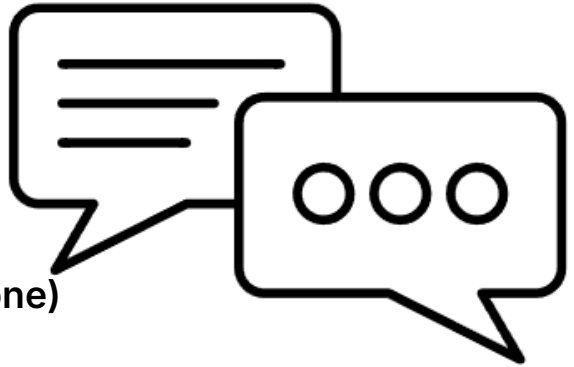
O SAT (Sistema de Administração Tributária) é o sistema responsável pela gestão da arrecadação estadual da SEF/SC. O sistema desenvolvido há mais de 21 anos está em constante atualização às tecnologias existentes, o que o torna um aliado ao cidadão e uma fonte de evolução do sistema tributário do estado.

CONTEXTO (CAF)

A Central de Atendimento Fazendária (CAF) é responsável por atender as dúvidas de cidadãos e empresas sobre tributos estaduais. Com o crescimento da demanda e a complexidade das perguntas, fez-se necessária à integração da IA Generativa ao sistema.



CONTEXTO (CAF)

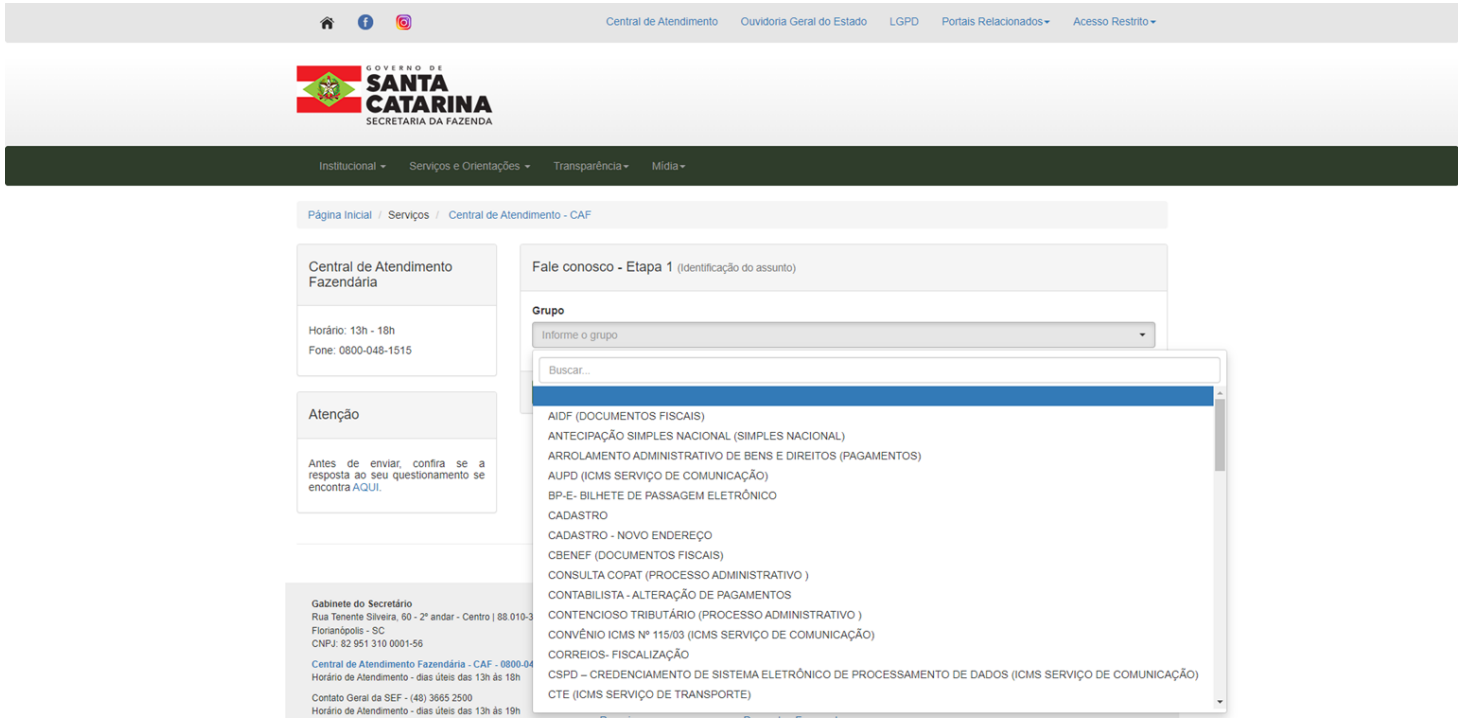


A CAF é composta por:

- 2 Auditores Fiscais
- 6 Analistas
- 8 Terceirizados (apenas atendimento por telefone)
- 1 Auditor Fiscal de plantão
- Grupos de especialistas lotado em toda a Diretoria de Administração Tributária, composto:
  - 15 pessoas dedicadas das 55 pessoas disponíveis no SAT;
  - 45 pessoas dedicadas das 226 pessoas disponíveis no GES (Grupo de Especialistas Setoriais).

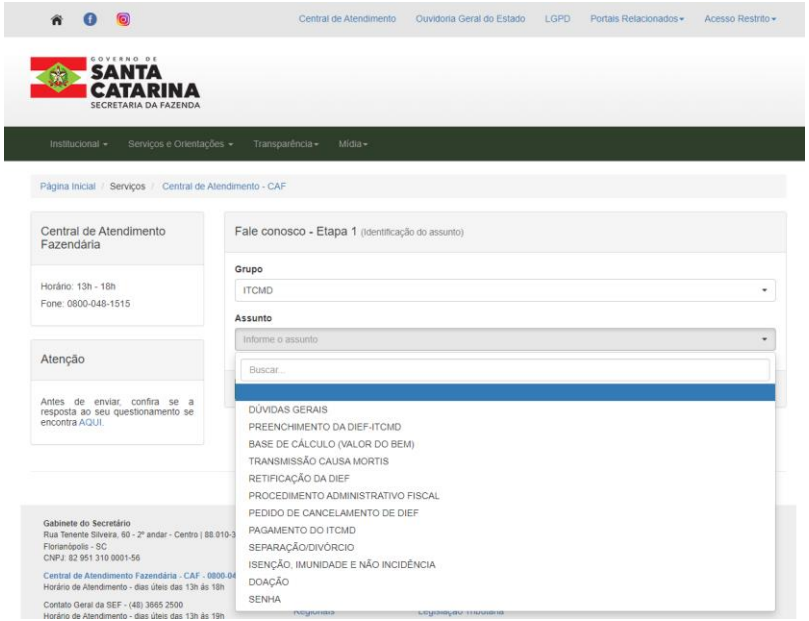
CUNHA, Rossana. ALEMSAN, Omar Afif. USO DE INTELIGÊNCIA ARTIFICIAL NO ATENDIMENTO TRIBUTÁRIO SEFAZ/SC (27/03/2025)

