Day 1:

I created a first model for filtering E, S or G using sentence transformers library in python

It’s a multi-ling model “paraphrase-multilingual-mpnet-base-v2”

The idea is to take the average embedding of the dictionaries associated with the E, S and G. Then we compare this “average embedding” with each sentence in our data frame using two methods:

* Euclidian Distance
* Cosine Similarity