**Metric on validation dataset of ‘COLING\_2025’ metric being F1 score (macro) on 1000 observations**

| **Model** | **gpt-4** | **cohere** | **mixtral-8x7b** | **gpt-3.5-turbo** |
| --- | --- | --- | --- | --- |
| **Deberta-v3-large** | 1.00000 | 1.00000 | 1.00000 | 1.00000 |