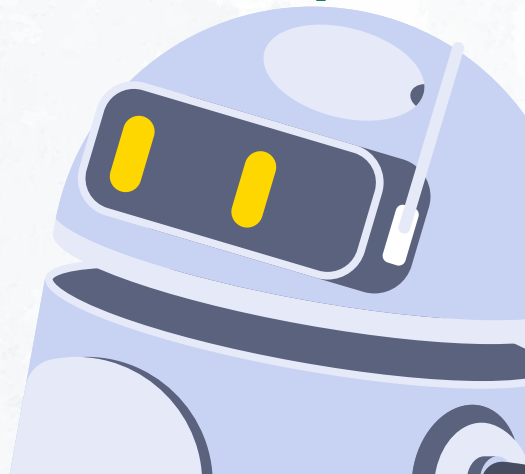


# Avaliação do Desempenho de Modelos LLaMa e Gemini na Correção de Redações do ENEM

Autor: Daniel Silva Lopes da Costa NUSP 11302720



# 01 Introdução

## Contexto

- A redação avalia cinco competências:
  - Domínio da norma culta(C1),
  - Compreensão da proposta(C2),
  - Organização de argumentos(C3),
  - Coesão textual(C4),
  - Proposta de intervenção(C5).
- A correção **manual é trabalhosa, dispendiosa e pode conter erros;**



# Objetivos

1. Dividir a análise em duas fases:  
**(1) comparar padrões de prompts;**  
**(2) integrar as saídas dos modelos em algoritmos supervisionados.**
2. Introduzir um dataset estendido com redações nota mil, reequilibrando distribuições de notas para análises mais precisas.

## Justificativa

- **LLMs:** O trabalho explora grandes modelos de linguagem (LLaMA e Gemini) em nova tarefa.
- **Correção automática de textos(AES):** A pesquisa preenche lacunas na avaliação automática de textos em português.
- **Social:** Ferramentas automáticas de correção democratizam o acesso à educação.



## 02 Revisão Literária

# Processamento de Linguagem Natural

Subárea da IA que permite máquinas entenderem e gerarem linguagem humana.

Evolução:

- **Métodos baseados em regras (anos 50-60):** gramáticas formais.
- **Métodos estatísticos (anos 90):** modelos probabilísticos (ex.: HMM, Naive Bayes).
- **Aprendizado profundo (atual):** redes neurais e Transformers (ex.: BERT, GPT).

## Transformers

Arquitetura baseada em atenção (self-attention), eficiente no processamento paralelo de sequências.

- **Encoder:** Gera representação da entrada. - Bert
- **Decoder:** Produz a saída baseada na entrada codificada. - GPT



# Correção Automática de Redações (AES)

Tecnologia que utiliza IA para avaliar textos com base em critérios como correção linguística, coerência e relevância semântica.

Evolução:

- 1960: Início com o PEG, baseado em características superficiais (ex.: comprimento de frases).
- 1990: Avanço com o uso de métodos estatísticos e PNL (ex.: e-rater, utilizado no GRE e TOEFL).

**AES em Português:** Escassez de corpora anotados e técnicas avançadas para o português brasileiro.





# 03 Ferramentas



## Modelos

### LLaMA-3.1 70B (Meta AI):

- **Parâmetros:** 70 bilhões em 80 camadas com 64 cabeças de atenção.
- **Janela de contexto:** 128.000 tokens. Suporte aprimorado para idiomas não-ingleses e tarefas de longo alcance.
- **Dados de treinamento:** 15T de tokens (50% conhecimento geral, 25% raciocínio/matемática, 17% código, 8% multilinguísticos).
- 

### Gemini 1.5-Flash (Google):

- **Janela de contexto:** Até 2 milhões de tokens.
- **Desempenho multimodal:** Integra texto, imagens, áudio e vídeo em um único fluxo.
- **Eficiência computacional:** Baixa latência com TPUs e cálculo paralelo.
- **Dados de treinamento:** Base multimodal abrangente.

# Modelos



## NILC-Metrix:

- **Descrição:** Ferramenta para análise de coesão, coerência e complexidade textual em português brasileiro.
- **Métricas:** Mais de 200 agrupadas em 14 categorias (ex.: frequência lexical, análise semântica).
- **Disponibilidade:** Código aberto (AGPLv3), interface web e API.

## XGBoost:

- **Descrição:** Algoritmo baseado em árvores de decisão otimizado para classificação/regressão.
- **Destaques:**
  - Estratégias como softmax para multi classificação.
  - Lida bem com dados desequilibrados.
- **Hiperparâmetros ajustáveis:** Profundidade máxima (max\_depth), taxa de aprendizado (learning\_rate).

# Datasets

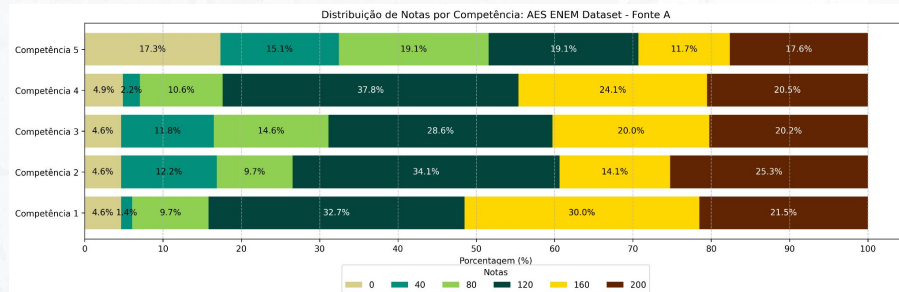
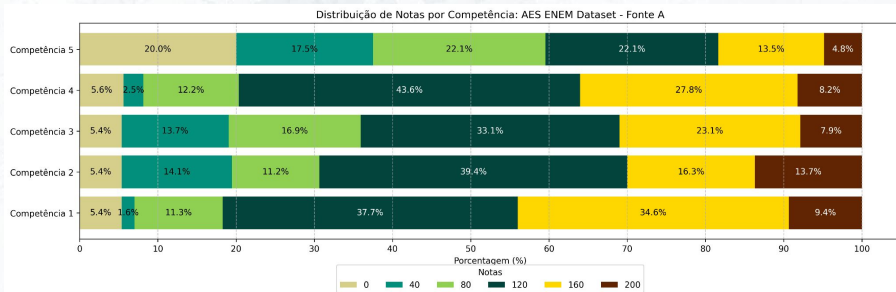
## AES ENEM Dataset - Fonte A:

- **Origem:** 1.629 redações (2015-2020) de um simulador do ENEM.
- **Limpeza:** 474 redações removidas por inconsistências, resultando em 1.155 textos finais.
- **Anotações:** Notas em 5 competências (0-200) e comentários qualitativos.

## Redações Nota Mil:

- **Origem:** 179 redações coletadas de fontes públicas (G1, O Globo, Lucas Felpi).
- **Processamento:**
  - Uso da métrica de Levenshtein para eliminar duplicatas.
  - Validação manual para garantir precisão.
- **Integração:** Incorporadas ao AES Dataset para balancear notas altas.

