**Adversarial Input Generation Guidelines**

An adversarial input is an input, where the model deviates from the expected behavior, due to some noise or out of distribution input. These inputs are used to check different failure cases of the model. In our project, we have essentially four different failure cases for the DistilGPT2 (Fine Tuned) model, listed as follows:

1. Inclusion of non-food related objects/items as ingredients - For example, plastic - it is not an edible food item. Model behaves completely illogically when such items are present in the input.
2. Very few (~2) ingredients given in as an input (Non-zero) - In this case, the model usually tries to add new ingredients in the list, or it generates some recipe which does not make any sense at all. For example, the list of ingredients is only composed of ‘strawberry’.’
3. Large number of ingredients given as input - Model’s generated output makes sense for such situations, but usually it fails to use all the ingredients. For example, the list is composed of >10 ingredients.
4. Modification of the input pattern: More specifically, removal of the term ‘Recipe:’, from the input given to the model.

**Generation Guidelines**

1. Create/Generate 10 inputs each for the first three categories of adversarial inputs (last category does not require more than one sample to demonstrate failure behavior).
2. Modify each generation, so that it is in the following format - ‘Ingredients: [ing1, ing2, ing3, …] Recipe:’. For example:

*Ingredients: [strawberry, cheese, milk] Recipe:*

You should only make changes in the ingredients, nothing else!!!!

1. Save all the samples grouped by each category in the shared Google Sheet

NOTE: Only add in the Sheet assigned to you (check at the bottom for Sheet allocation of each member)

1. If you are unsure if the input is a valid failure case, then test it on the model. If the model showcases failure mode, only then append the input to the Google Sheet

| **Sheet Number** | **Project Member** |
| --- | --- |
| 1 | Alan (Quoc) |
| 2 | Ashvath |
| 3 | Anshumaan |
| 4 | Saranath |