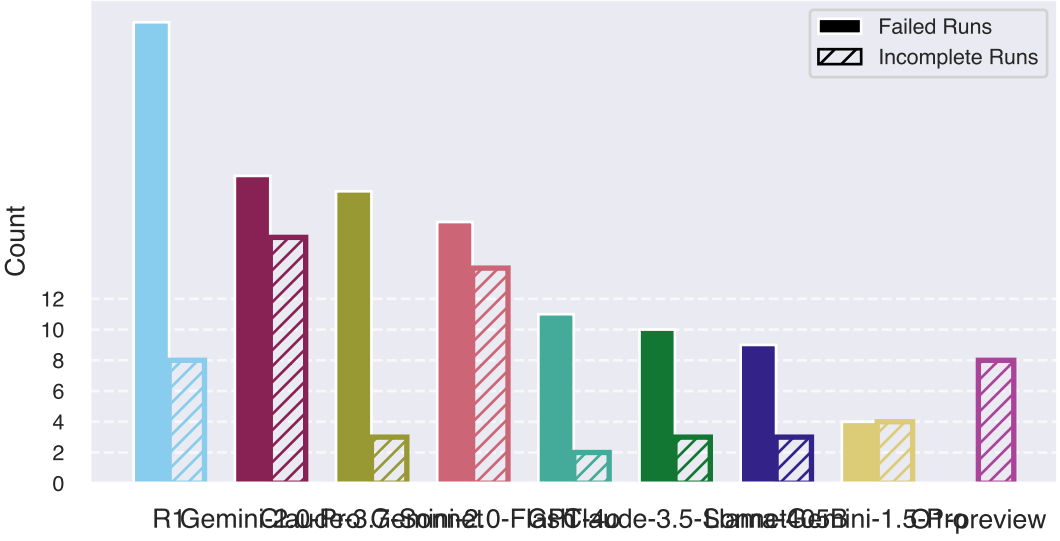
Failed Runs
Incomplete Runs
Count
0
2
4
6
8
10
12
R1
Gemini-2.0-Pro
Claude-3.7-Sonnet
o3-mini-2025-01-31
Gemini-2.0-Flash
GPT-4o
Claude-3.5-Sonnet
Llama-405B
Gemini-1.5-Pro-preview