	AES ENEM Dataset	Redações Nota Mil	Dataset Estendido
Treino	744	129	873
Teste	216	27	243
Validação	195	23	219
Total	1155	179	1334





# Métricas

## Quadratic Weighted Kappa (QWK)

- Mede a concordância entre classificações ordinais.
- Penaliza discrepâncias maiores entre categorias.
- Varia de -1 (discordância total) a 1 (concordância perfeita).

## Root Mean Square Error (RMSE)

- Avalia a média das diferenças quadráticas entre valores previstos e reais.
- Penaliza mais fortemente grandes erros.
- Valores mais baixos indicam maior precisão.

## Acurácia

- Proporção de predições corretas sobre o total.
- Calculada com correspondência exata entre notas previstas e reais.
- Resultados variam de 0 a 1 (1 = perfeição).

## Precisão do Intervalo Absoluto (PIA)

- Mede a proximidade entre notas previstas e reais dentro de um intervalo tolerado.
- Tolerância padrão: 80 pontos (seguindo critério do INEP).

# 04

# Metodologia e Resultados

## Fase 1 - Engenharia de Prompt

- Teste de LLaMA e Gemini utilizando **chamadas de API**;
- Todos os experimentos foram feitos em **zero-shot** sem dados de treinamento;
- Por limitações de tempo e recurso foi possível executar cada prompt para a mesma redação **20 vezes com o Gemini e 5 vezes com o LLaMA - Resultados validados com teste estatístico de Kruskal-Wallis**;



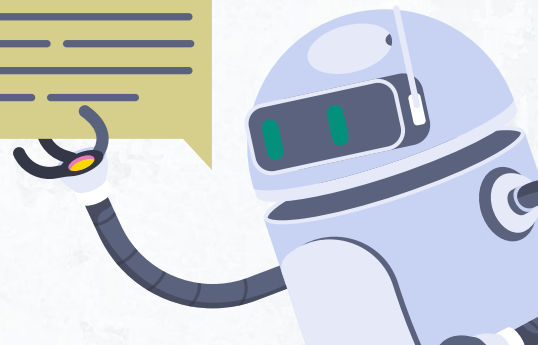
# Experimento 1 - Prompt Base

Corrija a seguinte redação, seguindo o método de avaliação do ENEM, entregue apenas a nota quantitativa entre 0 e 200 para cada competência e a nota final no seguinte formato: [x, y, z, w, t, x+y+z+w+t]. Não quero texto na resposta, apenas as notas numéricas para cada competência e o total.

Tema: <TEMA>

Textos de suporte: <SUPORTE>

Redação: <REDAÇÃO>



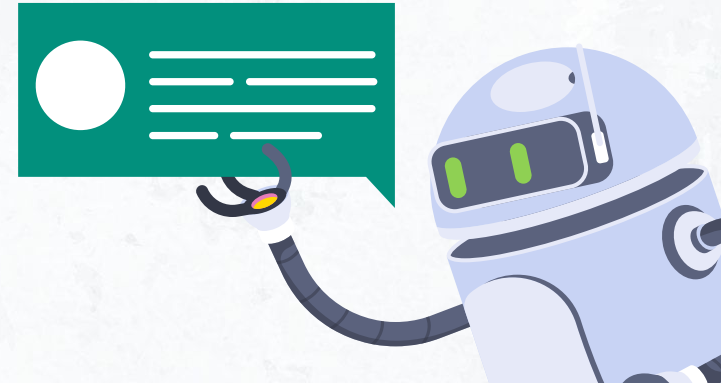
# Experimento 2 - Prompt Base em Inglês

Correct the following writing, following the ENEM assessment method, only deliver the quantitative grade between 0 and 200 for each competency and the final grade in the following format: [x, y, z, w, t, x+y+z+w+t]. I don't want text in the answer, just the numerical grades for each competency and the total.

Theme: <TEMA>

Supporting texts: <SUPORTE>

Writing: <REDAÇÃO>



# Experimento 3 - Prompt Base Melhorado

Corrija a redação abaixo seguindo o método de avaliação do ENEM. Forneça apenas as notas quantitativas entre 0 e 200 para cada competência, e a nota final, no seguinte formato:[C1,C2,C3,C4,C5,Total], onde "Total" é a soma das notas das cinco competências.

Tema:

.....

<TEMA>

.....

Textos de suporte:

.....

<SUPORTE>

.....

Redação:

.....

<REDAÇÃO>

.....

Não forneça texto na resposta, apenas as notas numéricas para competência e o total.



# Experimento 4 - Prompt Persona

**Como um professor de ensino médio especializado na correção de redações do ENEM.** Corrija a redação abaixo seguindo o método de avaliação do ENEM. Forneça apenas as notas quantitativas entre 0 e 200 para cada competência, e a nota final, no seguinte formato:[C1,C2,C3,C4,C5,Total], onde "Total" é a soma das notas das cinco competências.

Tema:

""

<TEMA>

""

Textos de suporte:

""

<SUPORTE>

""

Redação:

""

<REDAÇÃO>

""

Não forneça texto na resposta, apenas as notas numéricas para competência e o total.

# Experimento 5 - Prompt Cadeia de Pensamento

Qual o método de avaliação utilizado na correção de redações do ENEM?

Corrija a redação abaixo seguindo o método de avaliação do ENEM. Atribuindo notas quantitativas entre 0 e 200 para cada uma das cinco competências. Forneça uma explicação detalhada que justifique cada nota.

Tema:

.....

<TEMA>

.....

Textos de suporte:

.....

<SUPORTE>

.....

Redação:

.....

<REDAÇÃO>

.....

Não tenho dados adicionais que possam contribuir, use apenas a sua própria base de dados para fornecer as notas seguindo o formato do ENEM.

Com os dados da resposta anterior, forneça apenas as notas quantitativas entre 0 e 200 para cada competência, e a nota final, no seguinte formato:[C1,C2,C3,C4,C5,Total], onde "Total" é a soma das notas das cinco competências.

Não forneça texto na resposta, apenas as notas numéricas para competência e o total.

# Experimento 6 - Prompt Contextualizado

Conheça as cinco competências cobradas pelo Inep na redação do ENEM:

## 1. Domínio da escrita formal da língua portuguesa

É avaliado se a redação do participante está adequada às regras de ortografia, como acentuação, ortografia, uso de hífen, emprego de letras maiúsculas e minúsculas e separação silábica. Ainda são analisadas a regência verbal e nominal, concordância verbal e nominal, pontuação, paralelismo, emprego de pronomes e crase.

São seis níveis de desempenho:

**200 pontos** - Demonstra excelente domínio da modalidade escrita formal da língua portuguesa e de escolha de registro. Desvios gramaticais ou de convenções da escrita serão aceitos somente como excepcionalidade e quando não caracterizarem reincidência.

