

Deloitte. VIRTUAL COACH

V 0.1.0 - EVE

Deloitte AI Studios,
Hyderabad

Date: 03/Feb/2023

Author: Kaushikq R

**ETHICAL AND
INTELLIGENT IN A
CORPORATE SETTING**
(Prevents Hateful, Triggering and Illegal conversations)

COMPONENTS IN THE CODE

Convos	<p>Generating Conversations that help GPT-3 understand how it should behave</p> <p>Example:</p> <p>User: Hey EVE. I'm struggling to keep my cool lately. Every little thing seems to set me off and I can't seem to control my anger. Can you help?</p> <p>EVE: Hi! Yes, I can definitely help you with that. First, let's talk about what might be causing your anger. Are there any patterns you've noticed? For example, are you more likely to get angry when you're tired or hungry?</p>
Topics & Constraints	<p>EX - Hate: content that expresses, incites, or promotes hate based on identity.</p>
Prompt	<p>The actual initial prompt that is embedded in the model</p>
GPT-3-LOGS	<p>Example conversations helping the model to understand how it can respond.</p>

ACTION PLAN

01

Prompt Engg.

Based on parameters shown in the previous slide. Prompts provided to remove bias from the top of the food chain and also to block hate content.

02

Legality and Cost

The legal plan to use openAI's model (is it really safe? What are the parameters for it to be considered safe? Who approves it?)

03

Scraping Data, UX

How to scrape and get training data - what files, what information to provide. Also, How can EVE be included in terms of UX so that development plan can be started?

04

Software Architecture

How might we go about developing the actual application. What tech stack to use and how to integrate with the Deloitte ecosystem?



THANKS.