Problem Statement :

One of the interesting problems in NLP for healthcare is in handling medical terms and their association with coding dictionaries. We are going to mimic the problem with hangman/dump charades like NLP word games. Your NLP model should predict the correct word from an intentionally obscured word based on its description (Hint). Our evaluation set will have incomplete words and descriptions**.**

Example

Input masked word = DEM_G_A_HY and

Description = is the statistical study of populations, especially human beings.

Model prediction/output = DEMOGRAPHY

Details:

1. Write a simple deep learning model in python to identify the correct word from the obscured input and description
2. Can use any public datasets or synthetic dataset for training the model
3. Code preferably shared in Github. Colab is also fine. Must contain instructions to run code.
4. Dockerization of application
5. Please share the predictions of your model for this evaluation dataset - https://drive.google.com/file/d/16oFxJoRJpmtjyTxcbvLZBpFToVdKfIx2/view?usp=sharing

Inference pipeline - REST API is preferred.

Additional points

1. Add streamlit/gradio interface
2. Training is dockerized and repeatable for other datasets

Dont's:

Pretrained models from Libraries like Hugginface transformers, fastai, tfhub models are not allowed

Evaluation will be based on :

1. Experiment Design
2. Approach
3. Code quality
4. Accuracy on the evaluation set