**160 pontos** - Demonstra bom domínio da modalidade escrita formal da língua portuguesa e de escolha de registro, com poucos desvios gramaticais e de convenções de escrita.

**120 pontos** - Demonstra domínio mediano da modalidade escrita formal da língua portuguesa e de escolha de registro, com alguns desvios gramaticais e de convenções da escrita.

**80 pontos** - Demonstra domínio insuficiente da modalidade escrita formal da língua portuguesa, com muitos desvios gramaticais, de escolha de registro e de convenções da escrita.

**40 pontos** - Demonstra domínio precário da modalidade escrita formal da língua portuguesa, de forma sistemática, com diversificados e frequentes desvios gramaticais, de escolha de registro e de convenções da escrita.

**0 ponto** - Demonstra desconhecimento da modalidade escrita formal da língua portuguesa.

----

## 2. Compreender o tema e não fugir do que é proposto

Avalia as habilidades integradas de leitura e de escrita do candidato. O tema constitui o núcleo das ideias sobre as quais a redação deve ser organizada e é caracterizado por ser uma delimitação de um assunto mais abrangente.

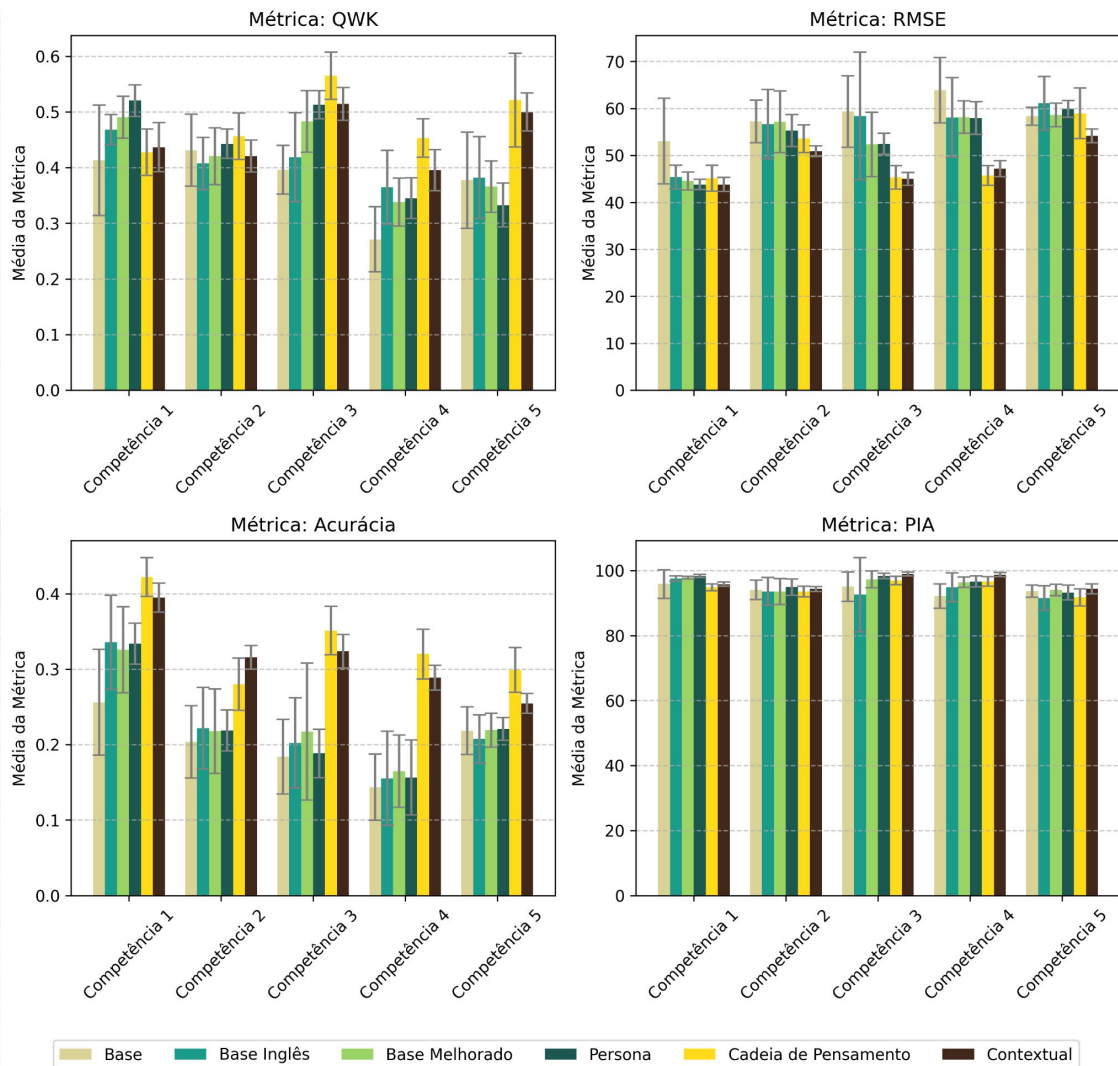
Eis os seis níveis de desempenho:

**200 pontos** - Desenvolve o tema por meio de argumentação consistente, a partir de um repertório sociocultural produtivo e apresenta excelente domínio do texto dissertativo-argumentativo. [...]