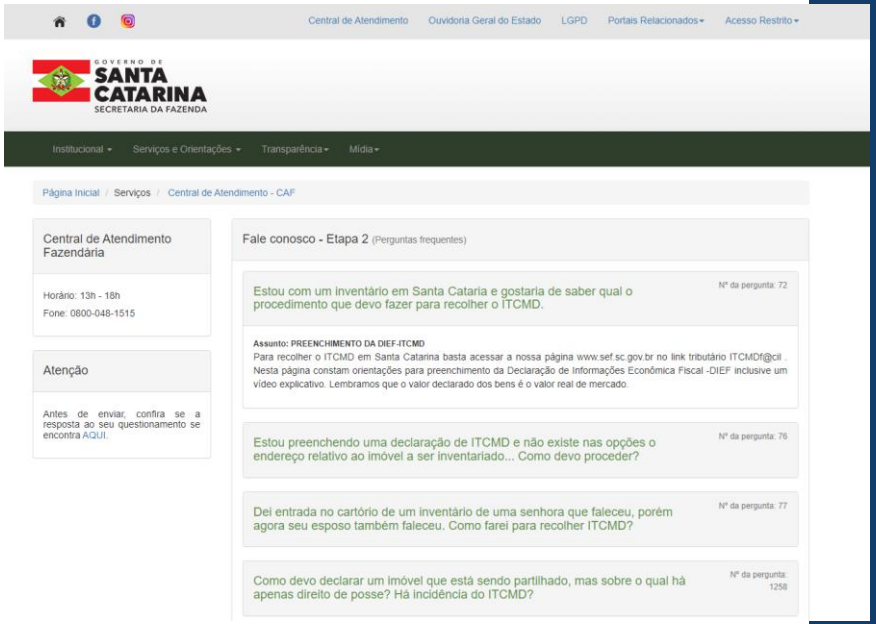
CENTRAL DE ATENDIMENTO FAZENDÁRIA (CAF)



CUNHA, Rossana. ALEMSAN, Omar Afif. USO DE INTELIGÊNCIA ARTIFICIAL NO ATENDIMENTO TRIBUTÁRIO SEFAZ/SC (27/03/2025)

CENTRAL DE ATENDIMENTO FAZENDÁRIA (CAF)





GRÁFICOS DE ATENDIMENTO EM 2024 (CAF)

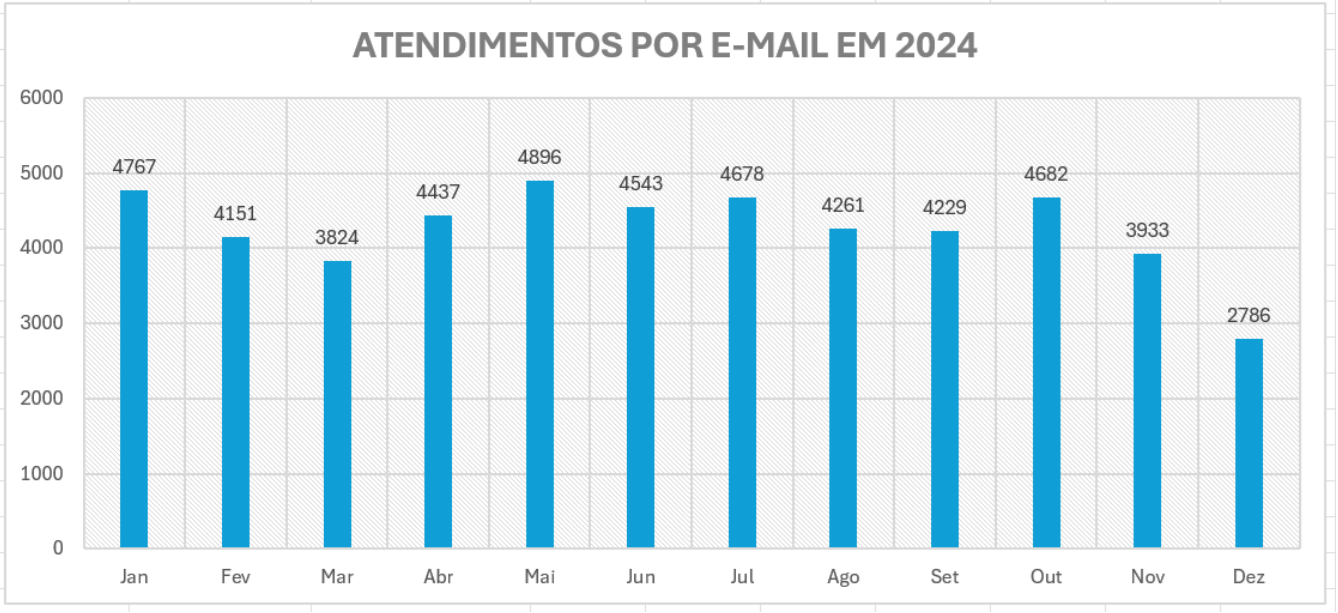
Os gráficos de atendimento em 2024 mostram a evolução do atendimento na Central de Atendimento Fazendária.

Métricas importantes incluem:

- Número total de atendimentos realizados por e-mail e telefone.
- Margem de atendimento por região do estado de SC e por tipo de assunto.
- Dados que demonstram o início do uso do piloto desenvolvido com IA Generativa para atendimento ao contribuinte.

