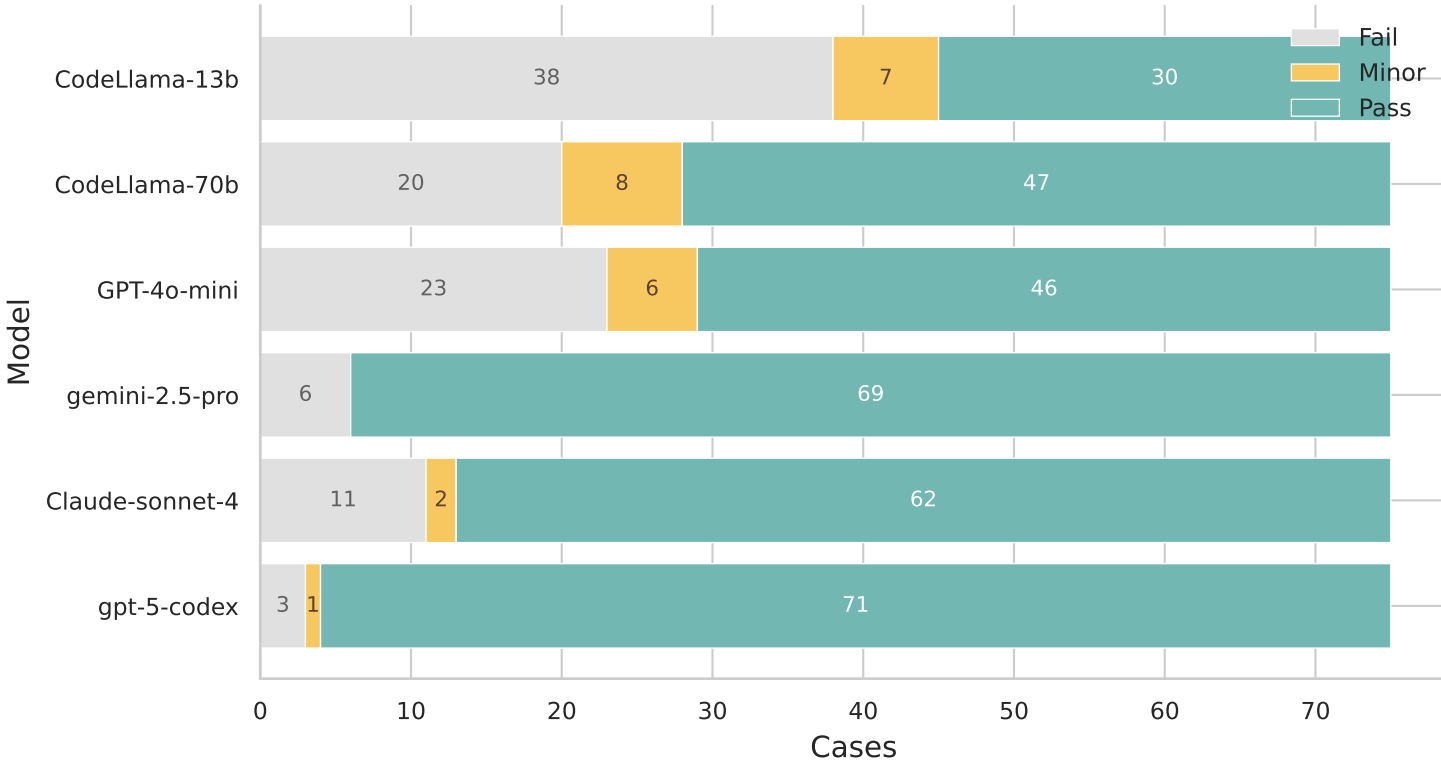
Fail
Minor
Pass
CodeLlama-13b
38
7
30
CodeLlama-70b
20
8
47
GPT-4o-mini
23
6
46
gemini-2.5-pro
6
69
Claude-sonnet-4
11
2
62
gpt-5-codex
3
1
71
Model
0
10
20
30
40
50
60
70
Cases