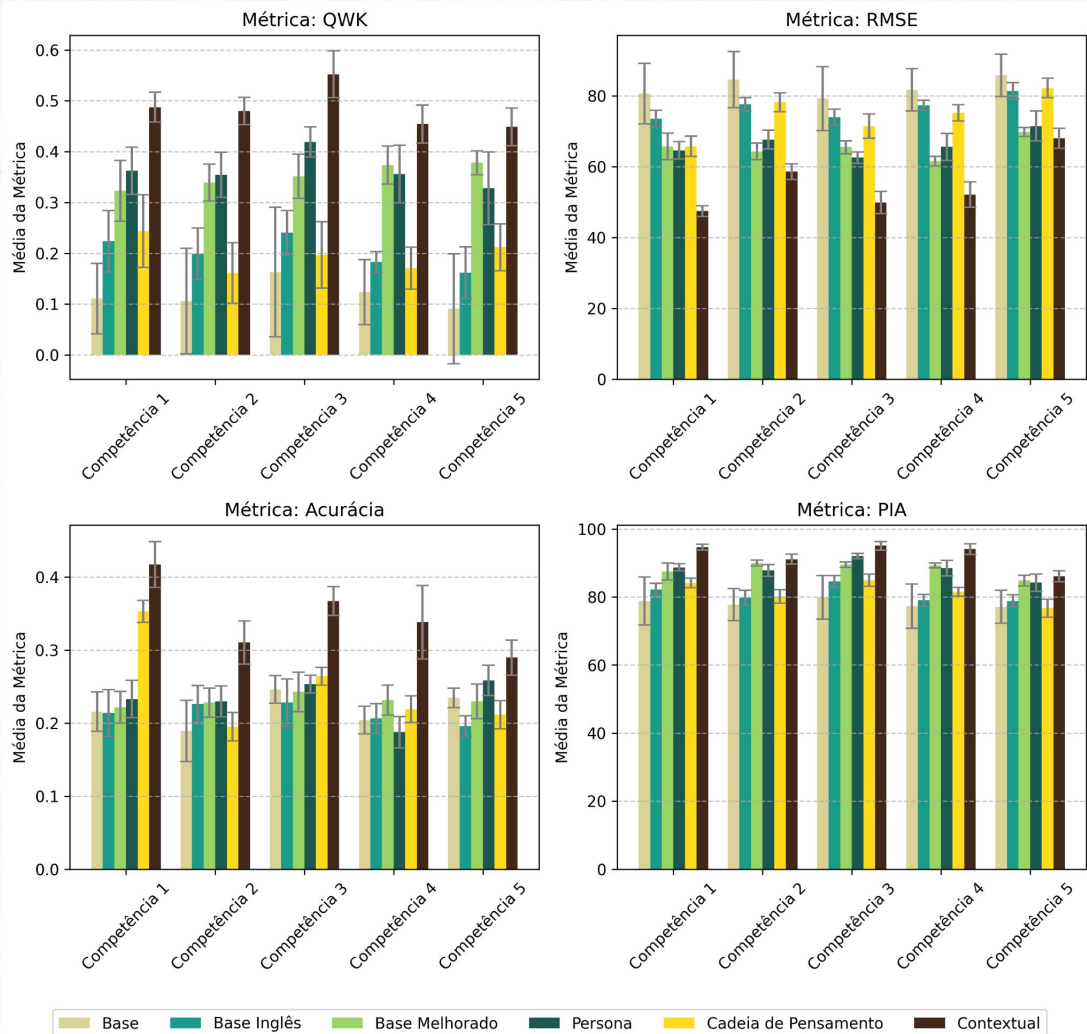
Gerado por IA

# Resultados Gemini



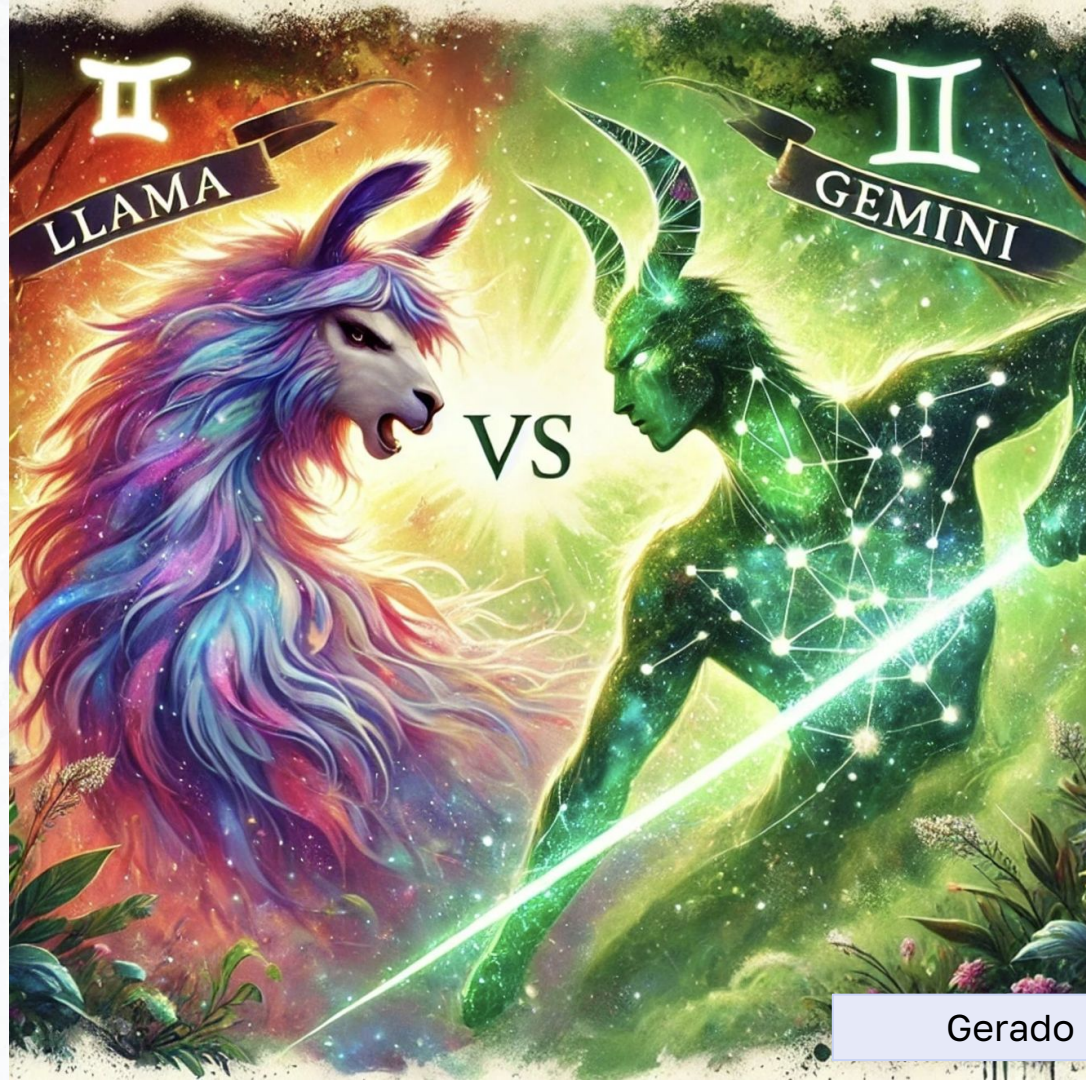


# Resultados LLaMA



Gerado por IA





Gerado por IA

Comp	Métrica	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5	Exp. 6
C1	QWK Gemini	0.413	0.468	0.490	0.520	0.427	0.437
	QWK Llama	0.111	0.224	0.323	0.362	0.243	0.487
	RMSE Gemini	53.026	45.400	44.553	43.801	45.179	43.808
	RMSE Llama	80.659	73.615	65.763	64.657	65.797	47.474
	Acurácia Gemini	0.256	0.336	0.326	0.334	0.422	0.395
	Acurácia Llama	0.216	0.214	0.221	0.233	0.353	0.417
C2	PIA Gemini	95.863	97.568	97.822	98.313	94.832	95.802
	PIA Llama	78.848	82.222	87.490	88.724	84.115	94.650
	QWK Gemini	0.431	0.407	0.420	0.443	0.456	0.420
	QWK Llama	0.106	0.199	0.339	0.354	0.161	0.480
	RMSE Gemini	57.201	56.681	57.133	55.288	53.547	50.914
	RMSE Llama	84.622	77.658	64.334	67.662	78.214	58.640
C3	Acurácia Gemini	0.204	0.222	0.218	0.219	0.280	0.315
	Acurácia Llama	0.189	0.226	0.228	0.230	0.195	0.310
	PIA Gemini	94.019	93.583	93.571	94.897	93.504	94.300
	PIA Llama	77.778	79.835	89.959	87.819	80.165	91.111
	QWK Gemini	0.396	0.419	0.483	0.513	0.565	0.515
	QWK Llama	0.163	0.240	0.351	0.419	0.197	0.552
C5	RMSE Gemini	59.318	58.387	52.328	52.477	45.312	45.002
	RMSE Llama	79.276	73.970	65.496	62.612	71.508	49.895
	Acurácia Gemini	0.184	0.202	0.217	0.188	0.351	0.324
	Acurácia Llama	0.246	0.228	0.243	0.253	0.264	0.367
	PIA Gemini	95.014	92.545	97.304	98.354	96.929	98.930
	PIA Llama	79.918	84.609	89.547	92.016	84.938	95.062