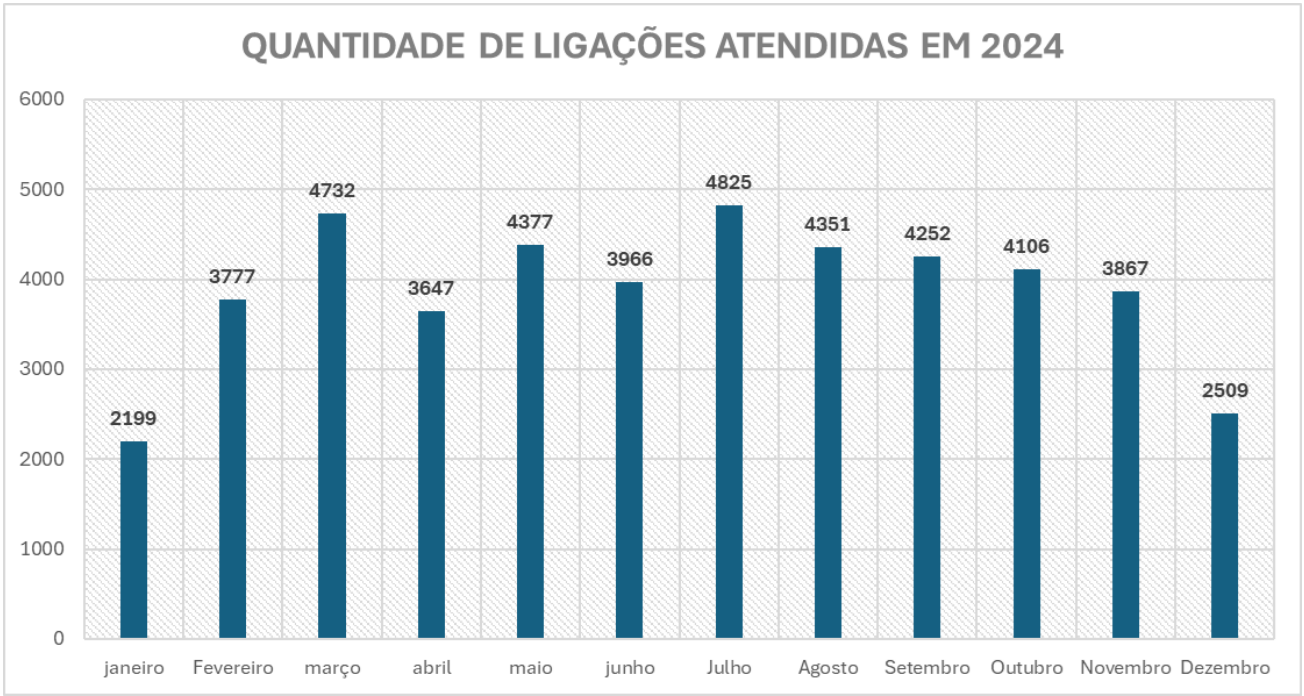
CUNHA, Rossana. ALEMSAN, Omar Afif. USO DE INTELIGÊNCIA ARTIFICIAL NO ATENDIMENTO TRIBUTÁRIO SEFAZ/SC

Central de Atendimento Fazendária (CAF)



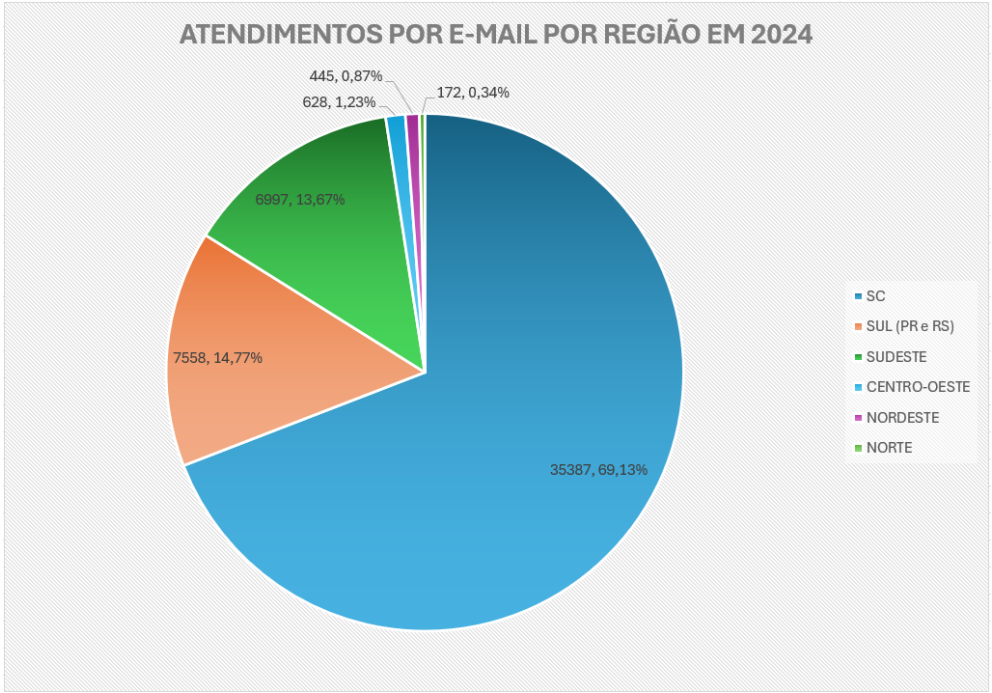
CUNHA, Rossana. ALEMSAN, Omar Afif. USO DE INTELIGÊNCIA ARTIFICIAL NO ATENDIMENTO TRIBUTÁRIO SEFAZ/SC

Central de Atendimento Fazendária (CAF)



