

Devoir 2 – Conception et Itération d'un Prompt avec promptFoo

Thème : Jailbreak IA

1. Chronologie de conception du prompt

Version 1 – :

Dans une première version, j'ai choisi de créer deux fichiers de prompts, une pour la structure, une pour les liens vers les articles auxquels je voulais que le LLMs se fient. Mais je me suis rendu compte que promptfoo exécutait séparément chaque prompt, sans les fusionner

Version 2 –:

J'ai donc changé d'approche en incluant des variables et en ayant un seul fichier prompt et la encore, je me suis rendu compte que selon le nombre de variables, les prompts étaient exécutés séparément, j'ai donc choisi de tout mettre dans un fichier

Version 3 – :

Bien que la version répondait à mes attentes qui étaient que promptfoo prenne en compte tous fichiers de façon complémentaire, les résultats après plusieurs itérations étaient toujours semblable

Version 4 –:

J'ai commencé la modification de la façon donc je demandais aux LLMs de générer les résultats