BẢNG SO SÁNH CÁC MODEL

1. Tốc độ xử lí

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Trung bình (ms) | Điểm (10) | Điểm mạnh |
| **Gemini** | 6,000–7,100 | 10 | Rất nhanh, phản hồi ổn định, phù hợp cho real-time. |
| **Devstral-Small** | ~9,500–10,200 | 8 | Tốc độ tốt với model miễn phí, phản hồi chuẩn JSON. |
| **Mistral-7B-Instruct** | ~10,100–11,800 | 7 | Ổn định, không quá chậm, nhưng không tối ưu. |
| **Qwen-Coder-32B** | 16,800–21,200 | 5 | Có khả năng xử lý tốt các câu hỏi logic cao. |
| **Deepseek-R1** | 26,000–49,000 | 2 | Quá chậm, dễ timeout, không phù hợp cho tác vụ real-time. |

1. Độ ổn định

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Thành công / Tổng | Điểm (10) | Điểm mạnh |
| **Gemini** | 5/5 | 10 | Không lỗi parse, JSON chuẩn, API phản hồi đều. |
| **Devstral-Small** | 3/3 | 9 | JSON chuẩn, không gặp lỗi định dạng. |
| **Qwen-Coder-32B** | 3/3 | 9 | Phản hồi tốt với độ dài input lớn, không lỗi hệ thống. |
| **Mistral-7B-Instruct** | 2/3 | 6 | Một lần lỗi JSON format, đôi khi lẫn ngữ cảnh. |
| **Deepseek-R1** | 1/3 | 3 | Thường xuyên lỗi parse JSON hoặc không có output. |

1. Hiệu suất khi input nhiều

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Điểm (10) | Lỗi thường gặp | Điểm mạnh |
| **Gemini** | 9 | Không có lỗi nổi bật khi input dài/ngắn. | Xử lý nhiều độ khó, scale tốt. |
| **Qwen-Coder-32B** | 8 | Tốc độ chậm dần khi input dài nhưng không lỗi. | Vẫn giữ được logic và trả lời đầy đủ. |
| **Devstral-Small** | 7 | Chưa test kỹ với file cực dài, nhưng không timeout. | Cân bằng giữa tốc độ và tính ổn định. |
| **Mistral-7B-Instruct** | 6 | Dễ sinh lỗi JSON, mất context nếu đoạn dài. | Xử lý cơ bản ổn với input vừa. |
| **Deepseek-R1** | 2 | Timeout, lỗi JSON, không phản hồi khi input vừa/dài. | Không phù hợp cho input lớn. |

1. Phủ sóng nội dung

|  |  |  |
| --- | --- | --- |
| Model | Điểm (10) | Điểm mạnh |
| **Gemini** | 10 | Bao phủ đầy đủ: Toán, Lý, Hóa, Python, cấu trúc dữ liệu. |
| **Qwen-Coder-32B** | 9 | Tạo được câu hỏi logic phức tạp, hiểu rõ khái niệm toán/lập trình. |
| **Devstral-Small** | 8 | Bao phủ đủ 3 môn cơ bản và lập trình phổ thông. |
| **Mistral-7B-Instruct** | 7 | Có thể lệch sang nội dung không có trong tài liệu. |
| **Deepseek-R1** | 6 | Câu hỏi đúng nhưng chưa khai thác được nội dung khó hoặc nâng cao. |

1. Độ chính xác

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Điểm (10) | Lỗi thường gặp | Điểm mạnh |
| **Gemini** | 10 | Không có lỗi sai nội dung, trả lời đúng, đầy đủ giải thích. | Chính xác tuyệt đối, kể cả toán và lập trình. |
| **Qwen-Coder-32B** | 9 | Đôi khi giải thích rườm rà, nhưng đáp án đúng. | Phân tích sâu, logic tốt. |
| **Devstral-Small** | 8 | Hiếm gặp lỗi, chỉ hơi đơn giản hóa giải thích. | Trả lời ngắn gọn, đúng trọng tâm. |
| **Mistral-7B-Instruct** | 6 | Một số câu sai ngữ cảnh (vd: cosine, trigonometry), lỗi diễn giải. | Hiểu cú pháp Python và toán cơ bản khá ổn. |
| **Deepseek-R1** | 5 | Dữ liệu đúng nhưng lỗi định dạng khiến không sử dụng được trong hệ thống. | Khi đúng thì câu hỏi có chất lượng. |

KẾT LUẬN

Gemini: Độ chính xác cao, tốc độ nhanh và ổn định tốt.

Qwen-Coder-32B: Phù hợp cho phân tích logic cao và open source mạnh.

Devstral-Small: Cân bằng các yếu tố, phù hợp hệ thống nhẹ hoặc về học thuật.

Mistral-7B: Tệ, nhỏ, ko nên sử dụng cho production.

DeepseekR1: Có thể dùng cho production nhưng ko nên dùng cho hệ thống tạo câu hỏi do quá nhiều lỗi parse và timeout.