

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | NULLCLASS EDTECH PRIVATE LIMITED | |  | | |
|  |  |  | | |  | |
|  |  | Performance Analysis Report | | |  | |
|  | This report aims to assess the performance of Three different LLms that are :-   1. gemma3 :-   This is the large language model with parameters with 1 billion to 27 billion. Here the model is 12 billion parameters used and needed 32 RAM to run the model.   1. phi4 :-   This is the large language model with parameters with 14 billion. Here the model needed 32 RAM to run the model.   1. llama2 :-   This is the large language model with parameters with 7 billion to 70 billion.Here the model is 13 billion parameters used and needed 16 RAM to run the model. | | | | |  |
|  | * Metrics used to assess the model :-   Here the large language model assess the articles based on following metrics :-  Fluency :- Fluency is the metrics which follows the idea of easy to follow and well writtenness, structure clarity and logical flow of concepts.  Relevancy :- Relevancy is idea of defining that whether the give article or text and related to given topic or concepts.  Creativity :- Creativity is idea of unique perspective, surprising details and originality of the text.  Factual Accuracy :- Factual accuracy deals with correctness of the facts and depth of research and knowledge mentioned in the articles   * Performance assessment:-   Here Performance is assessed by large language model named llama3 and given assessment range of (1-10) given below:-   * gemma3 :-  |  |  |  |  | | --- | --- | --- | --- | | Metrics used | Trial 1 | Trial 2 | Trial 3 | | Fluency | 9 | 9 | 9 | | Relevance | 9 | 8.5 | 8.5 | | Creativity | 7 | 7 | 7 | | Factual Accuracy | 9 | 9.5 | 9.5 |  * phi4 :-  |  |  |  |  | | --- | --- | --- | --- | | Metrics used | Trial 1 | Trial 2 | Trial 3 | | Fluency | 9 | 9 | 9 | | Relevance | 8.5 | 9.5 | 9 | | Creativity | 7.5 | 7.5 | 7 | | Factual Accuracy | 9.5 | 9.5 | 9 |  * llama2 :-  |  |  |  |  | | --- | --- | --- | --- | | Metrics used | Trial 1 | Trial 2 | Trial 3 | | Fluency | 9 | 9 | 9 | | Relevance | 8 | 9 | 8.5 | | Creativity | 7 | 6 | 7.5 | | Factual Accuracy | 9 | 9 | 9.5 |   As per above tables, we can state that Fluency and Factual accuracy of the all the tables are almost same with slight difference in the score assessment. Where as phi4 and gemma3 may show slightly more creativity than llama2. Phi4 may take slight lead in relevance and creativity.   * Conclusion :-   As per the assessment we can conclude that although all three models may show strong fluency and factual consistency, Phi4 may take slight lead in Creativity and Relevance. So here we may say that Phi4 is best model to generate article among them. | | | | |  |
|  |  | |  | | | |
|  |  | | | | |  |