

O Azure OpenAI playground fornecido pela Microsoft trata-se de um aplicativo web onde desenvolvedores podem utilizar as ferramentas de chat no-code Azure OpenAI para testar parâmetros em suas implementações, testar modelos e desenvolver prompts.

Ao lidarmos com o playground devemos ter em mente sobre o gasto com tokens, que se tratam de representações, de certa forma, numéricas, das palavras que serão utilizadas para gerar os resultados esperados. É quase uma regra que, palavras em português utilizam mais tokens que no inglês. Tratam-se de parcelas de uma palavra que são transformadas para uma linguagem compreensível pela Inteligência Artificial. Mas não trata-se apenas de mudar a linguagem, é necessário ainda otimizar seu prompt de forma a ser mais objetivo possível na entrada para obter uma saída com a resposta desejada.

Um dos parâmetros a ser avaliado no âmbito da inteligência regenerativa é a temperatura, que varia de 0 à 1. O “0” da temperatura se traduz em resultados previsíveis por parte da IA, e por consequência “1” seriam os resultados não esperados, ou seja, trata-se de um índice de “Criatividade”, onde os resultados passam a ser mais aleatórios.

Já o parâmetro Top-p restringe as palavras utilizadas na resposta, seria como um limitador de criatividade.

O Frequency Penalty e o Presence Penalty tratam-se de parâmetros que controlam e ajustam a qualidade e a variedade da resposta.