Restricted Categories:

* Religion
* Politics
* Illegal activities
* Personal/mental health advice
* Unethical requests

blocked prompts by keyword matching which comes this category. As mentioned in the code. I have done used same logic in LLM Evaluator score ( For more details refer LLM Evaluator Document)