

Teste de Métodos de Raciocínio em Inteligência Artificial

Maio 2025

1 Introdução

Este documento apresenta a descrição, testes, resultados e conclusões de nove métodos de raciocínio em IA, aplicados a um problema matemático. Inclui referências bibliográficas.

2 Descrição dos Métodos

- **Ajuste Fino Totalmente Supervisionado:** Treina modelo com dados rotulados, ajustando pesos. Ideal para tarefas específicas.
- **Prompting e In-context Learning:** Usa prompts com exemplos, sem ajustar pesos, aproveitando conhecimento pré-treinado.
- **Rationale Engineering:** Prompts com explicações para guiar raciocínio, aumentando transparência.
- **Chain-of-Thought Prompting:** Solicita raciocínio passo a passo, eficaz para tarefas complexas.
- **Amostragem de Autoconsistência:** Gera múltiplas respostas, seleciona a mais consistente, reduzindo erros.
- **Decomposição de Problemas:** Divide problemas em subproblemas, resolvidos separadamente.
- **Raciocínio com Ferramentas Auxiliares:** Integra ferramentas externas (ex., calculadoras) para precisão.
- **Raciocínio com Memória e Contexto:** Usa informações passadas para decisões atuais, simulando memória.
- **Model Context Protocol (MCP):** Protocolo para conectar modelos a ferramentas externas, gerenciando contexto.

3 Testes Realizados

3.1 Tarefa

Resolver: “Um trem sai às 08:00 a 60 km/h, outro às 08:30 a 80 km/h, mesma direção. Quando e onde se encontram?” (Resposta: 10:00, 120 km).

3.2 1. Ajuste Fino Supervisionado

Metodologia: Modelo treinado com 1000 problemas de movimento uniforme.
Prompt: “Resolva o problema.”

Resultado: “10:00, 120 km.” Rápido, preciso, sem explicação.

Observação: Alta precisão, exige dados e treinamento.

3.3 2. Prompting e In-context Learning

Metodologia: Prompt com 3 exemplos resolvidos:

- Exemplo 1: “Trem A sai às 09:00 a 50 km/h, Trem B às 09:30 a 70 km/h. Encontro: 11:00, 100 km. (A: $50 \times 0,5 = 25$ km, pois $09:30 - 09:00 = 0,5$ h; B alcança em $25 \div (70 - 50) = 1,25$ h; $09:30 + 1,25$ h = 11:00; $50 \times 2 = 100$ km).”
- Exemplo 2: “Trem A sai às 07:00 a 40 km/h, Trem B às 07:15 a 60 km/h. Encontro: 08:00, 80 km. (A: $40 \times 0,25 = 10$ km, pois $07:15 - 07:00 = 0,25$ h; B alcança em $10 \div (60 - 40) = 0,5$ h; $07:15 + 0,5$ h = 08:00; $40 \times 1 = 80$ km).”
- Exemplo 3: “Trem A sai às 10:00 a 30 km/h, Trem B às 10:30 a 50 km/h. Encontro: 12:00, 90 km. (A: $30 \times 0,5 = 15$ km, pois $10:30 - 10:00 = 0,5$ h; B alcança em $15 \div (50 - 30) = 0,75$ h; $10:30 + 0,75$ h = 12:00; $30 \times 3 = 90$ km).”

Prompt: “Resolva: [problema].”

Resultado: “10:00, 120 km.” Correto.

Observação: Rápido, depende de exemplos claros.

3.4 3. Rationale Engineering

Metodologia: Prompt: “Explique a lógica antes de responder: [problema].” Exemplo: “Trem A sai às 09:00 a 50 km/h, B às 09:30 a 70 km/h. Lógica: A percorre 25 km até 09:30; B alcança em 1,25 h; encontro às 11:00, 100 km.”

