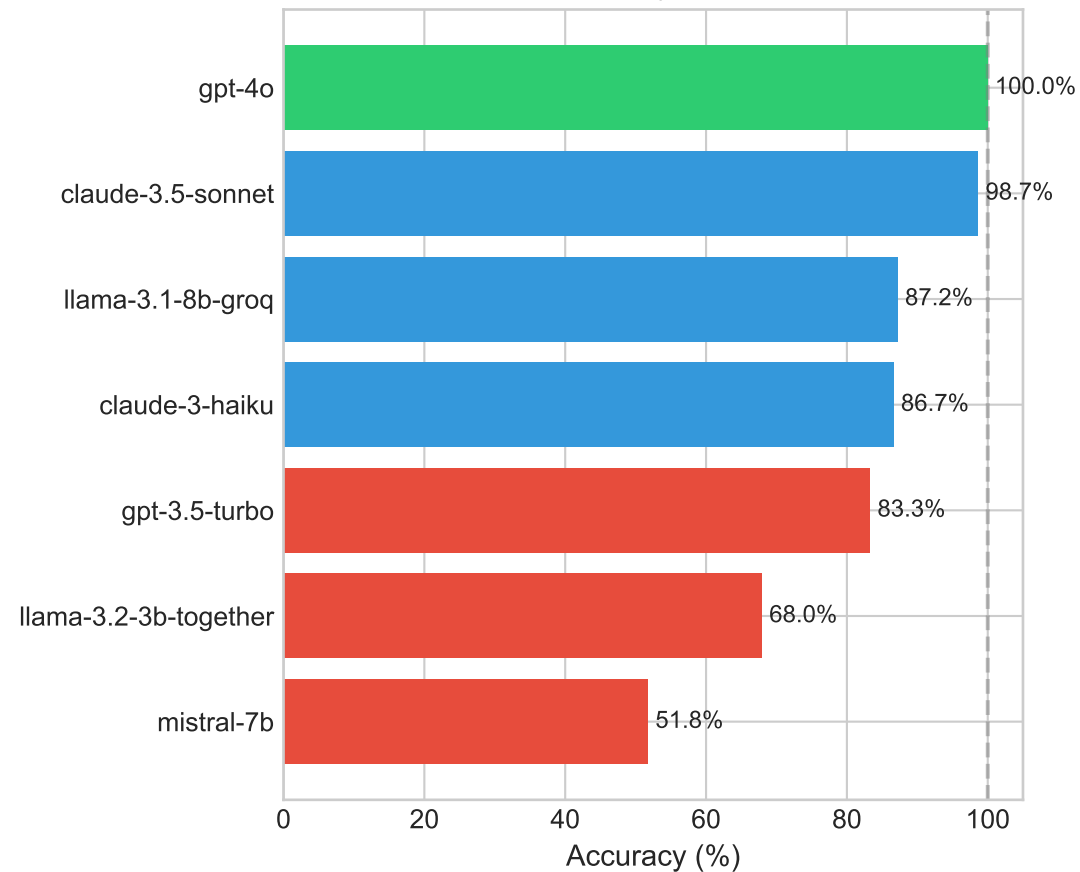


A. Model Accuracy on Probe Items



B. Misconception Alignment (When Wrong)

