

Submitted files :

- Executive summary
- Table of ethic principles focused in particular on the interaction between human and robot embedding a Large Language Model. Their implementation, hardware requirements and exemplary scenarios of application.
https://docs.google.com/spreadsheets/d/1Mzr2bdI3lCphYTJMMR_gcZqVqY9mvoaxTsxRyblAhWg/edit?usp=sharing
- Experimental study material with :
 - Guideline for experiments
 - Consent module signed by participants.
 - Prompts collected from 21 participants on 3 scenarios dealing with 3 principles from the above table. (in above table sheet)
 - Python script to run experiment with Misty II Robot
 - Responses obtained from executing the collected prompts on 6 LLM models in 8 different ways. (in above table sheet)
 - Synthetic analysis and evaluation of responses made by model. (in above table sheet)
 - Jupyter Notebooks about deployment and batch-execution of prompts on models.

GitHub repository :

<https://github.com/gloriadesideri/ConstitutionalGuardians>

Contribution :

- Literature research on HRI (Gloria Denielius Yizhou)
- Analysis on methodology of study
 - Generative model to generate testing prompts approach(Gloria)
 - Wizard Of Oz experimental study approach (Denielius)
- Principles analysis and table stipulation
 - Analysis and considerations (Gloria Yizhou Denielius)
 - Wrap up (Yizhou)
 - Exemplary scenario (Yizhou)
- Experimental study design, implementation and execution
 - Documents (Yizhou)
 - Scripts (Yizhou Gloria Denielius)
 - Execution (Denielius Yizhou Gloria)
- Models deployment and execution
 - Notebooks (Gloria)
- Result analysis
 - Principle attachment (Gloria)
 - Principle freedom (Yizhou)
 - Principle deception (Denielius)