Resultado: “Lógica: 1. Primeiro trem sai às 08:00, até 08:30 (0,5 h, pois $08:30 - 08:00 = 30$ min = 0,5 h), percorre $60 \times 0,5 = 30$ km. 2. Velocidade relativa: $80 - 60 = 20$ km/h. 3. Tempo para B alcançar: $30 \div 20 = 1,5$ h. 4. $08:30 + 1,5$ h = 10:00. 5. Distância: $60 \times 2 = 120$ km. Resposta: 10:00, 120 km.” Correto, claro.

Observação: Transparente, similar a CoT.

3.5 4. Chain-of-Thought Prompting

Metodologia: Prompt: “Pense passo a passo: [problema].”

Resultado: “1. Primeiro trem sai às 08:00, até 08:30 (0,5 h, pois $08:30 - 08:00 = 30$

min = 0,5 h), percorre $60 \times 0,5 = 30$ km. 2. Velocidade relativa: $80 - 60 = 20$ km/h. 3. Tempo para B alcançar: $30 \div 20 = 1,5$ h. 4. 08:30 + 1,5 h = 10:00. 5. Distância: $60 \times 2 = 120$ km. Resposta: 10:00, 120 km.” Correto, detalhado.

Observação: Eficaz, aumenta tempo de resposta.

3.6 5. Amostragem de Autoconsistência

Metodologia: Gera 5 respostas com prompt: “Resolva: [problema].” Seleciona mais comum por votação.

Resultado:

- Resposta 1: “10:00, 120 km.”
- Resposta 2: “10:00, 120 km.”
- Resposta 3: “10:00, 120 km.”
- Resposta 4: “10:30, 130 km.” (errada)
- Resposta 5: “10:00, 120 km.”

Final: “10:00, 120 km.” (4/5 correto).

Observação: Reduz erros, custo computacional alto.

3.7 6. Decomposição de Problemas

Metodologia: Divide em:

1. Distância inicial: Calcular distância do primeiro trem até 08:30.
2. Tempo de alcance: Tempo para o segundo trem alcançar o primeiro.
3. Ponto de encontro: Calcular horário e distância final.

Prompt: “Resolva cada parte separadamente.”

Resultado:

1. Distância inicial: Até 08:30 (0,5 h, pois $08:30 - 08:00 = 30$ min = 0,5 h), $60 \times 0,5 = 30$ km.
2. Tempo de alcance: Velocidade relativa = $80 - 60 = 20$ km/h; Tempo = $30 \div 20 = 1,5$ h; $08:30 + 1,5$ h = 10:00.
3. Ponto de encontro: Tempo total desde 08:00 = 2 h; Distância = $60 \times 2 = 120$ km.

Resposta: “10:00, 120 km.” Correto.

Observação: Estruturado, escalável.

3.8 7. Raciocínio com Ferramentas Auxiliares

Metodologia: Usa calculadora simulada. Prompt: “Use ferramenta para cálculos: [problema].” Calculadora executa: multiplicação ($60 \times 0,5$), subtração ($80 - 60$), divisão ($30 \div 20$), adição ($08:30 + 1,5$ h), multiplicação (60×2).

Resultado: “Distância inicial: Até 08:30 (0,5 h, pois $08:30 - 08:00 = 30 \text{ min} = 0,5 \text{ h}$), $60 \times 0,5 = 30 \text{ km}$ (calculadora). Velocidade relativa: $80 - 60 = 20 \text{ km/h}$ (calculadora). Tempo: $30 \div 20 = 1,5 \text{ h}$ (calculadora). Horário: $08:30 + 1,5 \text{ h} = 10:00$. Distância: $60 \times 2 = 120 \text{ km}$ (calculadora). Resposta: 10:00, 120 km.” Correto, preciso.

Observação: Exato, requer integração.

