

# Informações do Aluno

**Nome:** Ramos de Souza Janones

**E-mail:** [ramosinfo@gmail.com](mailto:ramosinfo@gmail.com)

**Data de Entrega:** 20/05/2025

## ÍNDICE ANALITICO

### [Informações do Aluno](#)

### [Exercício: Estratégias de Raciocínio em LLMs](#)

#### [Prompting e In-context Learning](#)

[Zero-shot:](#)

[One-shot:](#)

[Few-shot:](#)

[1.2 Resultado Obtido](#)

[2. Análise Crítica](#)

[2.1 Pontos Fortes](#)

[2.2 Limitações Encontradas](#)

[2.3 Possíveis Melhorias](#)

#### [Rationale Engineering](#)

##### [1. Implementação](#)

[1.1 Prompt Utilizado](#)

[1.2 Resultado Obtido](#)

[2. Análise Crítica](#)

[2.1 Pontos Fortes](#)

[2.2 Limitações Encontradas](#)

[2.3 Possíveis Melhorias](#)

[3 Conclusão](#)

#### [Prompting em Cadeia de Pensamento \(Chain-of-Thought\)](#)

##### [1. Implementação](#)

[1.1 Prompt Utilizado](#)

[1.2 Resultado Obtido](#)

[2. Análise Crítica](#)

[2.1 Pontos Fortes](#)

[2.2 Limitações Encontradas](#)

[2.3 Possíveis Melhorias](#)

[3 Conclusão](#)

#### [Amostragem de Autoconsistência \(Self-Consistency Sampling\)](#)

##### [1. Implementação](#)

[1.1 Prompt Utilizado](#)

[1.2 Resultado Obtido](#)

[2. Análise Crítica](#)

[2.1 Pontos Fortes](#)

[2.2 Limitações Encontradas](#)

2.3 Possíveis Melhorias	
3 Conclusão	
Decomposição de Problemas	
1. Implementação	
1.1 Prompt Utilizado	
1.2 Resultado Obtido	
2. Análise Crítica	
2.1 Pontos Fortes	
2.2 Limitações Encontradas	
2.3 Possíveis Melhorias	
3 Conclusão	
Raciocínio com Ferramentas Auxiliares (Tool-Augmented Reasoning)	
1. Implementação	
1.1 Prompt Utilizado	
1.2 Resultado Obtido	
2. Análise Crítica	
2.1 Pontos Fortes	
2.2 Limitações Encontradas	
2.3 Possíveis Melhorias	
3 Conclusão	
Raciocínio com Memória e Contexto (Memory and Contextual Reasoning)	
1. Implementação	
1.1 Prompt Utilizado	
1.2 Resultado Obtido	
2. Análise Crítica	
2.1 Pontos Fortes	
2.2 Limitações Encontradas	
2.3 Possíveis Melhorias	
3. Conclusão	
MCP (Model Context Protocol)	
1. Implementação	
1.1 Prompt Utilizado	
1.2 Resultado Obtido	
2. Análise Crítica	
2.1 Pontos Fortes	
2.2 Limitações Encontradas	
2.3 Possíveis Melhorias	
3. Conclusão	
Comparativo dos Métodos	
Discussão sobre a Escolha do Método	
Análise de Erros Comuns	
Considerações Éticas	
Conclusão	
Observações para Avaliação	

## Referências

## Exercício: Estratégias de Raciocínio em LLMs

Todos os prompts exemplos utilizados já são um ponto exploratório para o projeto de conclusão do curso. Para possíveis produções, no futuro, de melhores prompts para o trabalho de conclusão.

### Prompting e In-context Learning

Os prompts fornecem exemplos no contexto para guiar o modelo a aprender e responder adequadamente sem necessidade de treinamento adicional.

#### Zero-shot:

##### **[INPUT]**

Você é um assistente especializado em contabilidade fiscal. Analise a seguinte descrição de uma transação e determine o CFOP (Código Fiscal de Operações e Prestações) mais adequado, justificando sua escolha:

"Empresa varejista de São Paulo vende mercadorias para consumidor final no Rio de Janeiro, com entrega via transportadora."

#### One-shot:

##### **[INPUT]**

Exemplo:

Transação: "Indústria de São Paulo envia matéria-prima para industrialização em Minas Gerais, com retorno previsto em 60 dias."

Classificação: CFOP 6901 - Remessa para industrialização por encomenda. Este código é utilizado quando uma empresa envia materiais para serem processados por outra empresa em outro estado, com previsão de retorno dos produtos industrializados.

