




# Comparação das Métricas dos 3 modelos

## TechChallenge Grupo 18\_ Fase 3

### Resumo Comparativo dos Modelos Llama

Abaixo está o resumo comparativo dos três modelos treinados com base nos gráficos gerados no Weights & Biases. Foram analisadas métricas como *Loss de Treinamento e Validação*, *Runtime*, *Steps per Second* e *Samples per Second*.

Métrica	1º Modelo (13/12) 	2º Modelo (15/12) 	v4 (17/12) 
Train Loss	3.05 - 3.1	0.616 - 0.619	1.5 - 1.6
Eval Loss	3.05 - 3.1	0.616 - 0.619	1.5 - 1.6
Steps per Second	27.2 - 27.5	7.75 - 7.78	13.3 - 14.0
Runtime (s)	217 - 220	256.8 - 258	230 - 235
Samples per Second	108.5 - 110	31.05 - 31.15	50 - 52

### Análise Comparativa

Aspecto	Melhor Modelo	Explicação
Train/Eval Loss	2º Modelo (15/12)	Apresentou menor <i>loss</i> , indicando melhor adaptação ao conjunto de dados.
Steps per Second	1º Modelo (13/12)	Treinou com maior velocidade, sugerindo uma execução eficiente.
Runtime	1º Modelo (13/12)	Menor tempo de execução total comparado aos outros modelos.
Samples per Second	1º Modelo (13/12)	Maior taxa de amostras processadas por segundo.

### Observações

- 1º Modelo (13/12): Rápido em termos de execução, porém com *loss* mais alto.
- 2º Modelo (15/12): Melhor desempenho em termos de *loss*, indicando uma melhor qualidade de ajuste.
- v4 (17/12): Compromisso entre os dois, com desempenho intermediário em *loss* e velocidade.

Recomendação: Para melhor equilíbrio entre velocidade e qualidade de ajuste, o 2º Modelo (15/12) parece ser o mais adequado para respostas mais precisas.

Com esse rating podemos usar o modelo v2 para o Streamlit e seguir avançando nas melhorias.