**SBERT Clarification ranking Evaluation**

**Evaluation**

To evaluate the performance of our fined-tuned **SBERT model** the ranking of clarification question. **A custom dataset of** **75 ambiguous queries.** is created with GPT. These are **diverse set of queries** spans a wide varieties of topics like technology, environment, health, education, finance psychology and more to test the models semantic flexibility and its ability to generalise beyond its familiar domain. Each query was paired with their four candidate clarifying questions. Out of which one was the most contextually appropriate. The model was tested to see if I it could come up with the most contextually appropriate follow up and that it understands the intent of the query or not.

The model used is a fine-tuned version of the **all-MiniLM-L6-v2**, it is a compact SBERT model producing 384-dimensional sentence embeddings which was pretrained for ranking clarifying questions by understanding the intent. During fine-tuning the model was trained on **196 positive Query-clarification** pairs from the **Qulac** dataset using the **MultipleNegativesRankingLoss**. This setup enabled the model to rank multiple clarifications for a single query, which makes it ideal for retrieval and dialog system like RASA.

**Aim**

The aim of the evaluation is measuring model’s semantic understanding, confidence behaviour, and it’s overall suitability for a real-time integration into a retrieval-based dialog system. The performance was measured primarily through cosine similarity scores and confidence thresholds to determine how well the model aligned clarifications to the query intent.

Basically, the evaluation assesses whether the model can rank the best clarification for each query and also **understand user intent**, especially in ambiguous scenarios. A **confidence threshold of 0.40** is defined on cosine similarity to judge the reliability of the model’s top choice.

**Reference**

1. **Cosine Similarity Utility – SBERT:** [**https://www.sbert.net/docs/package\_reference/util.html**](https://www.sbert.net/docs/package_reference/util.html)
2. **MultipleNegativesRankingLoss – SBERT loss function documentation.**

[**https://www.sbert.net/docs/package\_reference/losses.html#multiplenegativesrankingloss**](https://www.sbert.net/docs/package_reference/losses.html#multiplenegativesrankingloss)

1. **all-MiniLM-L6-v2 pretrained model. Hugging Face Model Card.** [**https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2**](https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2)
2. **Hugging Face – SentenceTransformers Documentation.** [**https://www.sbert.net/**](https://www.sbert.net/)