Agora, classifique a seguinte transação:

"Distribuidora do Paraná vende mercadorias para revenda a uma loja de varejo em Santa Catarina, com substituição tributária."

#### Few-shot:

##### **[INPUT]**

Aqui estão alguns exemplos de classificações de transações fiscais:

Transação 1: "Indústria de SP vende produtos acabados para consumidor final no mesmo estado."

Classificação: CFOP 5101 - Venda de produção do estabelecimento dentro do estado. Este código é utilizado para vendas de produtos fabricados pelo próprio estabelecimento para destinatários dentro do mesmo estado.

Transação 2: "Comércio atacadista de MG transfere mercadorias para sua filial no RJ."

Classificação: CFOP 6152 - Transferência de mercadorias adquiridas para comercialização. Este código é utilizado quando há transferência de mercadorias entre estabelecimentos da mesma empresa em estados diferentes.

Transação 3: "Empresa de SP recebe devolução de mercadoria vendida para cliente no PR."

Classificação: CFOP 2202 - Devolução de venda de mercadoria adquirida. Este código é utilizado quando o estabelecimento recebe devolução de mercadorias que foram vendidas para outro estado.

Agora, classifique a seguinte transação:

"Empresa de serviços de TI do RS emite nota fiscal de serviço de manutenção de software para cliente no RS, sendo que o software foi desenvolvido pela própria empresa."

## 1.2 Resultado Obtido

As respostas apresentaram boa aderência ao contexto fornecido, com exemplos claros e aplicados diretamente às perguntas.

## 2. Análise Crítica

### 2.1 Pontos Fortes

Permite ao modelo generalizar a partir de poucos exemplos, facilitando respostas contextualizadas e precisas.

### 2.2 Limitações Encontradas

Dependência da qualidade e relevância dos exemplos no contexto; se mal escolhidos, podem levar a respostas imprecisas.

### 2.3 Possíveis Melhorias

Incluir exemplos mais variados e representativos para cobrir diferentes nuances do problema.

---

## Rationale Engineering

Os prompts incentivam o modelo a explicar seu raciocínio passo a passo antes de chegar à resposta final.

# 1. Implementação

## 1.1 Prompt Utilizado

### **[INPUT]**

Você é um auditor fiscal experiente. Analise a seguinte situação tributária e explique detalhadamente seu raciocínio:

"Uma empresa optante pelo Simples Nacional com faturamento anual de R\$ 1.800.000 realizou importação de mercadorias no valor de R\$ 250.000 e está em dúvida sobre o cálculo correto do ICMS devido nesta operação."

Para sua análise, siga estas etapas:

1. Identifique todos os impostos e contribuições envolvidos nesta operação
2. Explique como o regime tributário da empresa afeta o cálculo
3. Detalhe a base de cálculo do ICMS na importação
4. Explique se há diferença no tratamento por ser empresa do Simples
5. Calcule o valor aproximado devido, mostrando cada passo do cálculo
6. Indique quais documentos a empresa deve manter para comprovar a regularidade fiscal

Ao final, resuma sua conclusão em um parágrafo objetivo.

## 1.2 Resultado Obtido

Respostas mais transparentes e justificadas, facilitando a compreensão do processo de pensamento do modelo.

## 2. Análise Crítica

### 2.1 Pontos Fortes

Melhora a interpretabilidade e confiança nas respostas geradas.

### 2.2 Limitações Encontradas

Pode aumentar o tempo de resposta e gerar explicações redundantes ou superficiais.

### 2.3 Possíveis Melhorias

Refinar os prompts para focar em raciocínios mais relevantes e concisos.

## 3 Conclusão

A engenharia de raciocínio aumenta a transparência do modelo, sendo útil para aplicações que exigem justificativas claras, embora possa impactar a eficiência.

---

## Prompting em Cadeia de Pensamento (Chain-of-Thought)

O modelo é guiado a decompor problemas complexos em etapas sequenciais para facilitar o raciocínio.

### 1. Implementação

#### 1.1 Prompt Utilizado

##### **[INPUT]**

Uma empresa comercial precisa analisar a viabilidade fiscal de três cenários diferentes para sua expansão. Analise cada cenário passo a passo e determine qual é mais vantajoso do ponto de vista tributário.

Cenário 1: Abrir uma filial em outro estado, mantendo toda a estrutura atual.

Cenário 2: Criar uma nova empresa no mesmo estado, com regime tributário diferente.

