Task 3

Scenarios:

1. **If the entire network should be frozen.**
   If the entire network is supposed to be frozen then the model will be new to specific tasks and might not be accurate enough to answer precisely. One of the advantages of this can be training cost because we can load the pre-trained model and freeze the network so there won't be any training.

2. **If only the transformer backbone should be frozen.**
   If the transformer backbone is frozen then we can always load the model locally and train the last required layers for any specific task heads. In this case, we are masking our embeddings of example sentences to the pre-trained model. Core advantage of this is that the model can be specialized for any domain level task.

3. **If only one of the task-specific heads (either for Task A or Task B) should be frozen.**
   If it's trained on any specific task out of two then it will be trained for one and remain untrained in another. Merits of this scenario can be preservation of performance on core tasks and can experiment on another.

Consider a scenario where transfer learning can be beneficial.

Choice of a model will depend upon two factors, first is how knowledgeable the model is i.e., on what amount of data it has been trained and second will be how large is the model. I would consider both the factors as a trade-off because if the model is too big then loading models locally will be time consuming as well as costly.

Freezing will be considered at task level. If the task is very generic and the model is big enough to handle it without being trained additionally then we can freeze it but if it responds poorly on specific tasks then it should be open to be trained with a couple of more layers on top of pre-trained embeddings.