#### **Model 1: sentence-transformers/all-MiniLM-L6-v2**

| Question | Score (1–5) | Notes |
| --- | --- | --- |
| Was the answer complete? | 3 | Answered the core question but missed some details. |
| Was the answer correct? | 5 | The information provided was accurate. |
| Was the language clear? | 4 | Generally clear but slightly verbose. |
| Did the context feel on-topic? | 4 | The retrieved chunks were relevant but sometimes too broad. |
| Were the chunks concise and useful? | 3 | Pulled in the entire document as a single chunk, lacking focus. |

#### **Model 2: BAAI/bge-small-en-v1.5**

| Question | Score (1–5) | Notes |
| --- | --- | --- |
| Was the answer complete? | 5 | Provided a comprehensive and detailed answer. |
| Was the answer correct? | 5 | All facts were correct and directly supported by the text. |
| Was the language clear? | 5 | The synthesized answer was direct and easy to understand. |
| Did the context feel on-topic? | 5 | Retrieved highly specific sentences related to the query. |
| Were the chunks concise and useful? | 5 | The context was very focused, providing only the necessary info. |

#### **Model 3: sentence-transformers/all-mpnet-base-v2**

| Question | Score (1–5) | Notes |
| --- | --- | --- |
| Was the answer complete? | 4 | The answer was good but slightly less detailed than BGE. |
| Was the answer correct? | 5 | The information was accurate. |
| Was the language clear? | 4 | Clear and well-written response. |
| Did the context feel on-topic? | 4 | Context was relevant, similar to MiniLM but slightly more focused. |
| Were the chunks concise and useful? | 4 | Better than MiniLM at isolating relevant sections. |