Assignment 5

Srihari Inukurthi - 110712807

December 12, 2024

Question 1. (d) The Model accuracy on the dev set: 13.0 out of 500.0: 2.6%
Proof:

```
Correct: 13.0 out of 500.0: 2.6%
```

The word "London" accuracy: 25.0 out of 500.0: 5.0%
Proof:

```
Correct: 25.0 out of 500.0: 5.0%
```

Question 1. (f) The model accuracy on the dev set: 136.0 out of 500.0: 27.200000000000003%
Proof:

```
Correct: 136.0 out of 500.0: 27.200000000000003%
```

Question 1. (g)(ii)
The attention layers play a major role in the perceiver approach. Let us say the dimensionality of the input is n, then the complexity of the operation QKV is O(n*ℓ). In the cross attention mechanism, the K and V are the projections of the input whereas the Q is a projection of a learned vector reduced by m and m is very less than ℓ. So, now the complexity of the operation is O(n*m). If the same goes for self-attention, then the complexity is O(m^2). It scales quadratically. We know that the multihead attention comlexity is O($n^2 * \ell + \ell^2 * n$). Here, the complexity of the perceiver model is O($n*m + Lm^2$). This will be O($\ell*m + Lm^2$).

Question 2. (a)
The main reason behind that is that the model that is not pre-trained on started its training from scratch. This does not have the features and meaningful representations from training. That is why the accuracy is lower. The pre-trained model contains the generic meaningful representations and the features that it acquired from pre-training. So, the model has higher accuracy compared to the other.

Question 2. (b) One of the reasons is that if it makes up information on its own, it might be incorrect and this results in trust issues for the consumers of the application.
Another reason is that, the information that is made up may impact different industries like for example, stock markets, manipulating democratic elections, etc. This leads to legal troubles. Also, there are a lot of ethical implications too.

Question 2. (c) The reason that seems to be more valid is that it might be learning the associative inference by correlating the patterns it learned during pre-training and finetuning. Not only for this questions, if you train the model on different kind of data, it would be the same way that it tries to predict information that it has not seen before. This leads trust issues, legal and ethical implications, etc.