Comp	Métrica	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5	Exp. 6
C4	QWK Gemini	0.271	0.365	0.338	0.345	0.453	0.395
	QWK Llama	0.124	0.183	0.373	0.356	0.171	0.454
	RMSE Gemini	63.879	58.075	58.138	57.929	45.725	47.175
	RMSE Llama	81.738	77.290	61.580	65.623	75.243	52.124
	Acurácia Gemini	0.144	0.155	0.165	0.156	0.320	0.289
	Acurácia Llama	0.204	0.207	0.231	0.188	0.219	0.338
C5	PIA Gemini	92.127	94.825	96.379	96.626	96.617	98.704
	PIA Llama	77.284	79.012	89.300	88.477	81.564	94.074
	QWK Gemini	0.377	0.382	0.366	0.333	0.521	0.500
	QWK Llama	0.091	0.162	0.378	0.328	0.212	0.449
	RMSE Gemini	58.323	61.092	58.607	59.900	58.949	54.134
	RMSE Llama	85.821	81.426	69.810	71.532	82.258	68.046
C5	Acurácia Gemini	0.218	0.207	0.219	0.221	0.299	0.255
	Acurácia Llama	0.235	0.196	0.230	0.258	0.212	0.290
	PIA Gemini	93.635	91.575	93.988	93.210	91.728	94.403
	PIA Llama	77.119	78.848	84.856	84.198	76.708	86.091

Tabela 4.1: Comparação Métricas para as Competências por experimento, Gemini e Llama.



# Fase 2 - Engenharia de Características

## Passo 1: Extração de Métricas com NILC-Metrix

- 72 métricas textuais extraídas (estatísticas linguísticas, lexicais e estruturais).
- Exemplo: palavras por sentença, frequência lexical, conectores lógicos.

## Passo 2: Saídas de Modelos LLaMA e Gemini

- Utilizados melhores prompts (Cadeia de Pensamento e Contextual).
- Geraram notas por competência, notas finais e detalhamentos intermediários.

## Passo 3: Otimização com Grid Search

- Algoritmo: XGBoost ajustado com **Quadratic Weighted Kappa (QWK)** como métrica de desempenho.
- Testadas 19.683 combinações de hiperparâmetros.

## Passo 4: Execução de Experimentos

- Sete experimentos realizados, combinando diferentes características no treinamento: NILC-Metrix, LLaMA e Gemini.



**"Liberdade é muito pouco,  
o que eu quero ainda não  
tem nome."**

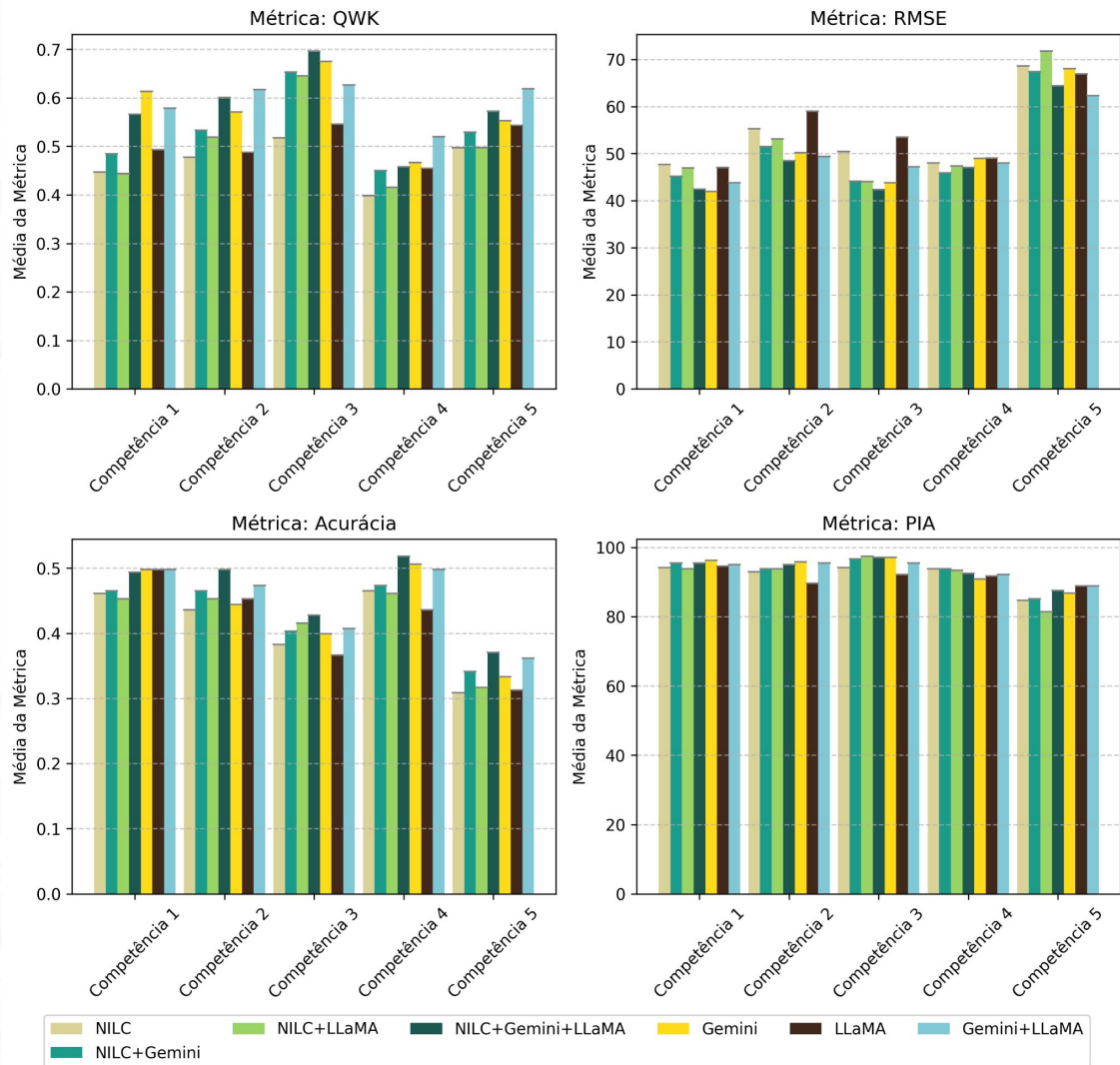
Métrica	Valor	Métrica	Valor
adjective_ratio	0.0	min_freq_brwac	5.135
adverbs	0.33333	min_cw_freq_bra	4.963
content_words	0.75	min_freq_bra	4.963
flesch	88.605	freq_brwac	6.14508
function_words	0.25	freq_bra	6.08708
sentences_per_paragraph	1.0	hypernyms_verbs	0.0
syllables_per_content_word	2.0	brunet	5.20226
words_per_sentence	12.0	honore	1079.18125
noun_ratio	0.16667	personal_pronouns	0.08333
paragraphs	1	ttr	1.0
sentences	1	conn_ratio	0.08333
words	12	add_neg_conn_ratio	0.08333
pronoun_ratio	0.25	add_pos_conn_ratio	0.0
verbs	0.25	cau_neg_conn_ratio	0.0
logic_operators	0.08333	cau_pos_conn_ratio	0.0
and_ratio	0.0	log_neg_conn_ratio	0.08333
if_ratio	0.0	log_pos_conn_ratio	0.0
or_ratio	0.0	tmp_neg_conn_ratio	0.0
negation_ratio	0.08333	tmp_pos_conn_ratio	0.0
cw_freq	1341415.33333	adjectives_ambiguity	0
cw_freq_brwac	6.01556	adverbs_ambiguity	4.75
cw_freq_bra	5.83889	nouns_ambiguity	4.0
min_cw_freq	65261.0	verbs_ambiguity	16.0
min_cw_freq_brwac	5.135	yngve	2.14286