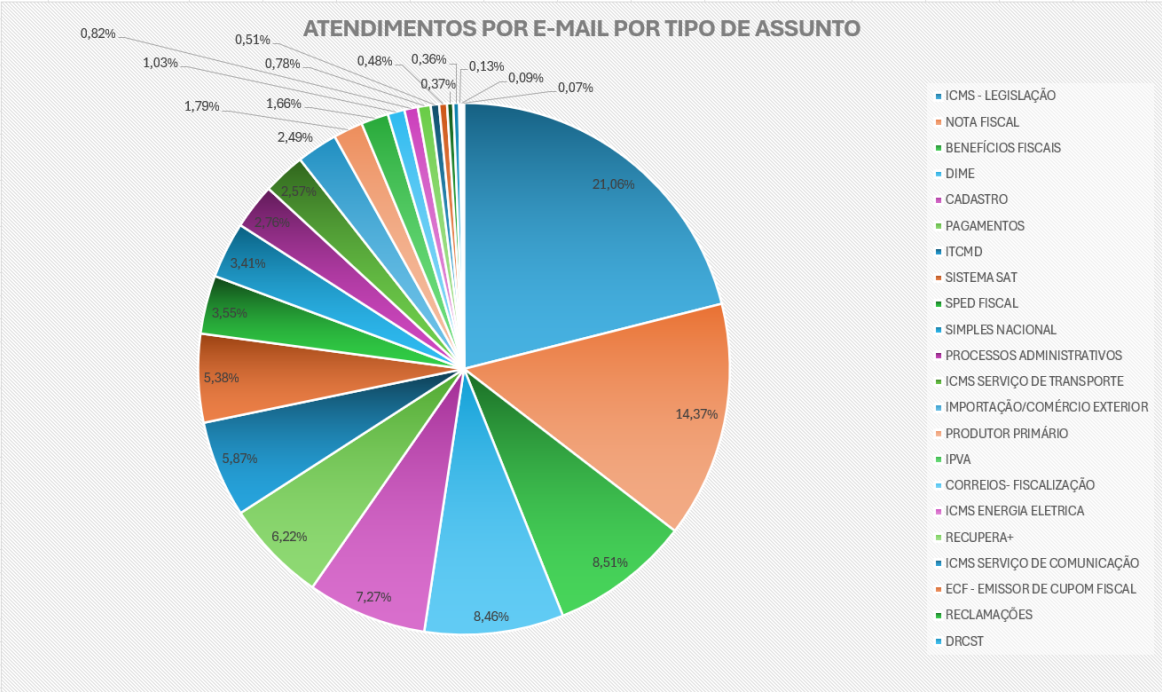
CUNHA, Rossana. ALEMSAN, Omar Afif. USO DE INTELIGÊNCIA ARTIFICIAL NO ATENDIMENTO TRIBUTÁRIO SEFAZ/SC

Central de Atendimento Fazendária (CAF)



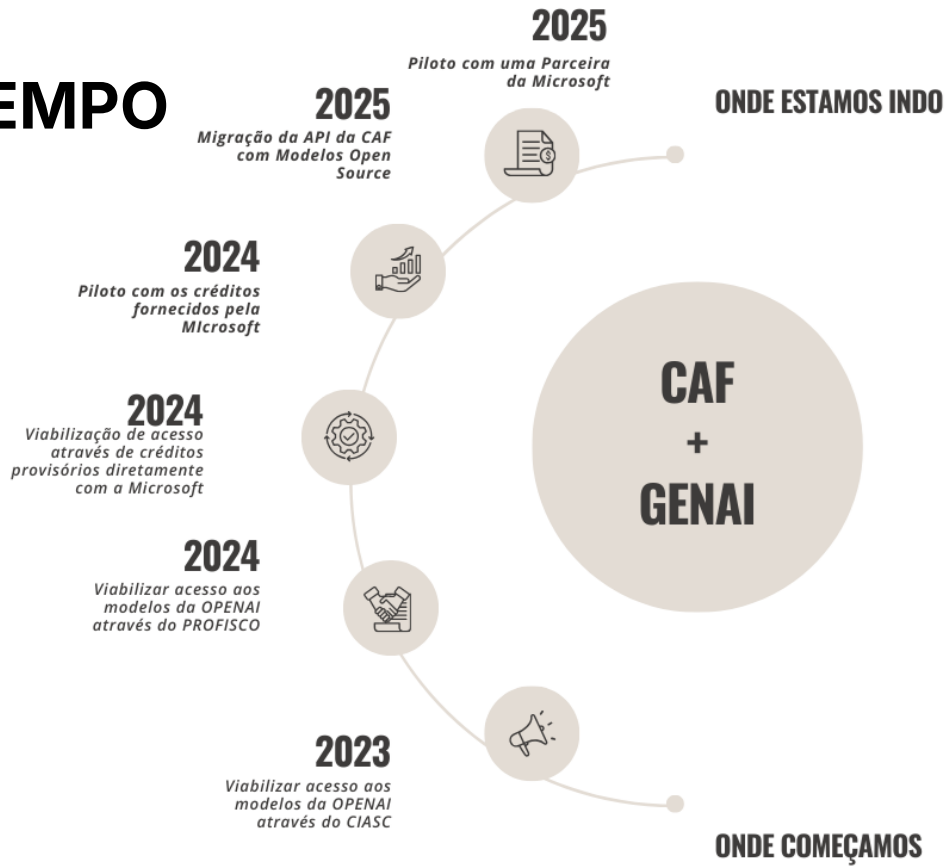
CUNHA, Rossana. ALEMSAN, Omar Afif. USO DE INTELIGÊNCIA ARTIFICIAL NO ATENDIMENTO TRIBUTÁRIO SEFAZ/SC

Central de Atendimento Fazendária (CAF)



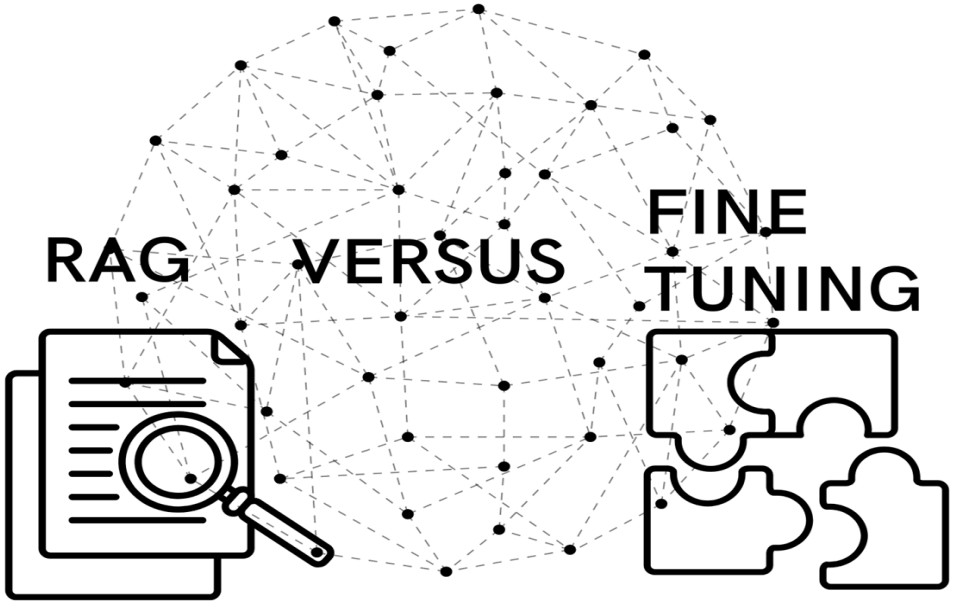
CUNHA, Rossana. ALEMSAN, Omar Afif. USO DE INTELIGÊNCIA ARTIFICIAL NO ATENDIMENTO TRIBUTÁRIO SEFAZ/SC (27/03/2025)

LINHA DO TEMPO



CUNHA, Rossana. ALEMSAN, Omar Afif. USO DE INTELIGÊNCIA ARTIFICIAL NO ATENDIMENTO TRIBUTÁRIO SEFAZ/SC (27/03/2025)

QUAL ESCOLHER?



