Llama 3.1:8b - métricas por conjunto de teste