Cenário 3: Fazer parceria com distribuidor local em outro estado.

Para cada cenário, analise:

1. Quais impostos seriam aplicáveis (ICMS, IPI, PIS/COFINS, ISS)
2. Como seria a tributação nas operações interestaduais
3. Impactos na escrituração fiscal e contábil
4. Riscos fiscais envolvidos
5. Estimativa de economia ou aumento da carga tributária

Mostre seu raciocínio completo para cada etapa antes de chegar à conclusão final.

#### 1.2 Resultado Obtido

Melhora significativa na resolução de problemas complexos, com respostas estruturadas e lógicas.

### 2. Análise Crítica

#### 2.1 Pontos Fortes

Facilita o raciocínio complexo e a solução de problemas que exigem múltiplas etapas.

#### 2.2 Limitações Encontradas

Pode ser sensível a erros em etapas iniciais que se propagam para o resultado final.

#### 2.3 Possíveis Melhorias

Implementar mecanismos de verificação em cada etapa para evitar propagação de erros.

### 3 Conclusão

A cadeia de pensamento é eficaz para problemas complexos, mas requer cuidados para garantir a precisão em cada etapa do raciocínio.

---

## Amostragem de Autoconsistência (Self-Consistency Sampling)

Geração de múltiplas respostas para a mesma pergunta, selecionando a resposta mais consistente entre elas.

### 1. Implementação

#### 1.1 Prompt Utilizado

**[INPUT]**

Responda à seguinte questão tributária complexa:

"Uma empresa industrial possui créditos acumulados de ICMS no valor de R\$ 350.000 e precisa determinar a melhor estratégia para sua utilização. Considerando a legislação tributária brasileira, quais são as possibilidades legais para aproveitamento desses créditos e qual seria a mais vantajosa?"

Apresente pelo menos três abordagens diferentes para resolver este problema, detalhando o raciocínio em cada uma. Ao final, indique qual seria a solução mais consistente e por quê.

#### 1.2 Resultado Obtido

Aumento da confiabilidade e precisão das respostas ao reduzir respostas errôneas isoladas.

### 2. Análise Crítica

#### 2.1 Pontos Fortes

Reduz a variabilidade e melhora a robustez das respostas.

#### 2.2 Limitações Encontradas

Requer maior custo computacional devido à geração múltipla de respostas.

#### 2.3 Possíveis Melhorias

Otimizar o número de amostras para balancear custo e qualidade.

### 3 Conclusão

A autoconsistência é uma técnica valiosa para aumentar a confiança nas respostas, especialmente em tarefas críticas, apesar do custo computacional.



---

## Decomposição de Problemas

Dividir problemas complexos em subproblemas menores para facilitar a resolução.

### 1. Implementação

#### 1.1 Prompt Utilizado

**[INPUT]**

Uma empresa de médio porte precisa implementar um sistema de conformidade fiscal digital para atender às obrigações do SPED (Sistema Público de Escrituração Digital). Decomponha este problema complexo em subproblemas menores e mais gerenciáveis.

Para cada subproblema identificado:

1. Explique o que precisa ser feito
2. Indique quais recursos seriam necessários
3. Estabeleça uma ordem lógica de implementação
4. Identifique possíveis desafios e como superá-los

Após a decomposição inicial, detalhe especificamente como resolver o subproblema da "Validação e consistência dos dados fiscais antes do envio ao SPED".

#### 1.2 Resultado Obtido

Respostas mais detalhadas e precisas, com melhor cobertura dos aspectos do problema.

### 2. Análise Crítica

#### 2.1 Pontos Fortes

Melhora a clareza e a organização do raciocínio.

#### 2.2 Limitações Encontradas

Pode aumentar a complexidade do prompt e o tempo de resposta.

#### 2.3 Possíveis Melhorias

Automatizar a decomposição para garantir consistência e eficiência.

### 3 Conclusão

A decomposição é fundamental para lidar com problemas complexos, tornando o raciocínio mais gerenciável e eficaz.

---

## Raciocínio com Ferramentas Auxiliares (Tool-Augmented Reasoning)

Uso de ferramentas externas ou módulos especializados para complementar o raciocínio do modelo.

