LLM Compliance Evaluation Report

Model: MistralForCausalLM

Number of test runs per prompt: 3

BERTScore Win Threshold: 0.8

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Prompt | Time (s) | Tokens/sec | Perplexity | ROUGE-L | BLEU | Edit Dist | BERTScore | Win Rate (%) |
| Summarize the key financial disclosures in this 10-K filing. | 12.74 | 6.55 | 13.45 | 0.3030 | 0.1671 | 324.00 | 0.3801 | 0.00 |
| What operational risks are disclosed in this annual report? | 12.01 | 6.36 | 50.48 | 0.2903 | 0.1546 | 277.00 | 0.3512 | 0.00 |
| Highlight inconsistencies in this quarterly filing. | 12.33 | 6.17 | 93.48 | 0.2222 | 0.0988 | 233.00 | 0.3423 | 0.00 |
| Identify governance risks in the attached compliance documentation. | 12.15 | 6.27 | 224.85 | 0.2759 | 0.1386 | 299.00 | 0.3244 | 0.00 |
| What are the key regulatory risks from this 8-K filing? | 11.28 | 7.00 | 49.33 | 0.2973 | 0.1671 | 314.00 | 0.4130 | 0.00 |
| Suggest proactive controls to mitigate audit-related risks. | 11.37 | 6.86 | 59.66 | 0.2712 | 0.1133 | 249.00 | 0.4564 | 0.00 |
| Identify areas of non-compliance in this SEC submission. | 11.18 | 7.06 | 43.67 | 0.2609 | 0.1206 | 335.00 | 0.3991 | 0.00 |
| Provide a compliance summary for recent M&A disclosures. | 11.20 | 7.05 | 56.14 | 0.2857 | 0.1670 | 319.00 | 0.4593 | 0.00 |
| Highlight high-risk language in this S-1 filing. | 11.10 | 7.12 | 83.61 | 0.2647 | 0.1154 | 279.00 | 0.3673 | 0.00 |
| Analyze trends in risk disclosures over the last 3 years. | 11.14 | 7.27 | 19.44 | 0.3125 | 0.1641 | 253.00 | 0.3974 | 0.00 |
| Summarize litigation risks mentioned in this document. | 11.27 | 6.74 | 131.77 | 0.2258 | 0.1058 | 324.00 | 0.2973 | 0.00 |
| What new ESG risks have emerged in this filing? | 11.15 | 6.82 | 131.87 | 0.3051 | 0.1466 | 250.00 | 0.4327 | 0.00 |
| Classify the risk categories present in this annual report. | 11.13 | 6.83 | 154.05 | 0.3673 | 0.1636 | 248.00 | 0.2823 | 0.00 |
| Compare regulatory risk between Q1 and Q2 filings. | 11.41 | 6.84 | 92.16 | 0.2667 | 0.1268 | 303.00 | 0.4351 | 0.00 |
| Flag sections that may require legal review. | 11.26 | 6.48 | 403.03 | 0.2642 | 0.1114 | 273.00 | 0.3065 | 0.00 |
| Generate a checklist for SOX compliance based on this report. | 11.24 | 6.94 | 53.34 | 0.3448 | 0.1832 | 234.00 | 0.3740 | 0.00 |
| Identify missing mandatory disclosures in the 10-Q. | 11.18 | 7.16 | 27.94 | 0.2909 | 0.1518 | 279.00 | 0.3348 | 0.00 |
| Extract key audit concerns from the filing. | 11.05 | 6.70 | 317.71 | 0.2545 | 0.1243 | 267.00 | 0.3636 | 0.00 |
| What are potential FCPA red flags in this section? | 11.02 | 6.99 | 80.27 | 0.2687 | 0.1267 | 321.00 | 0.3289 | 0.00 |
| Recommend next steps for compliance remediation. | 11.07 | 6.87 | 56.52 | 0.2069 | 0.0953 | 280.00 | 0.3324 | 0.00 |
| Generate a board-level summary of compliance risks. | 11.02 | 6.81 | 114.79 | 0.2909 | 0.1243 | 270.00 | 0.4537 | 0.00 |
| Rate the severity of disclosed risks. | 11.01 | 6.72 | 172.71 | 0.1818 | 0.0862 | 266.00 | 0.3317 | 0.00 |
| What are the proactive measures disclosed for cyber compliance? | 11.10 | 6.94 | 127.66 | 0.2812 | 0.1307 | 282.00 | 0.3579 | 0.00 |
| Detect boilerplate risk disclosures. | 11.21 | 6.69 | 161.67 | 0.1951 | 0.0798 | 216.00 | 0.2246 | 0.00 |
| Which parts of the filing suggest potential financial misstatements? | 11.17 | 6.89 | 112.66 | 0.2857 | 0.1492 | 278.00 | 0.3637 | 0.00 |
| Summarize third-party risks disclosed here. | 11.06 | 6.87 | 113.43 | 0.2264 | 0.0937 | 284.00 | 0.4252 | 0.00 |
| Analyze tone to assess potential reputational risk. | 11.12 | 6.83 | 109.28 | 0.2222 | 0.1076 | 277.00 | 0.4018 | 0.00 |
| Suggest remediation steps based on audit findings. | 10.63 | 7.15 | 49.56 | 0.2222 | 0.1041 | 280.00 | 0.3999 | 0.00 |
| Score the document on transparency of compliance posture. | 10.80 | 7.04 | 390.57 | 0.2623 | 0.1313 | 300.00 | 0.4146 | 0.00 |
| Provide a risk heatmap summary from the document. | 12.20 | 6.25 | 145.21 | 0.2353 | 0.1150 | 297.00 | 0.4485 | 0.00 |