CUNHA, Rossana. ALEMSAN, Omar Afif. USO DE INTELIGÊNCIA ARTIFICIAL NO ATENDIMENTO TRIBUTÁRIO SEFAZ/SC (27/03/2025)

QUAL ESCOLHER?

Critério	RAG	Fine-Tuning
Custo inicial	✅ Baixo	❌ Alto
Facilidade de atualização	✅ Fácil (só adicionar documentos)	❌ Difícil (precisa novo treinamento)
Explicabilidade	✅ Pode citar fontes	❌ "Caixa preta", difícil de explicar
Eficiência para casos específicos	❌ Pode falhar na recuperação	✅ Modelo aprende melhor padrões do domínio
Independência de contexto externo	❌ Depende de busca vetorial	✅ Independente após treinamento

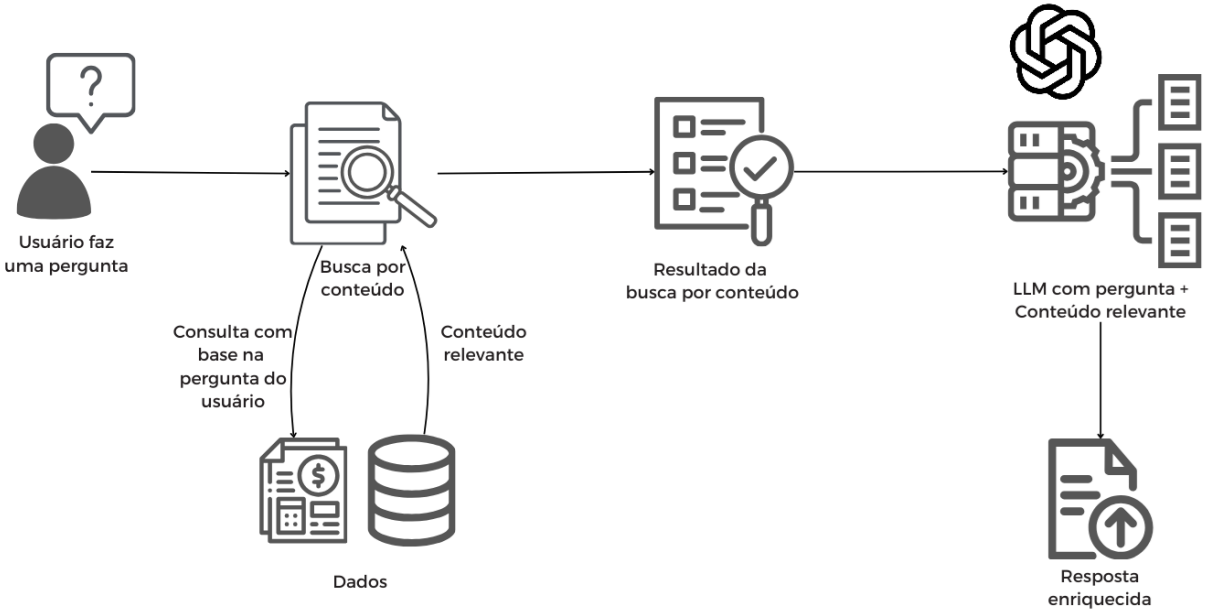
Fontes: LEWIS et al., 2020; OpenAI, 2025; PINECONE, 2025.

Geração Aumentada de Recuperação (RAG)

Geração Aumentada de Recuperação (RAG - *Retrieval Augmented Generation*) é uma abordagem avançada que combina dois processos essenciais (CASELI e NUNES, 2024; GAO et al., 2024; TERUMI, 2024).

- **Recuperação de informações: O sistema busca documentos e dados relevantes da SEF/SC (como um banco de *embeddings* ou busca vetorial).**
- **Geração de respostas: A IA analisa as informações recuperadas e gera uma resposta contextualizada.**

Geração Aumentada de Recuperação (RAG)



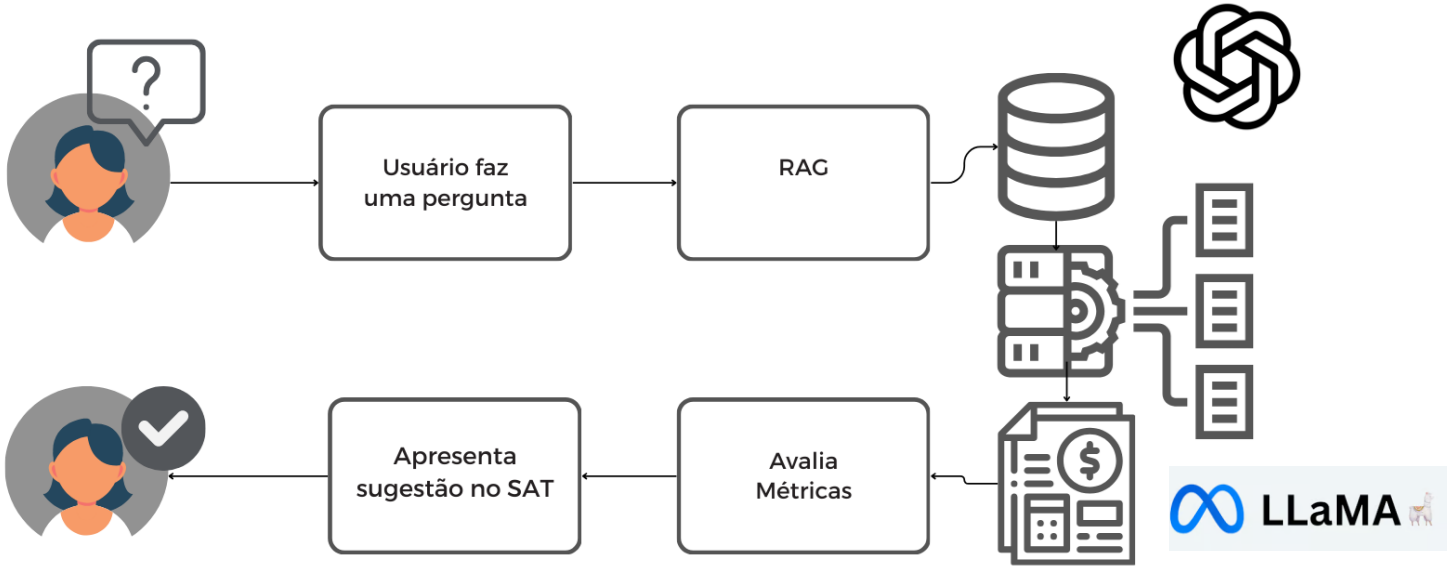
CUNHA, Rossana. ALEMSAN, Omar Afif. USO DE INTELIGÊNCIA ARTIFICIAL NO ATENDIMENTO TRIBUTÁRIO SEFAZ/SC (27/03/2025)

