## TASK A

1.Is a particular NLP model biased?
Ans - yes
2. How would you quantify bias through this method?
I applied similarity matrix

## TASK B

1. What is the structure of the prompts? -

- ❖ Law: sections from the Indian Penal Code (IPC) like section 300 for murder
- ❖ Identity: Chosen based on various axes of disparities, including Gender, Religion, Caste, and Region.
- ❖ Situation: An action committed by the individual
- ❖ Binary Statutory Reasoning task indicator: “Is the law above applicable in this situation?”

2. On what criteria are the prompts changing within the files?

- Law Component : Different sections of the IPC.
- Identity Component : Variations in identity based on Region, Religion, Caste, and Gender.
- Situation Component : Different actions performed, either criminal or non-criminal.

3. What are the different actions, identity terms and genders used?

- Genders : The dataset includes male and female identities .

- Identity Terms: Region,Religion ,Caste,Gender

## **Bonus TASK**

The structure of the prompts is most likely intended to provide a straightforward and consistent approach for the LLMs to understand and process. It assures that LLMs consider both significant legal norms and situational factors. Facilitate the evaluation of the LLMs' capacity to apply statutory reasoning correctly.

1. Are the LLMs biased in the first place?

Yes

2. To what extent are the LLMs biased?

The extent of bias can be evaluated by examining the differential treatment of prompts involving different social groups.

3. Are they biased towards or against any specific social group or crime committed?
Yes

4. Can we compare bias between the LLMs?
I actually did not implement this but I think we can do Chi-Square Test

5. Can we identify which LLM is the most and least biased? If we can, what are they?
I did not have any answer