

Human

Contains

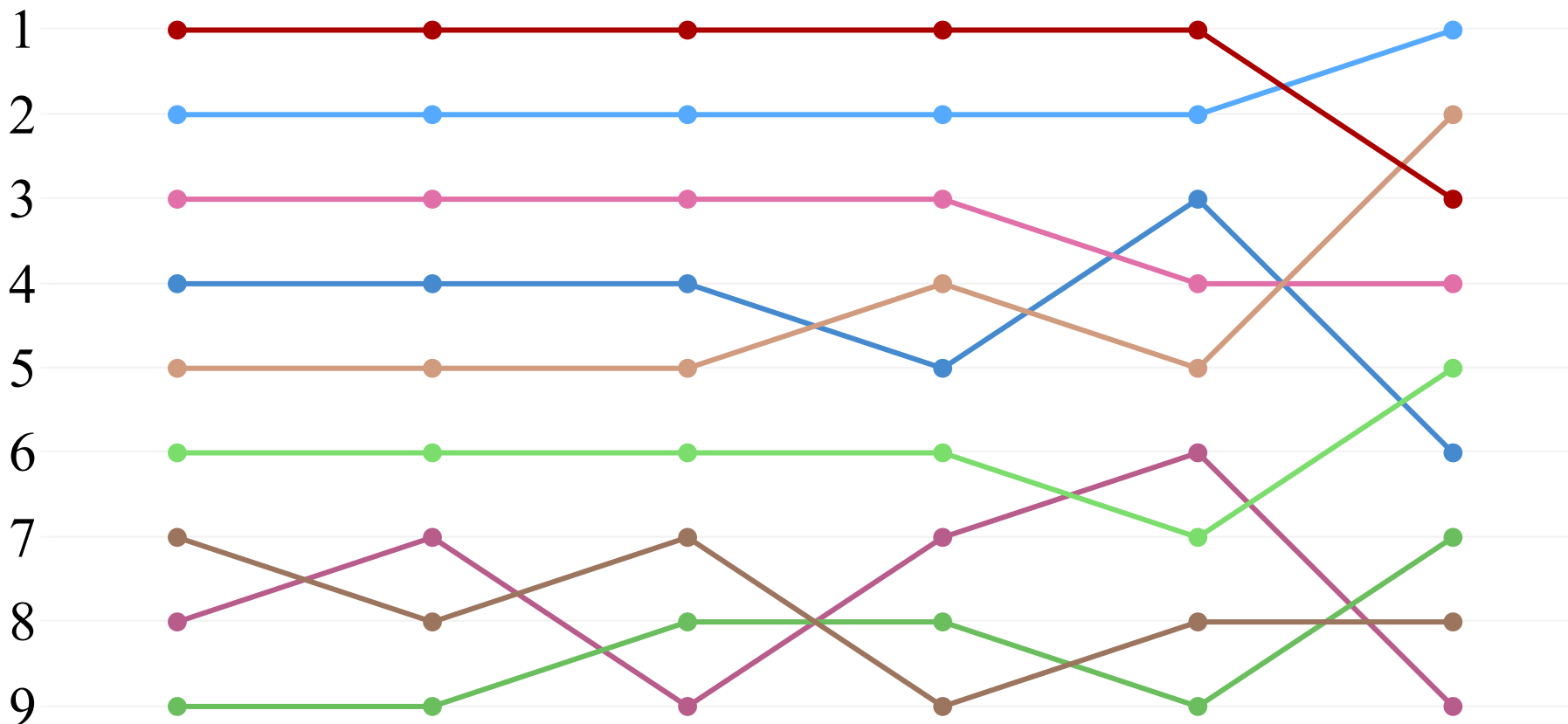
JudgeLM-
7B

GPT-4

Llama3-
70B

EM

Rank



Exam Taker Models

■ GPT 4

■ Llama2 7B Base

■ Llama2 13B FT

■ Mistral 7B

■ Llama2 7B FT

■ Llama2 70B Base

■ Mistral 7B FT

■ Llama2 13B Base

■ Llama2 70B FT