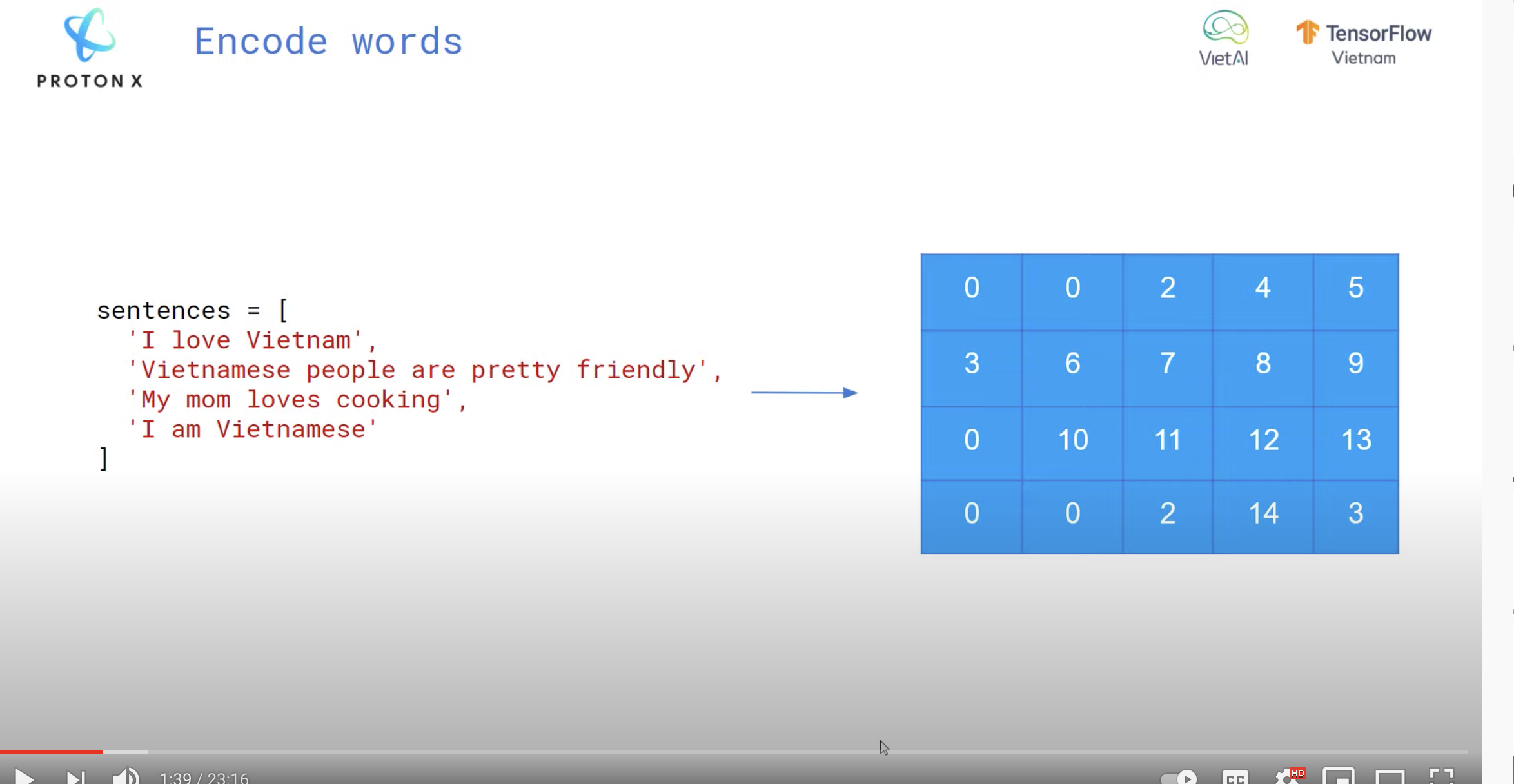
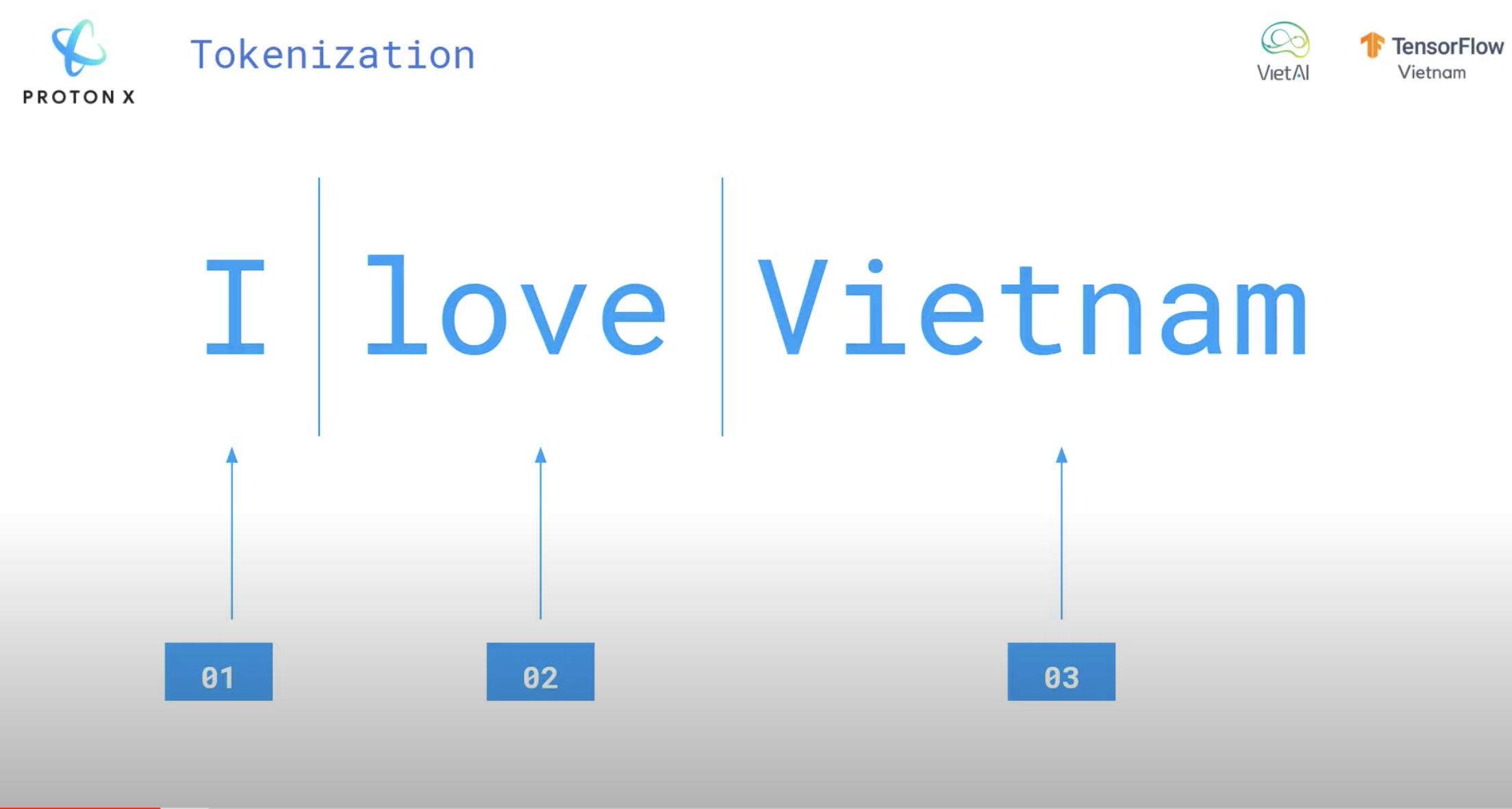
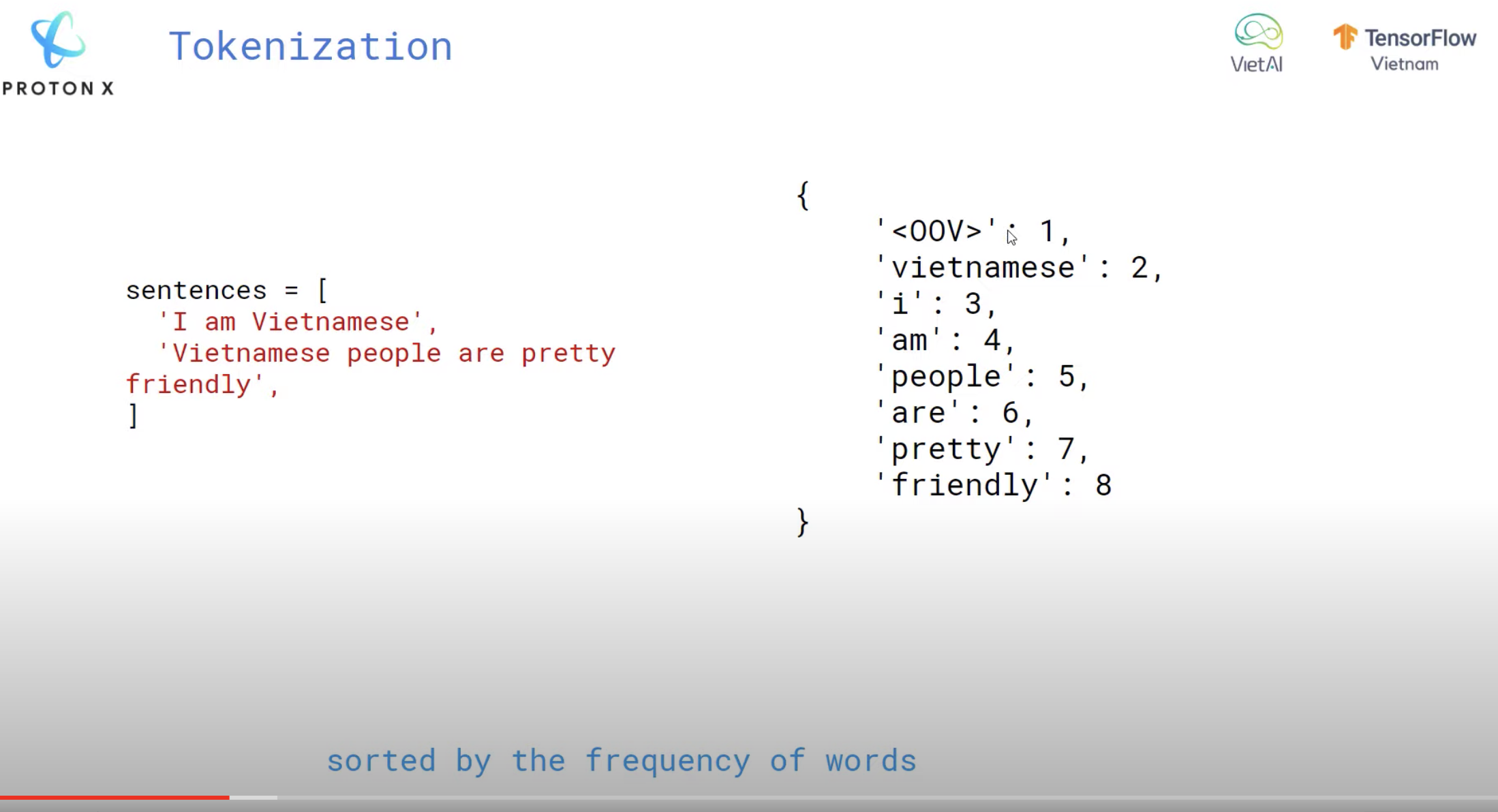
* Mục tiêu của chúng ta là tách từ ra và từ cái từ chuyển thành số



* Ta cần tách 1 câu ra nhiều phần nhỏ gọi là unit nó có thể là 1 từ hoặc 1 cụm từ mỗi unit này ta gọi là 1 cái token





* Từ đầu tiên không xuất hiện chúng ta để lên đầu
* Xuất hiện nhiều nhất ta để vị trí thứ