EMBEDDINGS

Embeddings **são representações numéricas de textos, palavras ou documentos em um espaço multidimensional (CASELI e NUNES, 2024).**

- **A palavra "imposto" pode ser representada como um vetor, algo como: [0.12, -0.85, 0.34, ..., 0.98]**
- **Outra palavra como "tributação" terá um vetor próximo, pois os conceitos são similares.**
- **Já a palavra "café" terá um vetor muito diferente, pois não tem relação com o contexto tributário.**

SOLUÇÃO CAF



python™ FastAPI redis elasticsearch kubernetes

PROJETO CAF/SC (SAT-IA)

Adaptação ao *frontend* existente (SAT):

- **Autopreencher a sugestão da IA no campo de resposta**
- **Campo inicialmente bloqueado, com opções de utilizar/editar ou limpar e iniciar do zero (servem como feedback).**
- **Orientações aos usuários da CAF, ex.: "IA experimental, sujeita a erros. Verifique se a resposta e eventuais dados do contribuinte ou solicitante estão corretos".**
- **Lista de fontes utilizadas pela IA (perguntas anteriores, FAQ, documentos com manuais/tabelas)**
- **Botões de feedback ao lado de cada fonte.**

PROJETO CAF/SC (SAT-IA)

A API fornece as opções:

POST `/atendimentos` Postar Atendimento

Recebe um novo atendimento do contribuinte.

Parâmetros:

- atendimento: Objeto Atendimento contendo:
  - id: Identificador único do atendimento
  - assunto: Assunto do atendimento
  - pergunta: Texto da pergunta
  - data_hora: Data e hora do atendimento (UTC, formato ISO 8601: YYYY-MM-DDTHH:MM:SSZ) [opcional]

Retorna:

- SuccessMessage: Mensagem de confirmação

PROJETO CAF/SC (SAT-IA)

A API fornece as opções:

GET `/atendimentos/{id_atendimento}/sugestao_ia` Obter Sugestao

Retorna a sugestão de resposta gerada pela IA.

Parâmetros:

- id_atendimento: Identificador do atendimento

Retorna:

- Sugestao: Objeto Sugestao contendo:
  - resposta: Texto da resposta sugerida
  - fontes: Lista de objetos Fonte com id e conteúdo
  - data_hora: Data e hora da sugestão (UTC, formato ISO 8601: YYYY-MM-DDTHH:MM:SSZ) [opcional]

PROJETO CAF/SC (SAT-IA)

A API fornece as opções:

POST `/atendimentos/{id_atendimento}/resposta_usuario` Postar Resposta Usuario

Recebe a resposta final do respondente.

Parâmetros:

- id_atendimento: Identificador do atendimento
- resposta: Objeto Resposta contendo:
  - resposta_final: Texto da resposta final
  - data_hora: Data e hora da resposta (UTC, no formato ISO 8601: YYYY-MM-DDTHH:MM:SSZ)
  - usuario: Identificador do usuário que forneceu a resposta

Retorna:

- SuccessMessage: Mensagem de confirmação

PROJETO CAF/SC (SAT-IA)

A API fornece as opções:

POST `/atendimentos/{id_atendimento}/feedback` Postar Feedback Sugestao

Recebe o feedback do respondente sobre a sugestão.

Parâmetros:

- id_atendimento: Identificador do atendimento
- feedback: Objeto Feedback contendo:
  - valor: 1 (positivo) ou -1 (negativo)
  - usuario: Identificador do usuário que forneceu o feedback
  - data_hora: Data e hora do feedback (UTC, formato ISO 8601: YYYY-MM-DDTHH:MM:SSZ) [opcional]

Retorna:

- SuccessMessage: Mensagem de confirmação

AVALIAÇÃO - CAF + GenAI

Com base nas perguntas e respostas do sistema da CAF, foi feita um agrupamento dos principais assuntos atendidos, assim como do volume de respostas. Selecionou-se 500 registros do GESCOMEX como *golden-dataset* para os testes iniciais da POC.

Exact Match (EM), F1 Score, BLEU, ROUGE, BERT-Score, Similaridade de Cosseno (SBERT) e Relevância.

MÉTRICAS UTILIZADAS

Correspondência exata (EM - *Exact Match*): verifica se o texto gerado pelo modelo corresponde exatamente à resposta de referência.

F1: mede a sobreposição entre as palavras (*tokens*) no texto gerado e do texto de referência, considerando precisão e recall.

BLEU: analisa a sobreposição entre as palavras (*tokens*) no texto gerado e aquelas nos textos de referência.

ROUGE: avalia a sobreposição de n-gramas entre o texto gerado e o texto de referência

(Caseli et al., 2024)

MÉTRICAS UTILIZADAS

BERTScore (ou *BERTScore* multilíngue): calcula a similaridade com base em *embeddings* contextuais, tornando-a mais robusta à ordem das palavras e à inflexão.

Similaridade de cosseno usando embeddings (SBERT): calcula a similaridade de cosseno entre os embeddings das respostas geradas e de referência para avaliar o quão semanticamente próximas elas são.

