

Desafio:

Análise de letras de músicas dos Engenheiros do Hawaii com técnicas de NLP e LLMs

Contexto

Vocês já concluíram a obtenção das letras dos Engenheiros do Hawaii, criando um dataset. Agora, o objetivo é: ver o que a máquina consegue 'perceber' nessas letras, mesmo sem que vocês treinem nenhum modelo complexo.

A ideia é explorar, brincar e refletir sobre como os algoritmos lidam com linguagem, estilo e significado.

Desafio

Escolham pelo menos três experimentos da lista abaixo. A proposta é produzir observações, não resultados 'corretos'. Cada experimento pode ser feito em um notebook, de forma independente.

Opções de experimentos

1. Similaridade entre letras: usar embeddings (sentence-transformers, OpenAI, etc.) para descobrir quais músicas tratam dos mesmos temas.
2. Palavras e temas dominantes: gerar nuvens de palavras, top-10 termos TF-IDF ou contagens simples. O que o vocabulário da banda revela?
3. Geração no estilo Engenheiros: usar dois modelos (ChatGPT, HuggingFace, Ollama...) para criar uma letra no estilo Engenheiros do Hawaii. Compare com as originais.
4. Classificação intuitiva: criar regras simples (ex: contar palavras tristes) ou um classificador para indicar as letras como melancólicas, otimistas ou filosóficas.
5. Chat interpretativo: fazer uma IA para explicar o significado de uma música e compare com a interpretação humana do grupo.

Entrega (até 09/12):

Cada grupo entregará:

a) 1 repositório no git contendo 1 notebook (.ipynb) (ou similar) por experimento.

b) 1 arquivo README.md explicando:

b.1) quais experimentos foram feitos,

b.2) prints ou gráficos obtidos,

b.3) quais resultados foram obtidos de forma quantitativa e qualitativa? Neste caso, apresentar as métricas utilizadas e seus resultados.

Avaliação

Critério	Descrição	Peso
Exploração e curiosidade	Fez perguntas relevantes, tentou entender o comportamento do modelo	30%
Clareza dos resultados	Visualizações e exemplos compreensíveis	30%
Reflexão final	Interpretação do grupo sobre o que a máquina “entendeu”	30%
Organização mínima do repositório	Estrutura clara, código limpo e documentado	10%