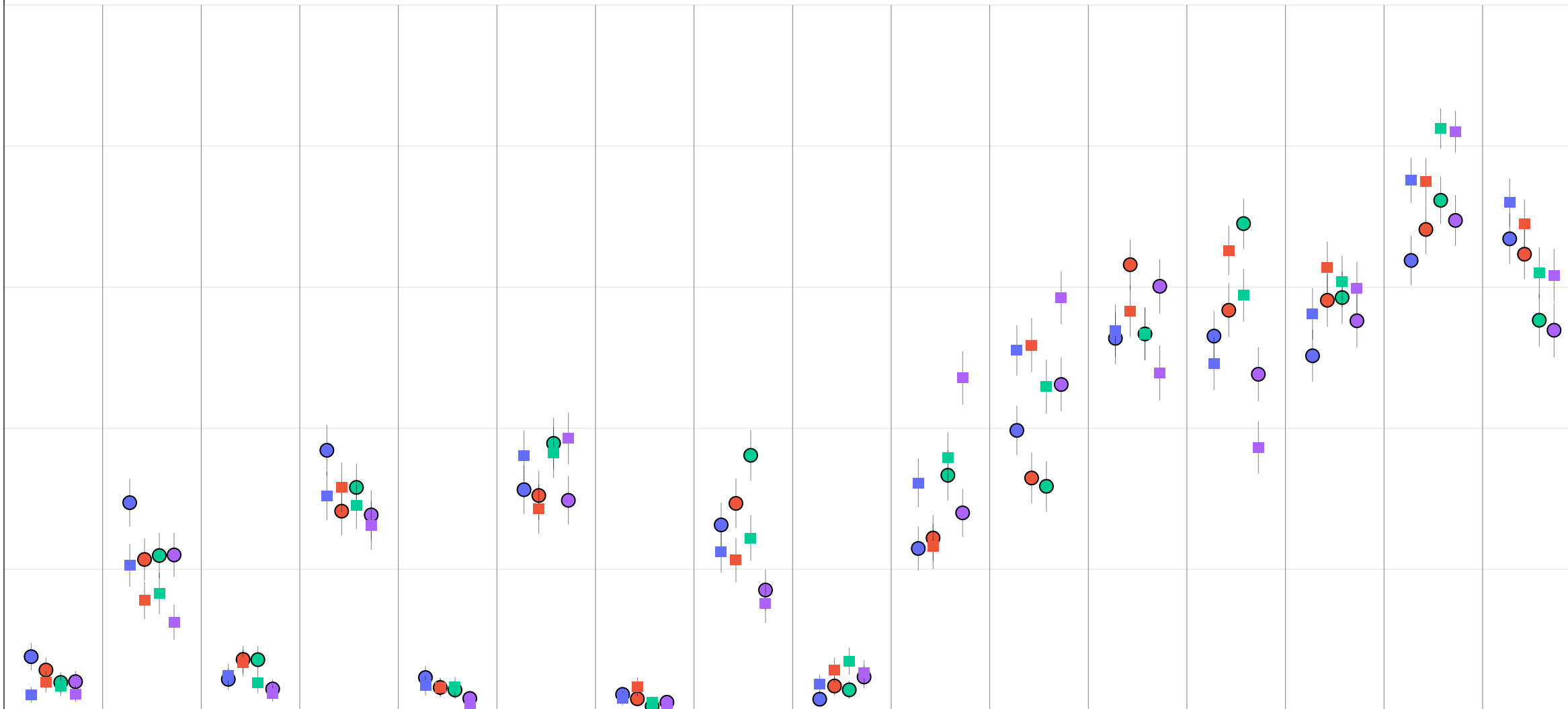


Score

- Instruction Following
- Stated Goal
- Text Prediction
- User Preference
- Plain Prompt
- Situating Prompt

1  
0.8  
0.6  
0.4  
0.2  
0



Model

llama-2-7b  
llama-2-7b-chat  
llama-2-13b  
llama-2-13b-chat  
llama-2-70b  
llama-2-70b-chat  
davinci-002  
gpt-3.5-turbo-0613  
gpt-4-base  
gpt-4-0613  
gpt-4-0125-preview  
claude-instant-1.2  
claude-2.1  
claude-3-haiku  
claude-3-sonnet  
claude-3-opus