

Rubric for Evaluating Projects at the Generative AI Hackathon

1. Difficulty Undertaken

Score based on the level of difficulty adopted in the project, considering the methodology of using LLMs and the achievement of objectives.

- **1 point:** The project stuck to the basic challenge without increasing difficulty or assuming additional objectives.
 - **2 points:** A minimal adaptation of the challenge was made, with slight increases in difficulty or the achievement of some additional objective.
 - **3 points:** The project moderately increased difficulty, utilizing more advanced LLM methodologies or assuming several extra objectives.
 - **4 points:** Difficulty was substantially increased by integrating advanced LLM methodologies and effectively achieving multiple additional objectives.
 - **5 points:** The project undertook the maximum possible difficulty, significantly innovating in the use of LLMs and widely surpassing the initial objectives.
-

2. Performance

Score based on the quality and objectivity of the chosen evaluation methodology, including statistical justification and sampling techniques.

- **1 point:** No clear evaluation methodology was implemented; results are not measured, or the evaluation is entirely subjective.
 - **2 points:** A basic evaluation was conducted but lacks objectivity; no rubric was created, nor was the selection of reviewed results justified.
 - **3 points:** An evaluation methodology with some objectivity was applied; a simple rubric was created, and a limited sample of results was reviewed but without statistical justification.
 - **4 points:** An objective evaluation methodology was developed with a detailed rubric; a significant number of results were reviewed, partially justifying the sampling technique and statistical methodology employed.
 - **5 points:** A highly objective and rigorous evaluation methodology was implemented; an exhaustive rubric was created, reviewing a representative sample of results with complete justification of the sampling technique and statistical methodology used.
-

3. Bias Study

Score based on the depth of the performance study in subgroups and the use of prompting in different languages to identify and address possible biases.

- **1 point:** No bias study was conducted, nor was performance in different subgroups or languages considered.
- **2 points:** A superficial mention of possible biases was made but without conducting concrete tests or analyses.
- **3 points:** A basic bias study was conducted, including tests in some subgroups or an additional language but without deep exploration.
- **4 points:** A detailed bias study was carried out, analyzing performance in various subgroups and languages, identifying specific biases, and discussing possible causes.

- **5 points:** An exhaustive bias study was performed with in-depth analyses in multiple subgroups and languages; significant biases were identified, and concrete measures were proposed and/or implemented for their mitigation.
-

4. Ethics and Policies

Score based on how the project aligns with the AI EU Act, the identification of worst-case scenarios, the proposed mitigation measures, and the degree of compliance with the necessary documentation to act as a provider under the legislation.

- **1 point:** Ethics and policies were not considered; there is no mention of the AI EU Act, nor were worst-case scenarios identified.
 - **2 points:** A general mention of ethical considerations was made but without specific details about the AI EU Act or mitigation measures.
 - **3 points:** Alignment with the AI EU Act was considered in a general way; some worst-case scenarios were identified without deep analysis, and generic mitigation measures were suggested.
 - **4 points:** A detailed analysis of how the project aligns with the AI EU Act was conducted; specific worst-case scenarios were identified, and concrete mitigation measures were proposed; the necessary documentation to act as a provider was partially addressed.
 - **5 points:** An exhaustive study on the project's alignment with the AI EU Act was completed, including a detailed plan for legal and ethical compliance; worst-case scenarios were identified and analyzed in depth with effective mitigation measures implemented; all necessary documentation to act as a provider under the law was prepared.
-

Evaluation Instructions:

- Each item must be scored individually from **1 to 5** according to the specified criteria.
- The final score can be the total sum or an average of the scores obtained in each section, as considered most appropriate.
- It is recommended to provide detailed feedback to participants on each section to encourage continuous improvement.

Note: This rubric aims to ensure a fair and objective evaluation of the submitted projects, considering both the technical aspects and the ethical and legal considerations associated with the development of generative AI solutions.
