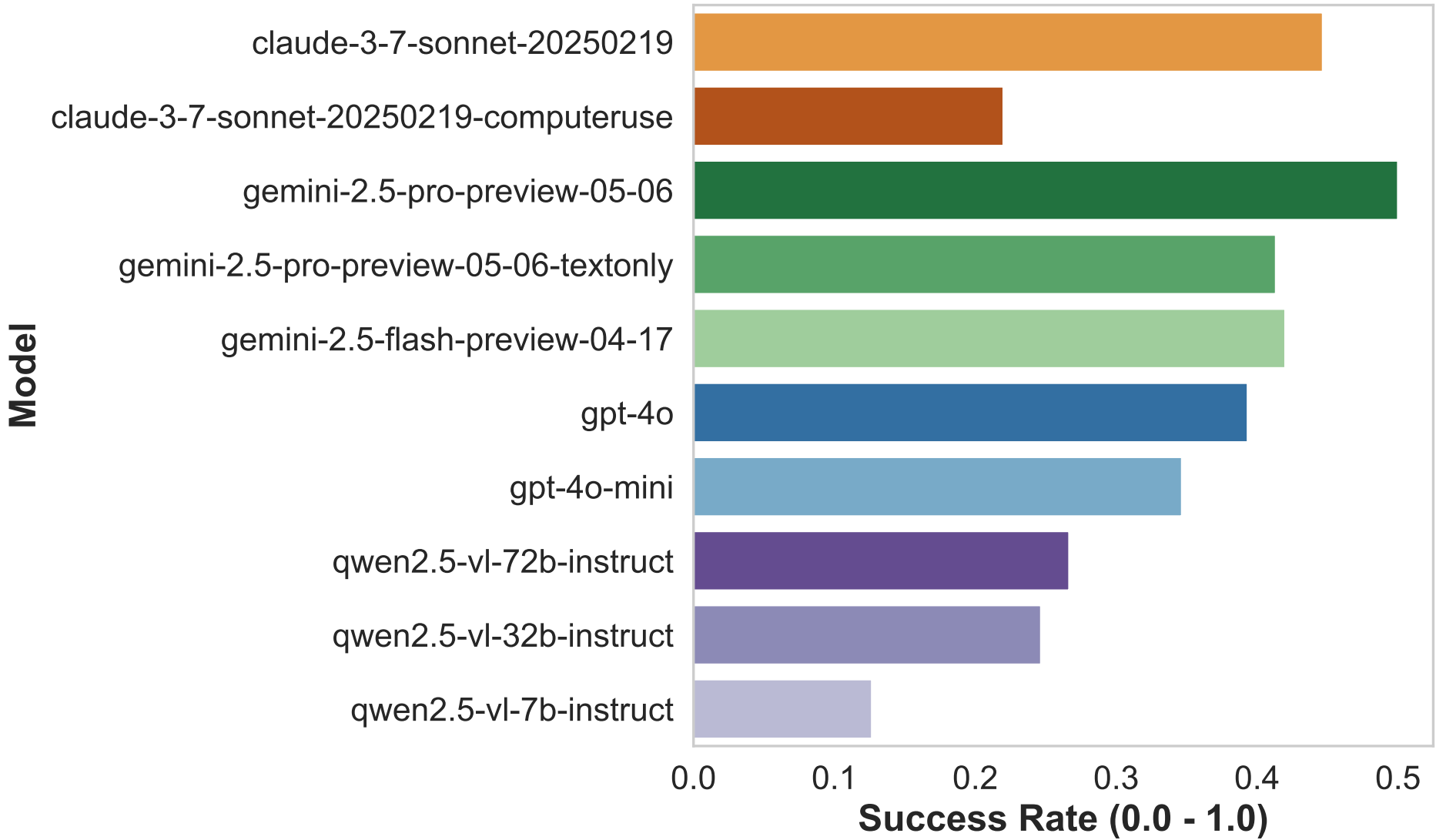
claude-3-7-sonnet-20250219
claude-3-7-sonnet-20250219-computeruse
gemini-2.5-pro-preview-05-06
gemini-2.5-pro-preview-05-06-textonly
gemini-2.5-flash-preview-04-17
gpt-4o
gpt-4o-mini
qwen2.5-vl-72b-instruct
qwen2.5-vl-32b-instruct
qwen2.5-vl-7b-instruct
Model
0.0
0.1
0.2
0.3
0.4
0.5
Success Rate (0.0 - 1.0)