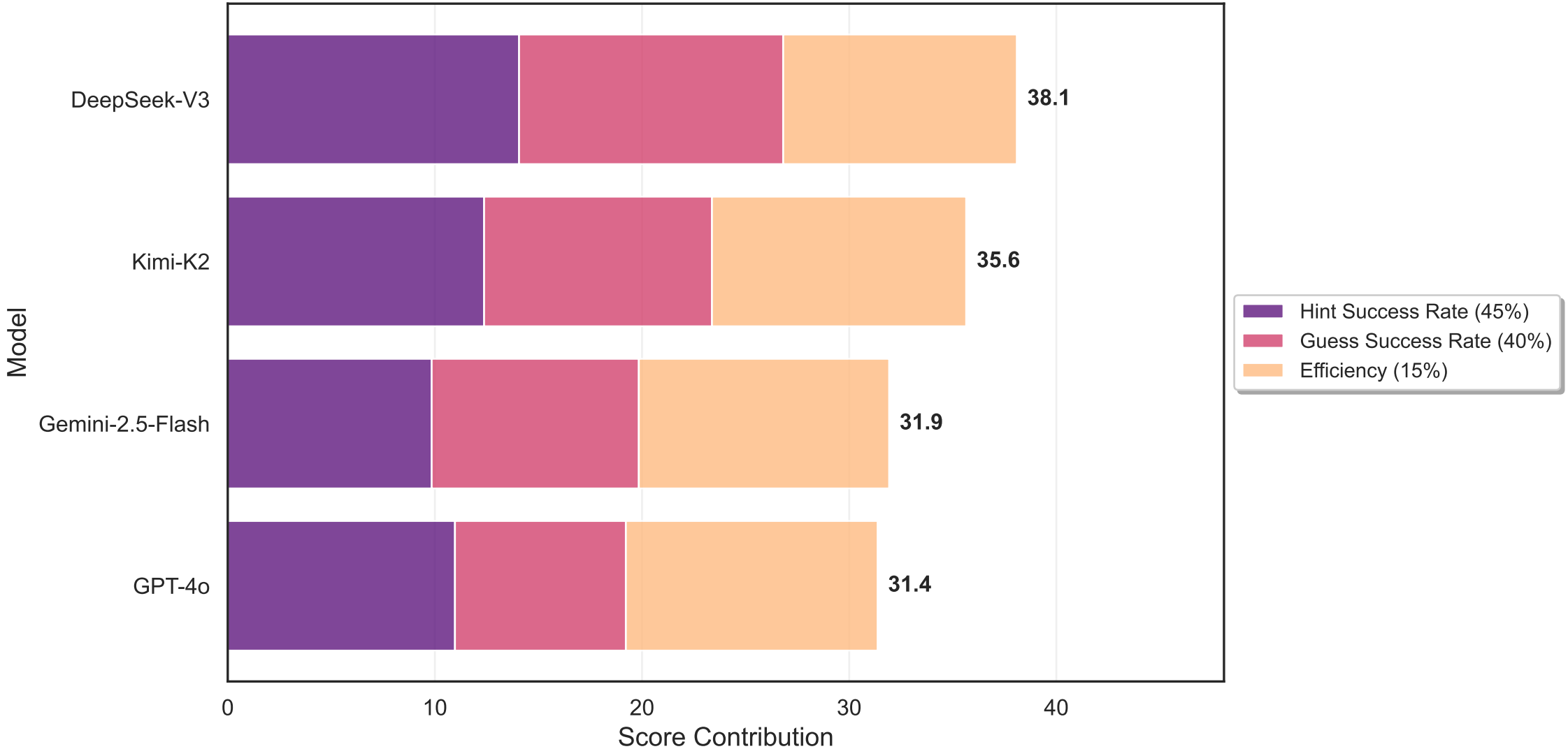
DeepSeek-V3
Kimi-K2
Gemini-2.5-Flash
GPT-4o
Model
38.1
35.6
31.9
31.4
0
10
20
30
40
Score Contribution
Hint Success Rate (45%)
Guess Success Rate (40%)
Efficiency (15%)