(a) Primeira metade das métricas.

(b) Segunda metade das métricas.

**Tabela 3.4:** Métricas extraídas de amostra, usando NILC-Metrix programa.

# Resultados XGBoost



# XGBoost



Experimento	Dados de Treinamento
1	Apenas métricas do NILC-Metrix
2	Métricas do NILC-Metrix + Resultados Gemini (Prompts Cadeia de Pensamento e Contextual)
3	Métricas do NILC-Metrix + Resultados LLaMA (Prompts Cadeia de Pensamento e Contextual)
4	Métricas do NILC-Metrix + Resultados Gemini e LLaMA (Prompts Cadeia de Pensamento e Contextual)
5	Resultados Gemini (Prompts Cadeia de Pensamento e Contextual)
6	Resultados LLaMA (Prompts Cadeia de Pensamento e Contextual)
7	Resultados Gemini e LLaMA (Prompts Cadeia de Pensamento e Contextual)

Tabela 3.6: Descrição dos dados de treinamento por experimento

Comp	Métrica	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5	Exp. 6	Exp. 7
C1	QWK	0.447	0.485	0.444	0.566	0.613	0.493	0.579
	RMSE	47.730	45.179	46.966	42.475	41.929	47.036	43.848
	Acurácia	0.461	0.465	0.453	0.494	0.498	0.498	0.498
	PIA	94.239	95.473	93.827	95.473	96.296	94.650	95.062
C2	QWK	0.478	0.534	0.519	0.601	0.571	0.488	0.617
	RMSE	55.273	51.512	53.086	48.483	50.218	59.018	49.358
	Acurácia	0.436	0.465	0.453	0.498	0.444	0.453	0.473
	PIA	93.004	93.827	93.827	95.062	95.885	89.712	95.473
C3	QWK	0.517	0.653	0.646	0.696	0.675	0.546	0.626
	RMSE	50.479	44.147	44.073	42.397	43.848	53.518	47.175
	Acurácia	0.383	0.403	0.416	0.428	0.399	0.366	0.407
	PIA	94.239	96.708	97.531	97.119	97.119	92.181	95.473
C4	QWK	0.399	0.450	0.416	0.458	0.466	0.455	0.520
	RMSE	48.005	45.902	47.384	47.036	49.023	49.090	48.005
	Acurácia	0.465	0.473	0.461	0.519	0.506	0.436	0.498
	PIA	93.827	93.827	93.416	92.593	90.947	91.770	92.181
C5	QWK	0.497	0.530	0.497	0.572	0.553	0.543	0.619
	RMSE	68.614	67.403	71.756	64.457	68.035	66.913	62.328
	Acurácia	0.309	0.342	0.317	0.370	0.333	0.313	0.362
	PIA	84.774	85.185	81.481	87.654	86.831	88.889	88.889

Tabela 4.3: Métricas para as Competências por experimento, usando aprendizado de máquina.

# Comparação Resultados Fase 1 e 2



## Competência 1 (C1):

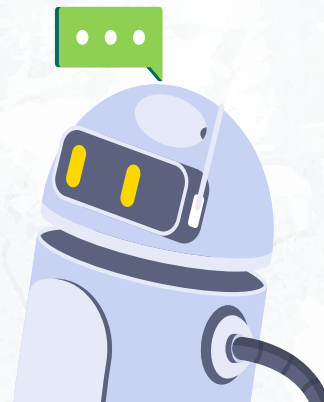
- Melhor QWK:
  - **0.613** (XGBoost com Gemini) - Fase 2
  - **0.520** (Gemini - Prompt Persona) - Fase 1
- Melhor RMSE:
  - **42.745** (XGBoost com Gemini + LLaMA + NILC) - Fase 2
  - **43.801** (Gemini - Prompt Persona) - Fase 1

## Competência 2 (C2):

- Melhor QWK:
  - **0.617** (XGBoost com Gemini + LLaMA + NILC) - Fase 2
  - **0.480** (LLaMA - Prompt Contextual) - Fase 1
- Melhor RMSE:
  - **48.483** (XGBoost com Gemini + LLaMA) - Fase 2
  - **50.914** (Gemini - Prompt Contextual) - Fase 1

## Competência 3 (C3):

- Melhor QWK:
  - **0.675** (XGBoost com Gemini) - Fase 2
  - **0.565** (Gemini - Prompt Cadeia de Pensamento) - Fase 1
- Melhor RMSE:
  - **42.397** (XGBoost com Gemini + NILC) - Fase 2
  - **45.00** (Gemini - Prompt Contextual) - Fase 1



# Comparação Resultados Fase 1 e 2

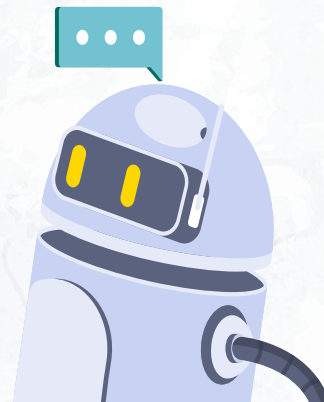


## Competência 4 (C4):

- Melhor QWK:
  - **0.520** (XGBoost com dados Gemini + LLaMA) - Fase 2
  - **0.454** (LLaMA - Prompt Contextual) - Fase 1
- Melhor RMSE:
  - **45.725** (Gemini - Prompt Cadeia de Pensamento) - Fase 1
  - **45.902** (XGBoost com dados Gemini + NILC) - Fase 2

