### Why are you designing the solution in this way?

The task given is to categorize fashion products given an image and corresponding description. Given this information, how would a human expert solve this? Do we need both image and text data?

Firstly, there might be missing image or description for a particular product. Secondly, image might hold more information that just the product itself. For example, a woman wearing a dress and carrying a bag. Thirdly, the text description such as “Cotton fabric, round neck” might be a little too abstract to figure out the product category.

It is therefore clear that the solution should leverage data from both the image as well as the product description. For training the models, I used the [Fashion Product Images Dataset](https://www.kaggle.com/paramaggarwal/fashion-product-images-dataset) from Kaggle and extracted the classes required for this task after some data exploration and cleaning.

### What are the aspects that you considered when designing?

* Only image data or only text data may be a little ambiguous. Models are needed for both image data as well as text data
* The ‘OTHERS’ category can’t really be trained for. One way to figure out if the image/ description doesn’t fit into any of the given classes is if the prediction probability is low for both image and text data

### What are the cases that your solution covers, how are they covered and why are they important?

* Missing text/ image data

There are many cases where either of them might be missing.

* ‘OTHERS’ category

If the prediction probability is low for

### What are the cases your solution doesn’t cover and what are the ways you can extend your current solution for them?

* Modelling a better relationship between the results of the text and image models

Better results might be obtained if a neural network is trained on the results of the text and image models.

* Complex descriptions such as: “sleeveless, cotton, goes well with jeans”

I think I should have concentrated more on getting better data. The data these models were trained on was less complex. The descriptions were obvious and straightforward.

* Poor accuracy for certain classes

The F1 scores on validation data is directly proportional to the number of training examples. More example would’ve definitely helped.