### 1. Implementação

#### 1.1 Prompt Utilizado

**[INPUT]**

Você é um consultor fiscal com acesso a ferramentas de cálculo, bases de dados legislativas e sistemas de análise. Um cliente apresenta o seguinte caso:

"Empresa varejista com faturamento anual de R\$ 4.200.000 está avaliando mudar do regime de Lucro Presumido para o Lucro Real. Precisa saber qual seria o impacto financeiro desta mudança considerando os seguintes dados do último ano:

- Receita bruta de vendas: R\$ 4.200.000
- Custo das mercadorias vendidas: R\$ 2.500.000
- Despesas operacionais: R\$ 800.000
- Despesas financeiras: R\$ 150.000"

Utilizando suas ferramentas:

1. Calcule os tributos devidos em ambos os regimes (IRPJ, CSLL, PIS, COFINS)
2. Compare os resultados com cálculos precisos
3. Consulte a legislação atual para verificar obrigações acessórias em cada regime
4. Analise cenários de crescimento para os próximos 3 anos
5. Recomende a opção mais vantajosa com base em dados concretos

Mostre como cada ferramenta auxiliou seu raciocínio e como integrou os resultados na sua análise final.

#### 1.2 Resultado Obtido

Respostas mais precisas em domínios específicos que exigem cálculos ou dados externos.

### 2. Análise Crítica

#### 2.1 Pontos Fortes

- Expande as capacidades do modelo além do conhecimento interno.

#### 2.2 Limitações Encontradas

Integração complexa e dependência da disponibilidade das ferramentas.

## 2.3 Possíveis Melhorias

Desenvolver interfaces mais robustas e integradas para ferramentas auxiliares.

## 3 Conclusão

O uso de ferramentas auxiliares amplia o alcance do modelo, sendo essencial para tarefas que requerem dados ou cálculos externos.

---

# Raciocínio com Memória e Contexto (Memory and Contextual Reasoning)

## 1. Implementação

Incorporação de informações de contexto e memória para manter coerência em diálogos longos.

### 1.1 Prompt Utilizado

#### **[INPUT]**

Vamos simular uma consultoria fiscal contínua. Eu serei o cliente e você o consultor tributário. Ao longo de nossa conversa, você deve se lembrar de todas as informações que eu fornecer sobre minha empresa e usar essas informações para dar orientações consistentes e contextualizadas.

Primeira consulta: "Olá, sou proprietário de uma rede de farmácias com 5 lojas em São Paulo. Estamos faturando cerca de R\$ 9 milhões por ano e operamos no regime de Lucro Real. Recentemente, abrimos uma nova unidade em Minas Gerais e estou com dúvidas sobre como proceder com a apuração do ICMS interestadual."

[Após sua resposta, farei novas consultas que exigirão que você se lembre do contexto anterior]

### 1.2 Resultado Obtido

Melhora na consistência das respostas ao longo da interação.

## 2. Análise Crítica

### 2.1 Pontos Fortes

- Permite raciocínio contextualizado e continuidade na conversa.

## 2.2 Limitações Encontradas

Limitações na capacidade de memória do modelo podem afetar o desempenho.

## 2.3 Possíveis Melhorias

Implementar mecanismos de resumo e recuperação eficiente de contexto.

## 3. Conclusão

A gestão eficaz de memória e contexto é crucial para diálogos prolongados e raciocínio coerente.

---

# MCP (Model Context Protocol)

## 1. Implementação

Protocolo para gerenciar o contexto do modelo, definindo como e quando o contexto é atualizado e utilizado.

### 1.1 Prompt Utilizado

#### **[INPUT]**

Imagine que você é parte de um sistema integrado de consultoria fiscal que utiliza o Model Context Protocol para compartilhar informações entre diferentes módulos especializados. Você precisa implementar um fluxo de trabalho para análise de documentos fiscais que envolve:

1. Módulo de extração de dados de notas fiscais
2. Módulo de validação fiscal e tributária
3. Módulo de classificação contábil
4. Módulo de geração de relatórios gerenciais

Descreva detalhadamente:

1. Como estruturar o contexto compartilhado entre os módulos
2. Quais informações cada módulo deve extrair e armazenar no contexto
3. Como garantir a consistência das informações entre os módulos
4. Como o sistema utilizaria o contexto acumulado para fornecer insights mais precisos ao longo do tempo

Demonstre com um exemplo concreto como seria o fluxo de processamento de uma nota fiscal eletrônica desde sua recepção até a geração do relatório final.

1.2 Resultado Obtido

Melhora na gestão do contexto, evitando perda de informações relevantes. Todos os modelos simulam um MCP.

2. Análise Crítica

2.1 Pontos Fortes

Organiza o fluxo de informações para otimizar o uso do contexto.