3.9 8. Raciocínio com Memória e Contexto

Metodologia: Simula conversa:

- Contexto: “Você resolveu: Trem A sai às 09:00 a 50 km/h, B às 09:30 a 70 km/h. Encontro: 11:00, 100 km. (A: $50 \times 0,5 = 25 \text{ km}$, pois $09:30 - 09:00 = 0,5 \text{ h}$; B alcança em $25 \div (70 - 50) = 1,25 \text{ h}$; $09:30 + 1,25 \text{ h} = 11:00$; $50 \times 2 = 100 \text{ km}$).”
- Prompt: “Use esse conhecimento para resolver: [problema].”

Resultado: “Baseado no problema anterior: Primeiro trem sai às 08:00, até 08:30 (0,5 h, pois $08:30 - 08:00 = 30 \text{ min} = 0,5 \text{ h}$), $60 \times 0,5 = 30 \text{ km}$. Velocidade relativa: $80 - 60 = 20 \text{ km/h}$. Tempo: $30 \div 20 = 1,5 \text{ h}$. $08:30 + 1,5 \text{ h} = 10:00$. Distância: $60 \times 2 = 120 \text{ km}$. Resposta: 10:00, 120 km.” Correto.

Observação: Útil em continuidade, depende de contexto claro.

3.10 9. Model Context Protocol (MCP)

Metodologia: Simula MCP conectando modelo a calculadora via protocolo. Prompt: “Acesse ferramenta via MCP para resolver: [problema].” MCP envia cálculos ($60 \times 0,5$, $80 - 60$, $30 \div 20$, $08:30 + 1,5 \text{ h}$, 60×2) à ferramenta e recebe resultados.

Resultado: “Ferramenta calcula: Distância inicial: Até 08:30 (0,5 h, pois $08:30 - 08:00 = 30 \text{ min} = 0,5 \text{ h}$), $60 \times 0,5 = 30 \text{ km}$. Velocidade relativa: $80 - 60 = 20 \text{ km/h}$. Tempo: $30 \div 20 = 1,5 \text{ h}$. Horário: $08:30 + 1,5 \text{ h} = 10:00$. Distância: $60 \times 2 = 120 \text{ km}$. Resposta: 10:00, 120 km.” Correto.

Observação: Padronizado, similar a ferramentas.

4 Resultados e Conclusões

Conclusões:

- **Ajuste Fino:** Melhor para tarefas específicas, caro.
- **In-context:** Rápido, menos robusto.
- **Rationale e CoT:** Excelentes para transparência.
- **Autoconsistência:** Reduz erros, aumenta custo.
- **Decomposição:** Ideal para problemas grandes.
- **Ferramentas e MCP:** Garantem precisão, dependem de infraestrutura.

Método	Acurácia	Clareza	Tempo	Observação
Ajuste Fino	Alta	Baixa	Rápido	Preciso, exige treinamento.
In-context	Média	Média	Rápido	Depende de exemplos.
Rationale	Alta	Alta	Médio	Transparente, similar a CoT.
CoT	Alta	Alta	Lento	Ideal para lógica complexa.
Autoconsistência	Alta	Média	Lento	Reduz erros, custo alto.
Decomposição	Alta	Alta	Médio	Estruturado, escalável.
Ferramentas	Alta	Alta	Médio	Preciso, requer integração.
Memória	Alta	Média	Rápido	Útil em continuidade.
MCP	Alta	Alta	Médio	Padronizado, similar a ferramentas.

Table 1: Comparação dos Métodos

- **Memória:** Eficaz em cenários contínuos.

Aplicações: Escolha depende de precisão, transparência ou rapidez.

5 Referências

- Wei, J., et al. (2022). “Chain-of-Thought Prompting Elicits Reasoning in Large Language Models.” *arXiv:2201.11903*.
- Wang, X., et al. (2022). “Self-Consistency Improves Chain of Thought Reasoning in Language Models.” *arXiv:2203.11171*.
- Gao, L., et al. (2023). “Tool-Augmented Reasoning.” *arXiv:2306.02408*.
- Anthropic. (2023). “Model Context Protocol.” anthropic.com.