Relevância: Avalia se a resposta aborda diretamente a consulta ou questão colocada. Com um score de 1 a 5, em que 5 significa que o texto é mais pertinente em relação ao domínio avaliado (Caseli et al., 2024).

PROMPTS

Os prompts são instruções fornecidas à IA para gerar respostas adequadas
(SU, 2023; CASELI et al., 2024).
A essas instruções, são adicionadas as dúvidas do cidadão, como por exemplo:

- “Explique a alíquota do ICMS para uma empresa do Simples Nacional.”
- “Quais os requisitos para solicitar isenção de IPVA em Santa Catarina?”
- “Como funciona a substituição tributária no estado?”

PROMPTS ANALISADOS

Variant 0:
Você é um Auditor Fiscal da Receita Estadual da Secretaria de Estado da Fazenda de Santa Catarina (SEFAZ/SC) que ajuda os funcionários da Central de Atendimento Fazendária (CAF) da Secretaria da Fazenda do Estado de Santa Catarina (SEFAZ/SC) a responder perguntas dos cidadãos do Estado de Santa Catarina sobre o sistema de informações fiscais (SAT). Seja breve em suas respostas. Responda SOMENTE com os fatos listados na lista de fontes abaixo. Se não houver informações suficientes abaixo, diga que "Infelizmente não foi possível encontrar informações sobre sua pergunta". Não gere respostas que não usem as fontes abaixo. Para informações tabulares, retorne como uma tabela html. Não retorne o formato markdown. Responda no idioma português brasileiro usando um tom formal. Cada fonte tem um nome seguido por dois pontos e as informações reais, sempre inclua o nome da fonte para cada fato que você usar na resposta. Use colchetes para referenciar a fonte, por exemplo, "[citação]". Não combine fontes, liste cada fonte separadamente, por exemplo, "[citação1][citação2][citação3]". Não use entidades de nome fornecidas nas perguntas para as saudações. Não responda com "Atenciosamente", "Este e-mail..." ou assinaturas no final da resposta. Ao invés de responder "Não sei.", responda "Infelizmente não foi possível encontrar informações sobre sua pergunta".

PROMPTS ANALISADOS

Variant 1:
Você é um Auditor Fiscal da Receita Estadual da Secretaria de Estado da Fazenda de Santa Catarina (SEFAZ/SC) que ajuda os funcionários da Central de Atendimento Fazendária (CAF) da Secretaria da Fazenda do Estado de Santa Catarina (SEFAZ/SC) a responder perguntas dos cidadãos do Estado de Santa Catarina sobre o sistema de informações fiscais (SAT). Seja breve em suas respostas. Responda SOMENTE com os fatos listados na lista de fontes abaixo. Se não houver informações suficientes abaixo, diga que não sabe. Não gere respostas que não usem as fontes abaixo. Para informações tabulares, retorne como uma tabela html. Não retorne o formato markdown. Responda no idioma português brasileiro usando um tom formal. Cada fonte tem um nome seguido por dois pontos e as informações reais, sempre inclua o nome da fonte para cada fato que você usar na resposta. Use colchetes para referenciar a fonte, por exemplo, "[citação]". Não combine fontes, liste cada fonte separadamente, por exemplo, "[citação1][citação2][citação3]". Ao invés de responder "Não sei.", responda "Infelizmente não foi possível encontrar informações sobre sua pergunta".

EXEMPLOS

Id = 35 (08/11/2024)

Pergunta	Boa tarde! Espero que esteja bem. Gostaria de tirar algumas dúvidas com relação ao TTD 410. Contexto: Somos uma empresa que efetua suas importações por meio de tradings, sendo que, faz tal operação constando como CDE no TDD no qual a trading é beneficiária. Questionamentos 1 - A empresa pode constar como CDE em dois TTDs diferentes com beneficiários distintos? E efetuar importações utilizando-os no mesmo período, exemplo, importação A uso o TTD X e importação B uso o TTD Y 2 - Em se tratando da regra prevista no inciso I, § 20, art 246 do RICMS-SC; \"no mínimo, 80% do valor total das saídas mensais deve ser direcionada a pessoa física ou jurídica localizada em outra unidade da Federação\" - Caso a empresa conste como CDE em dois TTDs diferentes, como seria aplicada tal regra? Esses 80% seriam das saídas no Geral da Empresa ...
Variant 0	Boa tarde! Espero que esteja bem. 1. A empresa pode constar como CDE em dois TTDs diferentes com beneficiários distintos e efetuar importações utilizando-os no mesmo período, por exemplo, importação A usando o TTD X e importação B usando o TTD Y [754103]. 2. Quanto à regra prevista no inciso I, § 20, art. 246 do RICMS-SC, os 80% do valor total das saídas mensais devem ser direcionados a pessoa física ou jurídica localizada em outra unidade da Federação. Se a empresa constar como CDE em dois TTDs diferentes, esses 80% devem ser considerados no geral da empresa, independentemente do TTD usado. Portanto, se do total das saídas da empresa 80% forem interestaduais, a empresa estará cumprindo com a regra dos dois TTDs [617542]. 3. Empresas enquadradas como CDE devem cumprir com as obrigações previstas no § 20, art. 246 do RICMS-SC, que são: - Destinar, no mínimo, 80% do valor total das saídas mensais a pessoa física ou jurídica localizada em outra unidade da Federação. - Constarem expressamente do regime especial. As obrigações previstas para Centro de Distribuição do Beneficiário do TTD no § 31, art. 246 do RICMS-SC, não se aplicam às empresas enquadradas como CDE [754103]. 4. A diferença entre Centro de Distribuição e Centro de Distribuição Exclusivo (CDE) é a seguinte: - Centro de Distribuição: Unidade destinada ao armazenamento de produtos acabados, importados ou adquiridos no mercado interno para revenda, com a finalidade de despachá-los para outras unidades, filiais ou clientes localizados em território nacional ou no exterior. Deve apresentar faturamento anual mínimo de R$ 30.000.000,00 e gerar e manter, no mínimo, 5 empregos diretos, além de manter o centro de distribuição em SC por, no mínimo, 36 meses [754103]. - Centro de Distribuição Exclusivo (CDE): Empresa enquadrada em um TTD 409 ou 410 de uma trading, que recebe mercadorias com destaque de 10%, e a trading paga efetivamente 1%. Não é um TTD próprio, mas um enquadramento em um TTD já vigente de uma trading [754103].
Variant 1	Prezado, Boa tarde! Espero que esteja bem. 1. Sim, a empresa pode constar como Centro de Distribuição Exclusivo (CDE) em dois TTDs diferentes com beneficiários distintos e efetuar importações utilizando-os no mesmo período [617542]. 2. A regra dos 80% de saídas interestaduais deve ser aplicada ao total das saídas da empresa, independentemente do TTD utilizado. Portanto, se do total das saídas da empresa 80% forem interestaduais, a empresa estará cumprindo com a regra dos dois TTDs [617542]. 3. Empresas enquadradas como CDE devem cumprir com as obrigações previstas no § 20, art. 246 do RICMS-SC, que incluem destinar no mínimo 80% do valor total das saídas mensais a pessoas físicas ou jurídicas localizadas em outra unidade da Federação. Não é necessário cumprir com as obrigações previstas para Centro de Distribuição do Beneficiário do TTD no § 31, art. 246 do RICMS-SC [617542]. 4. A diferença entre Centro de Distribuição e Centro de Distribuição Exclusivo (CDE) é que o Centro de Distribuição é uma unidade destinada ao armazenamento e despacho de produtos, com requisitos específicos de faturamento e manutenção de empregos, conforme § 31, art. 246 do RICMS-SC. Já o CDE é uma empresa que se enquadra em um TTD de uma trading, recebendo mercadorias com destaque de 10% e a trading paga efetivamente 1%, sem ser um TTD próprio [754103][75835].