2.2 Limitações Encontradas

Requer configuração cuidadosa para evitar sobrecarga ou perda de contexto.

2.3 Possíveis Melhorias

Automatizar ajustes do protocolo com base no tipo de tarefa e interação.

3. Conclusão

O MCP é uma abordagem estruturada para maximizar a eficiência do uso do contexto, essencial para aplicações complexas.

---

Comparativo dos Métodos

Método	Eficácia	Custo Computacional	Facilidade de Implementação	Aplicabilidade
Prompting e In-context Learning	Alta	Baixo	Alta	Tarefas com exemplos claros e repetitivos
Rationale Engineering	Média	Moderado	Moderada	Aplicações que exigem explicações detalhadas
Chain-of-Thought	Alta	Moderado	Moderada	Problemas complexos que exigem raciocínio
Self-Consistency Sampling	Alta	Alto	Baixa	Tarefas críticas que exigem alta confiabilidade

Método	Eficácia	Custo Computacional	Facilidade de Implementação	Aplicabilidade
Decomposição de Problemas	Alta	Moderado	Moderada	Problemas complexos divididos em etapas
Tool-Augmented Reasoning	Alta	Variável	Baixa	Tarefas que requerem dados externos ou cálculos
Memory and Contextual Reasoning	Média	Moderado	Moderada	Diálogos longos e raciocínio contextualizado
MCP	Média	Baixo	Moderada	Gestão eficiente do contexto em interações

## Discussão sobre a Escolha do Método

A escolha do método de prompting depende da complexidade da tarefa, recursos computacionais disponíveis e necessidade de explicação ou interação. Por exemplo, para questões fiscais complexas, o Chain-of-Thought é preferido pela sua capacidade de decompor problemas, enquanto para tarefas que exigem alta confiabilidade, a amostragem de autoconsistência é recomendada.

## Análise de Erros Comuns

- **Chain-of-Thought:** Propagação de erros em etapas iniciais.
- **Self-Consistency Sampling:** Custo computacional elevado.
- **Tool-Augmented Reasoning:** Falhas na integração das ferramentas.
- **Memory and Contextual Reasoning:** Perda de contexto em diálogos longos.

Mitigações incluem verificação em etapas, otimização de amostras, desenvolvimento robusto de interfaces e mecanismos de resumo de contexto.

## Considerações Éticas

O uso de agentes autônomos para reasoning deve considerar a transparência das decisões, responsabilidade pelos resultados e mitigação de vieses. É fundamental garantir que as respostas sejam auditáveis e que haja supervisão humana em decisões críticas.

## Conclusão

Este trabalho demonstrou que diferentes métodos de prompting oferecem vantagens específicas para aplicações em consultoria fiscal. A combinação e adaptação desses métodos, aliada a melhorias contínuas, são essenciais para maximizar a eficácia e confiabilidade dos agentes autônomos.

## Observações para Avaliação

- **Este trabalho foi desenvolvido individualmente.** No E-mail enviado de formação de grupo veio somente o meu nome e e-mail sem os outros membros do grupo e sem respostas ao e-mail quando questionado.
- Todas as fontes externas utilizadas foram devidamente citadas
- **O prompt foi testado em** GROK 3, GPT4 e GEMINI 2.5 PREVIEW.

# Referências

- Wei, J., et al. (2022). Chain of Thought Prompting Elicits Reasoning in Large Language Models. \*arXiv\*.
- Nye, M., et al. (2021). Show Your Work: Scratchpads for Intermediate Computation with Language Models. \*ICLR\*.
- Kojima, T., et al. (2022). Large Language Models are Zero-Shot Reasoners. \*arXiv\*.
- Wang, X., et al. (2022). Self-Consistency Improves Chain of Thought Reasoning in Language Models. \*NeurIPS\*.
- Clark, P., et al. (2020). Transformers as Soft Reasoners over Language. \*EMNLP\*.
- Shinn, T., et al. (2023). Toolformer: Language Models Can Teach Themselves to Use Tools. \*arXiv\*.
- Lample, G., et al. (2022). Augmented Language Models: A Survey. \*arXiv\*.
- Weston, J., et al. (2014). Memory Networks. \*ICLR\*.
- Khandelwal, U., et al. (2020). Generalization through Memorization: Nearest Neighbor Language Models. \*ICLR\*.
- Chen, M., et al. (2023). Context Management in Large Language Models: Protocols and Practices. \*arXiv\*.
- Li, X., et al. (2022). Efficient Context Handling for Language Models. \*ACL\*.