## Competência 5 (C5):

- Melhor QWK:
  - **0.619** (XGBoost com dados Gemini + LLaMA) - Fase 2
  - **0.521** (Gemini - Prompt Cadeia de Pensamento) - Fase 1
- Melhor RMSE:
  - **54.134** (Gemini - Prompt Contextual) - Fase 1
  - **62.328** (XGBoost com dados Gemini + LLaMA) - Fase 2



## Avaliação dos Modelos:

- LLaMA e Gemini foram eficazes na correção, principalmente em cenários com poucos dados.
- Prompts detalhados, como **os contextualizados e em cadeia de pensamento, aumentaram a consistência e eficiência das saídas.**

## Integração com Aprendizado de Máquina:

- Combinar as saídas dos modelos com o XGBoost e as métricas do NILC-Metrix resultou em melhorias significativas.
- **Demonstração do valor de grandes modelos de linguagem como insumos para técnicas supervisionadas.**

## Contribuições:

- **Criação de um dataset estendido com redações nota mil**, enriquecendo a base de dados e corrigindo desequilíbrios de notas.
- Ampliação do entendimento sobre a avaliação automática de textos em português usando Grande Modelos conversacionais.

## Desafios Identificados:

- **Escassez de redações com notas baixas.**
- **Limitações de recursos** no uso das API Privadas e de hardware para teste local.

## Perspectivas Futuras:

- Repetir experimentos para aumentar a robustez das análises.
- Explorar outras técnicas de engenharia de prompts.
- Adicionar a análise de novo modelos e versões, como o LLaMA 3.1 - 405B.



# Bibliografia

- [al. 2023] Sidney Leal et al. **NILC-Metrix: assessing the complexity of written and spoken language in Brazilian Portuguese**. 2023. url: <https://arxiv.org/pdf/2201.03445.pdf>
- [Dubey et al. 2024] Abhimanyu Dubey et al. **“The llama 3 herd of models”**. Em: arXivpreprint arXiv:2407.21783 (2024)
- [Mizumoto e Eguchi 2023] Atsushi Mizumoto e Masaki Eguchi. **“Exploring the potential of using an ai language model for automated essay scoring”**. Em: *Research Methods in Applied Linguistics* 2.2 (2023), pg. 10005
- [Silveira et al. 2024] Igor Cataneo Silveira, André Barbosa e Denis Deratani Mauá. **“A new benchmark for automatic essay scoring in Portuguese”**. Em: *Proceedings of the 16th International Conference on Computational Processing of Portuguese - Vol. 1*. Ed. por Pablo Gamallo et al. Santiago de Compostela, Galicia/Spain: Association for Computational Linguistics, mar. de 2024, pgs. 228–237. url: <https://aclanthology.org/2024.propor-1.23>
- [Team, Georgiev et al. 2024] Gemini Team, Petko Georgiev et al. **“Gemini 1.5: unloc-king multimodal understanding across millions of tokens of context”**. Em: arXivpreprint arXiv:2403.05530 (2024)
- [White et al. 2023] Jules White et al. **“A prompt pattern catalog to enhance prompt engineering with chatgpt”**. Em: arXiv preprint arXiv:2302.11382 (2023) (citado nas pgs. 3, 18).



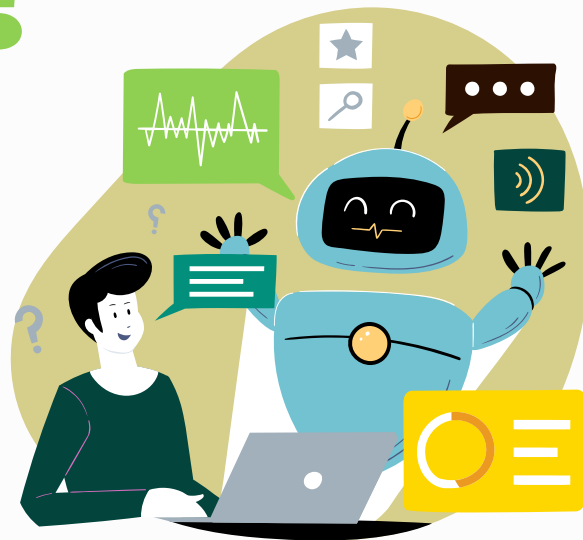
# OBIGADO !

**Você tem alguma dúvida?**

dslcosta2016@gmail.com

11980636381

<https://linux.ime.usp.br/~costa/tcc.html>





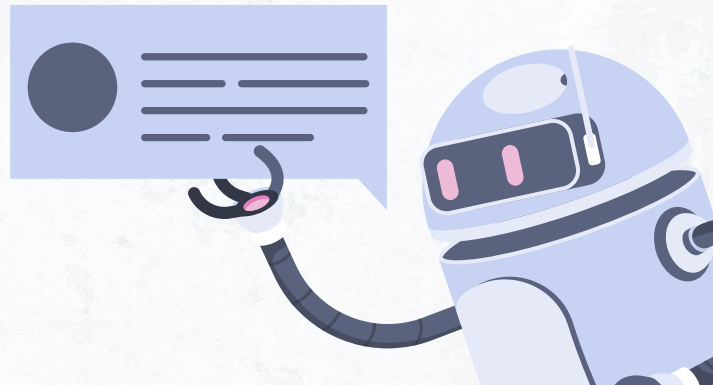
## 2 - Question refinement Pattern

(-) Melhores o seguinte prompt →

Usar o LLM para sugerir perguntas melhores ou mais refinadas que o usuário podeira usar no lugar na pergunta original.

Quando eu fizer uma pergunta sugira uma pergunta melhor e pergunte se eu gostaria de usá-la.

Atue como um especialista em linguagem, como foco em pesquisa na composição texto dissertativos.



### Original:

Corrija a seguinte redação, seguindo o método de avaliação do ENEM, entregue apenas a nota quantitativa entre 0 e 200 para cada competência e a nota final no seguinte formato: [x, y, z, w, t, x+y+z+w+t] Não quero texto na resposta, apenas as notas numéricas para cada competência e o total. Tema:<TEMA> \n Redação: <REDAÇÃO>

### Melhorado:

*Avalie a redação a seguir usando os critérios do ENEM, atribuindo uma nota quantitativa de 0 a 200 para cada competência. Responda apenas com as notas numéricas no formato [x, y, z, w, t, Total], onde "Total" é a soma das notas das cinco competências. Tema: <TEMA> \n Redação: <REDAÇÃO>*

# 3 - Audience Persona Pattern

(-) Explique em um formato que X consiga compreender →

Faz o large language model adaptar a resposta para um certo público ou nível de conhecimento da pessoa.

Assuma que eu sou um jovem no ensino médio, em busca de dicas e feedbacks para melhorar o meu texto.

