**Institute of Business Administration**

**Introduction to Text Analytics**

**Assignment 01 – LLM Response Assessment**

**Name: Zuha Aqib ID: 26106**

Name of 3B Model used: "Qwen/Qwen2.5-3B-Instruct"

Name of 7B Model used: "open-thoughts/OpenThinker-7B"

Name of 14B Model used: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Evaluate each model’s response using the mentioned criteria (on a scale of 1-5, 1 means completely irrelevant response while 5 means perfect response):

| **Task** | **Criteria** | **3B Model** | **7B Model** | **14B Model** | **Comments (if any)** |
| --- | --- | --- | --- | --- | --- |
| **Summarization** | Conciseness | 5 |  |  |  |
| Clarity | 5 |  |  |  |
| **Question Answering** | Accuracy | 5 |  |  |  |
| Completeness | 5 |  |  |  |
| **Keyword Extraction** | Completeness | 3 |  |  |  |
| Categorization | 2 |  |  |  |
| **Translation** | Fidelity | 5 |  |  |  |
| Fluency | 5 |  |  |  |
| Consistency | 4 |  |  |  |
| **Time Taken (seconds)** | Time taken to do all 4 prompts, and then convert French back to English | 25s 20ms + 10s 73ms = **36s** |  |  |  |

If you were to deploy one of these LLMs at IBA, which model would you choose, and what will be your justification?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Criteria Explanation**

Summarization:

* Conciseness: Is the summary brief yet comprehensive?
* Clarity: Is the summary clear and understandable?

Question Answering:

* Accuracy: Is the answer factually correct, as per the email?
* Completeness: Is the answer complete, providing all necessary details?

Keyword Extraction:

* Completeness: Are all keywords identified?
* Categorization: Are keywords correctly categorized (person, date, product, etc.)?

Translation (English to French to English):

* Fidelity: Does the back-translation convey the original message accurately?
* Fluency: Is the grammar of the back translation correct and fluent?
* Consistency: Is the tone and terminology consistent in the back translation?