

Análise Comparativa de LLMs na Elaboração de Resumos de Editais Públicos: Uma Abordagem com Estratégias de Engenharia de Prompt

Davi Costa Mendes da Silva
Ryan Souza Santana
Discentes – CCNT/UEPA
davi.cmdsilva@aluno.uepa.br
ryan.ssantana@aluno.uepa.br



ROTEIRO

- Problema
- Justificativa
- Hipótese
- Objetivos
- Metodologia
- Resultados esperados
- Cronograma
- Orçamento

PROBLEMA

- Pergunta norteadora: Qual LLM disponível ao público é mais confiável e melhor performa no resumo de editais aplicando diferentes técnicas de engenharia de prompt considerando critérios de coerência e integridade da informação?

JUSTIFICATIVA

No mundo onde somos rodeados por informação, o processo de sumarização se torna indispensável para conseguir condensar o conhecimento. Diante disso, muito se pesquisa sobre a criação de sumarização automáticas, mais especificamente, no uso de LLMs para resumos abstrativos. Essa prática já é viável com modelos modernos, no entanto, ainda se investiga sobre a qualidade e a veracidade das informações empregadas nos textos gerados.

Ademais, pouco se dedicou a avaliar criticamente a coerência lógica e a integridade factual dos resumos gerados por LLMs em cenários de uso real, isto é, não controlados por métricas acadêmicas ou conjuntos de problemas idealizados.

HIPÓTESE

- Considerando o avanço das LLMs na produção de texto, pressupõe-se que a ocorrência de variações significativas e erros na geração de conteúdo pelos modelos testados seja baixa. As principais discrepâncias nos resultados tendem a estar associadas às diferenças nos prompts utilizados.

OBJETIVOS

O objetivo dessa pesquisa é determinar, dentre 4 IAs generativas, ChatGPT, Gemini, DeepSeek e Copilot, qual performa melhor no resumo abstrativo de editais. Para tanto, os seguintes objetivos específicos deverão ser atingidos:

- Definição das melhores técnicas para construção dos prompts para uso na criação dos resumos;
- Avaliação dos resultados da interação cada um dos modelos com cada prompt construído e seleção os melhores resumos;
- Comparação dos melhores resumos e seleção da melhor opção de modelo e modelo/prompt para sumarização.

METODOLOGIA

- **Coleta dos editais:** Serão escolhidos 5 a 10 editais públicos representativos de diferentes áreas (concursos, licitações, bolsas, ingressos).
- **Construção dos prompts:** Serão desenvolvidos 3 instruções, uma simplificada, simulando um usuário normal, mais duas utilizando engenharia de prompt (Chain-of-Thought e Extrativa-Abstrativa), todas exigirão uma estrutura em tópicos (título, resumo geral, cargos e vagas, requisitos, critérios, datas, taxas e valores).
- **Geração dos resumos:** Será gerado um resumo para cada combinação de modelo/prompt. Serão feitos manualmente pelos sites oficiais com parâmetros padrões, simulando o uso do dia-a-dia. Se o resumo seguir a estrutura proposta, ele será catalogado e armazenado.

METODOLOGIA

- **Avaliação dos resumos:** Um formulário será desenvolvido para coletar a avaliação de cada resumo nas métricas de relevância, consistência, fluência e coerência. Utilizaremos o Google forms, montando um formulário para cada edital e suas seções representarão seus resumos. Na primeira parte do formulário consta instruções sobre a avaliação e uma descrição das métricas e um link para o edital que foi resumido. Nas seções, haverá um link para o resumo em questão, um para o texto original e uma estrutura para avaliar de 1 a 5 cada métrica. Todos os formulários serão preenchidos pelos avaliadores voluntários.

METODOLOGIA

Avaliação dos resumos do Edital

Instruções

- Leia cuidadosamente o Edital, esteja atento às informações que ele contém.
- Leia os resumos propostos.
- Avalie cada resumo em uma escala de 1 (pior) a 5 (melhor) com base em sua relevância, consistência, fluência e coerência.
- #Explicar caso de invalidez#

Definições

Relevância: A avaliação mede o quanto bem o resumo captura os pontos chave do artigo. Considere se todos os aspectos importantes (e somente eles) estão contidos no resumo.

Consistência: A avaliação mede se os fatos no resumo são consistentes com os fatos do artigo original. Considere se o resumo reproduz todos os fatos com precisão e não inventa informações falsas.

Fluência: Esta avaliação mede a qualidade das sentenças individualmente, se estão bem escritas e gramaticalmente corretas. Considere a qualidade das sentenças individualmente.

Coerência: Esta avaliação mede a qualidade de todas as sentenças coletivamente, se se encaixam e soam naturalmente. Considere a qualidade do resumo como um todo.

Edital analisado: ##Título do Edital## [@Link do Edital](#)

[Próxima](#)

[Limpar formulário](#)

Avaliação dos resumos do Edital

Resumo

Edital analisado: ##Título do Edital## [@Link do Edital](#)
Resumo avaliado: ##Identificação do Resumo## [@Link do Resumo](#)

	1	2	3	4	5
Relevância	<input type="radio"/>				
Consistência	<input type="radio"/>				
Fluência	<input type="radio"/>				
Coerência	<input type="radio"/>				

[Voltar](#)

[Próxima](#)

[Limpar formulário](#)

METODOLOGIA

- **Classificação das avaliações:** As avaliações serão transferidas para o Google Planilhas, iremos organizar os resumos, identificando cada um através da disposição deles em cada Formulário e separando eles por modelo e prompt. Ao fim desse processo, haverá uma tabela por edital, destacando as notas de cada resumo e segmentando por prompt e modelo.
- **Comparação e análise das melhores avaliações:** Realizaremos o ranqueamento dos modelos, dos prompts e de suas combinações. Para isso, será calculado as média das notas obtidas. Com esses dados, se identificará qual combinação, modelo isoladamente e prompt isoladamente obteve melhores resultados.

RESULTADOS ESPERADOS

Após o processo realizado pela metodologia apresentada, definimos três resultados que podem servir de marco da pesquisa realizada:

- Espera-se construir prompts eficientes tanto para um usuário leigo quanto para o profissional conhecedor de engenharia de prompt que atendam a demanda exigida.
- Espera-se obter uma base de resumos avaliados que cumpra seu papel nessa pesquisa e que sirva para futuros trabalhos.
- Com comparação dos resultados obtidos, espera-se ser extraído um ranqueamento dos modelos, prompts e combinações e comentários sobre os casos observados.

CRONOGRAMA

Etapas	MESSES					
	1	2	3	4	5	6
Coleta dos editais	X					
Construção dos <i>prompts</i>	X	X♦				
Geração dos resumos		X	X			
Avaliação dos resumos			X	X	X♦	
Classificação dos editais avaliados					X	
Comparação e análise						X
Classificação dos LLMs avaliados						X♦

ORÇAMENTO

- Não haverá necessidade de investimentos nesta pesquisa, visto que trata-se de um experimento que utiliza versões gratuitas e acessíveis ao usuário geral.