E continuando aqui, agora a gente vai falar um pouquinho sobre a parte da geração de tokens ou tokenização. Acho que é assim que se pronuncia, não ajuda muito, mas o que a gente precisa entender desse cenário? Quando a gente fala da geração de tokens, primeiro, eu estou falando na etapa do meu modelo onde eu estou particionando, aqui a gente usa a palavra decompor também, mas eu gosto de dizer particionando. Pelo menos para mim faz um pouco mais sentido porque eu tô trazendo aqui meio que blocos, para conseguir definir qual que é o meu valor. É tipo como se eu desse o valor para cada ... valor de importância, digamos assim, para cada palavra que existisse. Ah mas existem muitas. Pois então, por isso que a gente precisa ter cada vez mais a questão de uma base de dados e também o teste. Qual é a palavra que vem depois de tal palavra mais vezes?

Por exemplo, eu te amo. Então, assim, eu amo você. Então, assim, ah, mas isso é uma coisa, vai parecer só dia de namorado? Beleza, mas qual que é a estratégia a partir de uma base de dados gigante, a minha solução de inteligência artificial consegue entender isso. A gente entende de forma automática, mas a inteligência é artificial por um motivo, então, eu preciso dizer isso para ela.

Então, nessa primeira etapa da totalização, a gente vai ter o nosso modelo sendo treinado e a primeira coisa é decompor o texto ou participar.

Então, aqui, "Eu ouvi um cachorro latir alto para um gato". Beleza? Eu já sei pela análise de sentimentos aqui rapidamente que o cachorro está enlouquecido, ensandecido, querendo ir atrás desse gato, ok? Mas a minha inteligência, que é artificial, vamos de novo, não está nem aí para quem é cachorro quem é o gato quem vai bater em quem, mas ele quer entender. Nossa inteligência artificial precisa entender. E aí? O que a gente faz com essa informação? Então, vamos lá. Eu, ficou aqui taxado como "1", "2" é cachorro, vamos aqui de 1 a 8. Ou seja, conseguimos atribuir aqui um valor para cada

palavra, certo? O que eu faço com essa informação agora? Viu o ponto importante? Se a gente for trabalhar que cada uma dessas palavras recebeu um número. Então, aqui essa nossa frase fica representada de 1 a 8, ok? Beleza.

Agora o que onde que entra né a história da tokenização. Eu tô louca para errar o uso dessa palavra. Eu tô vendo tudo, que até o final da aula eu erro rss Mas, quando a gente fala isso ... deixa eu até voltar aqui ó. "Eu ouvi um cachorro latir alto para um gato". De 1 a 8. Beleza. Então, agora a gente tem aqui a nossa sequência de 1 a 8. Então aqui "um" está tokenizado como "3", apenas uma vez. Então, vamos voltar lá. "Um", aqui a palavra um está em terceiro, ok? Beleza? Da mesma forma, a frase: "Eu ouvi um gato", poderia ser representada com

as fichas 1, 2, 3 e 8. Vou até voltar aqui. "Eu ouvi um" troca "gato" e vai lá para o final.

Então, percebam que, a partir do momento que eu tenho o meu texto, aqui é só uma frase, mas digamos que nós temos aí uma dissertação de TCC, quem nunca, meu, acho que tinha...

Tinha pouquinho, mas tinha 120 poucas páginas. Um dos menores da turma acho que era o meu. Mas mesmo assim, deu trabalho para escrever. Não tinha chat GPT para me ajudar naquela época. Bom seria se fosse hoje, ah outros tempos.

Então, aqui a gente consegue fazer representação de uma frase literalmente que está dentro da outra né. Eu estou usando aqui informações "Euu ouvi um cachorro latir alto para o gato". Beleza. Ai que frase mais estranha. Aqui a ideia é testar e é isso que nos importa. Então eu peguei todo esse primeiro bloco "Eu ouvi um". Agora não é o mesmo cachorro, agora eu vou utilizar o "gato". Porque também faz sentido, também não vai deixar a frase estranha, não mais

que já está. Então, eu consigo aproveitar os valores que eu já tenho. Isso é a dita cuja da tokenização. Então, aqui a gente vai conseguir trabalhar... a nossa frase, digamos assim. Imagina que isso pode ser um texto também, num modelo de conjunto, podemos colocar isso também. Então imagine que você tem muitos textos e você vai treinando, vai separando, vai dando um valor para cada uma daquelas palavras e você vai testando qual é a próxima.

Então, aqui a gente já tem uma ideia de como que ele consegue fazer isso.

E agora vamos falar um pouquinho sobre como isso se encaixa nas inserções ou incorporações. Talvez você veja também em documentações falando como incorporações.