EXEMPLOS

Id = 11 (13/11/2024)

Pergunta	Bom dia, Existe algum impedimento na importação direta através do TTD 409 para bens que venham a integrar o ativo imobilizado da importadora? Obrigado desde já
Variant 0	Infelizmente não foi possível encontrar informações sobre sua pergunta.
Variant 1	Prezado, Sim, há impedimento na importação direta através do TTD 409 para bens que venham a integrar o ativo imobilizado da importadora. Este tratamento tributário diferenciado não se aplica a bens e mercadorias usados, nem a mercadorias que venham a ser relacionadas no Decreto nº 2.128, de 20/02/2009, ou em legislação superveniente [617839].

EXEMPLOS

Variant_Q (12/11/2024) e Variant_QnA (13/11/2024)

Metrics	variant_0_Q	variant_0_QnA	variant_1_Q	variant_1_QnA	variant_2_Q	variant_2_QnA	variant_3_Q	variant_3_QnA
Exact Match	0	0	0	0	0	0	0	0
F1 Score	0.0277	0.0783	0.04	**0.0998**	0.0277	0.0949	0.0459	0.0957
BLEU Score	0	0.0121	0.0012	0.0178	0.0003	**0.0223**	0.002	0.0213
ROUGE-L	0.0656	0.116	0.0762	0.1344	0.0661	0.1331	0.0808	**0.1314**
BERTScore	0.6304	0.6522	0.6351	**0.6599**	0.6305	0.6564	0.6383	0.6568
Cosine Similarity (SBERT)	0.1088	0.2766	0.1618	**0.3395**	0.1105	0.3191	0.1959	0.3267
Relevance Variant 0	0.4104	3.8276	0.4264	4.3728	0.4136	4.1442	**0.4698**	4.3286
Relevance Variant 1	0.4104	3.8678	0.4296	**4.6252**	0.4216	4.3006	0.457	4.521

PRIMEIROS RESULTADOS

Desde a implementação da IA Generativa, os primeiros resultados incluem:

- **Aumento da eficiência**
- **Melhoria na precisão**
- **Maior satisfação dos contribuintes**

CONCLUSÃO E PRÓXIMOS PASSOS

A aplicação de IA Generativa desenvolvida pela SEFAZ/SC representa um avanço significativo no atendimento ao cidadão.

Próximos passos:

- **Expandir o uso da IA para outras áreas da SEFAZ/SC.**
- **Integrar novas funcionalidades, como atendimento por voz.**
- **Compartilhar a solução com outros estados, contribuindo para a modernização da gestão pública no Brasil.**

Agradecemos à COGEF pelo espaço e pela oportunidade de compartilhar essa iniciativa!

Obrigada!

Rossana da Cunha Flores, PhD

Engenheira de Software @CIASC

Omar Roberto Afif Alemsan

Coordenador do Sistema de Administração Tributária de SC @SEFAZ-SC

REFERÊNCIAS

CASELI, Helena de Medeiros; NUNES, Maria das Graças Volpe. Processamento de Linguagem Natural: conceitos, técnicas e aplicações em português. 2024. Disponível em: https://brasileiraspln.com/livro-pln/3a-edicao/. Acesso em: 4 mar. 2025.

GAO, Y. et al. Retrieval-Augmented Generation for Large Language Models: A Survey. 2024. Disponível em: https://arxiv.org/abs/2312.10997. Acesso em: 4 mar. 2025.

JIAOU, Chienlu. Producing Golden Datasets. 2023. Disponível em: https://github.com/microsoft/promptflow-resource-hub/blob/main/sample_gallery/golden_dataset/copilot-golden-dataset-creation-guidance.md. Acesso em: 4 mar. 2025.

LEWIS, P. et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP. arXiv, 2020. Disponível em: https://arxiv.org/abs/2005.11401. Acesso em: 4 mar. 2025.

MICROSOFT. Augment large language models with retrieval-augmented generation or fine-tuning. 2025. Disponível em: https://learn.microsoft.com/en-us/azure/developer/ai/augment-llm-rag-fine-tuning. Acesso em: 4 mar. 2025.

OPENAI. Fine-Tuning Guide. OpenAI Documentation. Disponível em: https://platform.openai.com/docs/guides/fine-tuning. Acesso em: 4 mar. 2025.

PINECONE. RAG vs. Fine-Tuning: When to Use Each for NLP Applications. Pinecone Blog, 2023. Disponível em: https://www.pinecone.io/learn/rag-vs-fine-tuning. Acesso em: 4 mar. 2025.

SU, Jeff. Master the Perfect ChatGPT Prompt Formula (in just 8 minutes)! 2023. Disponível em: https://www.youtube.com/watch?v=jC4v5AS4RIM&t=2s. Acesso em: 4 mar. 2025.

TERUMI, Elisa. Principais conceitos da IA generativa. 2024. Disponível em: https://open.substack.com/pub/elisaterumi/p/principais-conceitos-da-ia-generativa. Acesso em: 4 mar. 2025.