Assuma que sou um professor universitário de linguística e estou avaliando sua capacidade de corrigir redações dissertativas corretamente





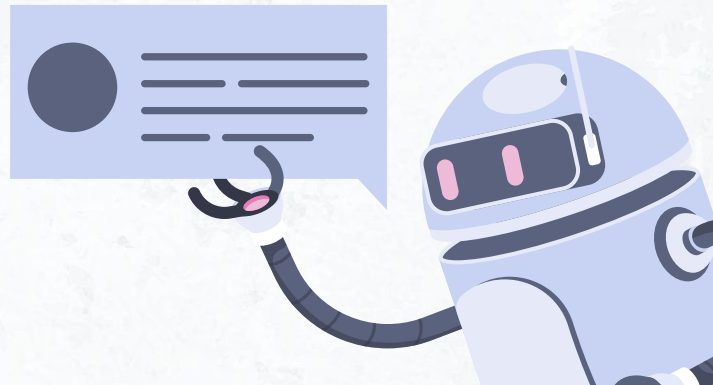
## 4 - Flipped Interaction Pattern

**(-) Me faça perguntas até ter informações para responder** —→

Usa a capacidade de iteração e conversação dos modelos. Fazemos o modelo nos fazer perguntas até ter todas as informações necessárias para fazer uma determinada tarefa.

Me faça perguntas sobre o método avaliativo usado no ENEM. Quando você tiver informações suficientes, avalie o seguinte texto:

Quero que você avalie textos dissertativos. Me faça perguntas sobre o método avaliativo e output esperado, até ter informações suficientes para fazer uma correção correta.



# 5 - Few-shot Pattern

(-) Input: xx, Outuput yyyyyy →

Não contamos ao modelo o que fazer, mas passamos exemplos com entradas e saídas. O modelo aprende um novo padrão e consegue completar a próxima saída.

Redação: xxxxxx  
Prompt: yyyyyy  
Grades: [a, b, c, d, e, total]

Essay: xxxxxx  
Think: Evaluate the grammar and vocabulary  
Action: Give a note 200 because there is no much grammar errors  
Grades: zzzzz



## 6 - ReAct Pattern

(-) Faça X, avalie Y, faça Z, avalie W, faça S



Gera um padrão lógico de pensamento que o modelo deve seguir, é passado um exemplo ou mais, com sequência de passos em que o LLM deve simular.

- 1 - Corrija a seguinte redação: xxxxx
- 2 - Avalie a conclusão, possui uma proposta de intervenção? Está claro quem, como e de que modo essa solução será implementada?
- 3 - Atribua uma nota para essas competências com base nessa avaliação.
- 4 - Avalie agora a coerência entre os parágrafos. O texto possui uma fluidez lógica?
- 5 = Atribua uma nota para esse competência



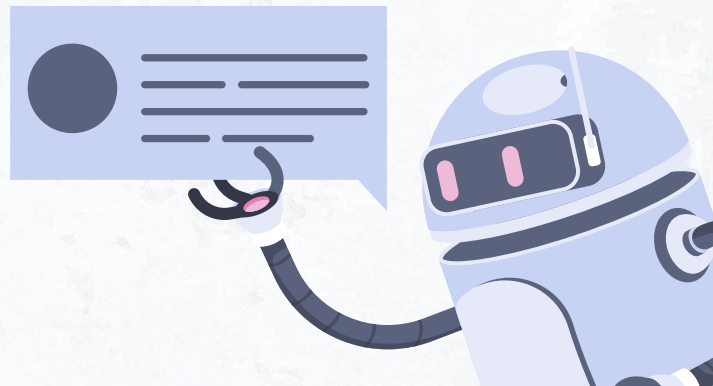
# 7 - Chain of Thought Pattern

**(-) Faça x e explique como você chegou nesse resultado** →

A resposta final, geralmente fica mais correta se pedimos para o modelo explicar uma resposta. Melhorar o raciocínio, quebrando o problema em passos.

Corrija a seguinte redação seguindo o método de correção do Enem. Dê uma explicação para os erros e acertos em cada competência e em seguida atribua uma nota para cada uma delas.

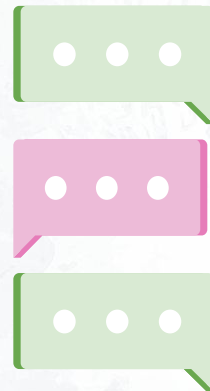
Corrija a seguinte redação, explicando a razão para cada nota que deve ser baseada no desconto de pontos por erros baseados em cada competência.





03

# Chain of thought



(AI)

04



# Técnicas adicionais

(AI)



# Técnicas adicionais

## (1) Adicionar novas informações →

Criar prompts iniciais como conteúdos que o modelo vai precisar usar para cumprir uma tarefa. Temos limitações sobre a quantidade de conteúdo, cabe ao usuário selecionar o que for mais relevante ou utilizar técnicas como sumarização.

## (2) Uso repetido →

Uso como uma conversa permite refinamento iterativo. Pode ser difícil conseguir a resposta correta com apenas um prompt.

## (3) Prompt raízes →

Alguns prompts são pre-configurados antes da sessão inicia, estabelecendo limites e regras para as interações, como as regras sobre temas sensíveis, ou temas envolvendo violência, ou discriminação.

# Prompt engineering tools

## (1) AIBM watsonx.ai Prompt Lab →

Permite o usuário realizar experimentos com prompts de diferentes modelos.

## (2) Spellbook →

Permite editar e testar prompts. Você pode acessar exemplos de prompts também.

## (3) Dust →

Provém uma interface web que permite escrever prompts agrupá-los juntos. Tem integração de API como outros modelos e serviços.

## (4) PromptPerfect →

Pode ser usado para otimizar prompts para diferentes LLMs. Diferentes modelos tem diferentes modelos de otimização.



Otimize esse prompt para Gemini: "Corrija a seguinte redação, seguindo o método de avaliação do ENEM, entregue apenas a nota quantitativa entre 0 e 200 para cada competência e a nota final no seguinte formato: [x, y, z, w, t, z+y+z+w+t] Não quero texto na resposta, apenas as notas numéricas para cada competência e o total. Tema:<TEMA> Redação: <REDAÇÃO>"



Optimizer

Vamos otimizar o prompt para Gemini, levando em consideração a clareza e os delimitadores:

1. Especificação da tarefa: corrigir a redação e calcular a nota de acordo com as competências do ENEM.
2. Clareza na formatação das notas a serem fornecidas.
3. Delimitadores claros para o tema e a redação.

## PROMPT

Corrija a redação abaixo seguindo o método de avaliação do ENEM. Forneça apenas as notas quantitativas entre 0 e 200 para cada competência, e a nota final, no seguinte formato: [C1, C2, C3, C4, C5, Total].

Tema:

----

<tema>

----

Redação:

----

<redação>

----

Não forneça texto na resposta, apenas as notas numéricas para cada competência e o total.

</redação></tema>