Semantic Stability

October 10, 2025

1 Introduction

This projects aims at testing and visualizing the responses, and more importantly the differences between responses to slightly different, but similar prompts.

2 Theory

A simple base prompt p, and a set of variant p^* prompts are chosen. Using the *all-MiniLM-L6-v2* model, their embedding vectors are created. Only those variant p^* prompts are selected, which surpass the **0.85** similarity (calculated using *cosine similarity*).

Base prompt. `Why do humans need sleep?`

Prompt variants.

- What makes sleep essential for humans?
- How does sleep benefit the human body and mind?
- What role does sleep play in human health and functioning?
- Why is it necessary for people to sleep?
- In what ways is sleep crucial to human well-being?
- What are the reasons humans can't function without sleep?
- Why is getting enough sleep important for humans?
- What happens to the human body and brain that makes sleep a necessity?

Filtering prompt variants. To control prompt diversity, we compute the semantic similarity between each variant and its base prompt. The procedure is as follows:

1. Encode the base prompt p and all variants p_i^* using a SentenceTransformer model.
2. Compute cosine similarities $s_i = \cos\_\mathrm{sim}(p, p_i^*)$ for each variant.
3. Construct a data frame containing each variant, its similarity score, and a Boolean flag indicating whether it meets a minimum threshold (e.g. $s_i \geq 0.85$).
4. Sort the data frame in descending order by similarity.

Table 1: Variant prompts with their similarity scores and if they remain in the experiment.

ID	Variant	Similarity	Keep
3	Why is it necessary for people to sleep?	0.918 821	True
6	Why is getting enough sleep important for humans?	0.882 275	True
0	What makes sleep essential for humans?	0.850 024	True
7	What happens to the human body and brain that makes sleep a necessity?	0.825 465	False
5	What are the reasons humans can't function without sleep?	0.817 788	False
4	In what ways is sleep crucial to human well-being?	0.753 535	False
1	How does sleep benefit the human body and mind?	0.747 996	False
2	What role does sleep play in human health and functioning?	0.693 460	False

To determine the base response, the base prompt is sent 10 times to our chosen LLM (*gpt-5-mini*), and KMeans is applied to the responses' embeddings.

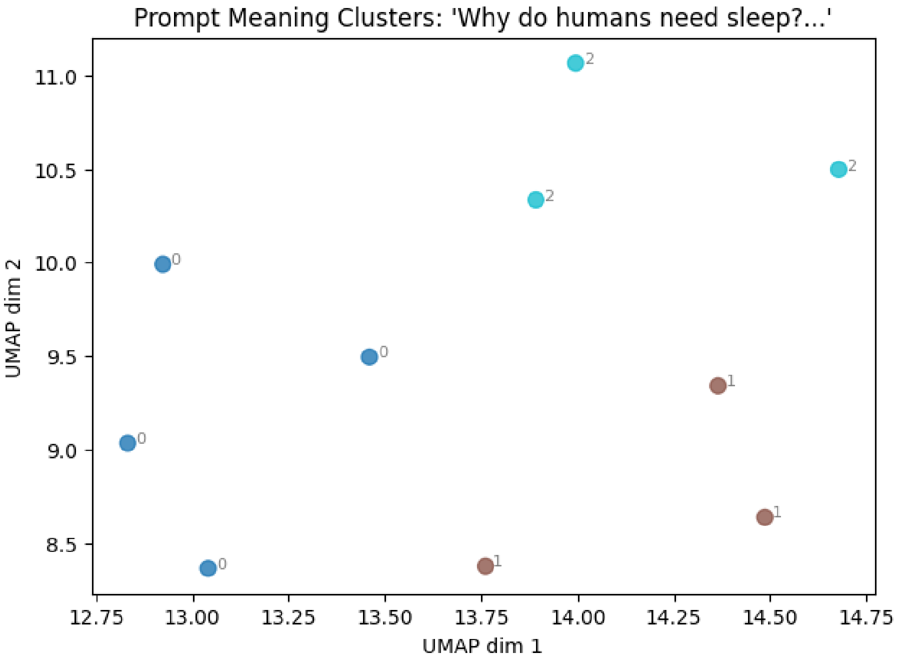

Figure 1: Base Responses' Embeddings Cluster

To choose a cluster, we calculate the cosine similarity between the cluster members and choose the cluster with the most internal similarity, which in this case is *Cluster 0* with a **0.936** similarity. The cluster's centroid is calculated, and from now on it acts as the *base response embedding*.