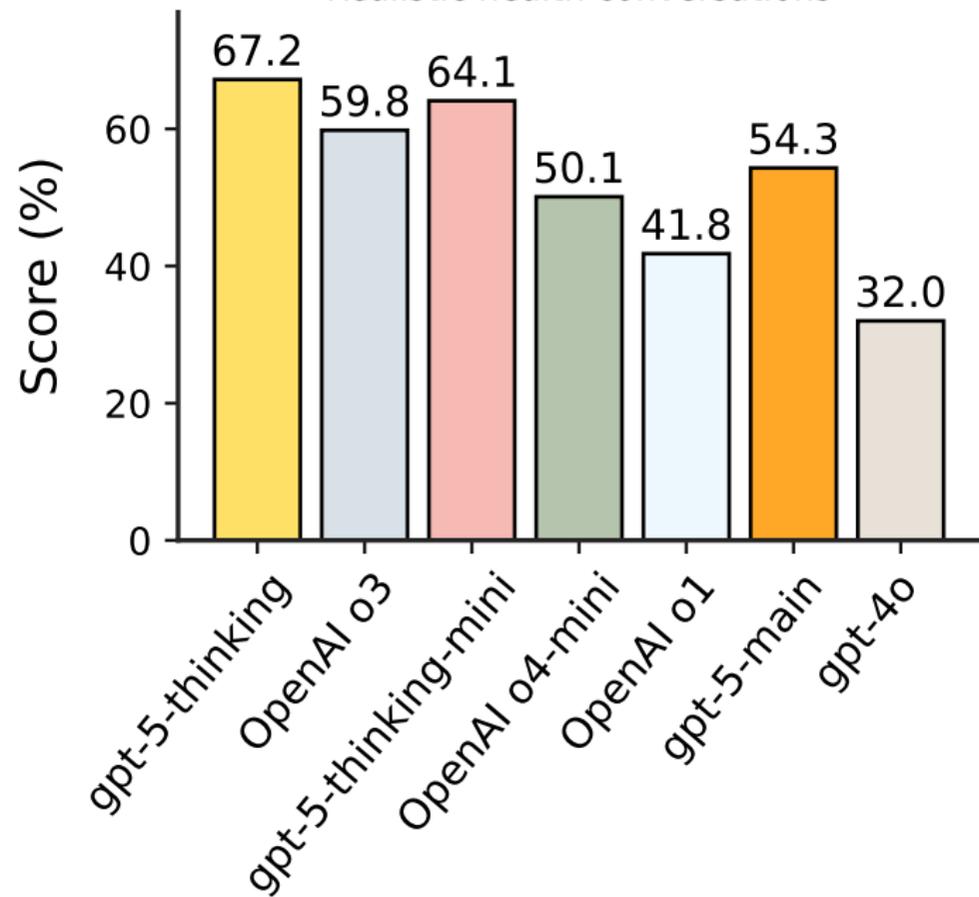


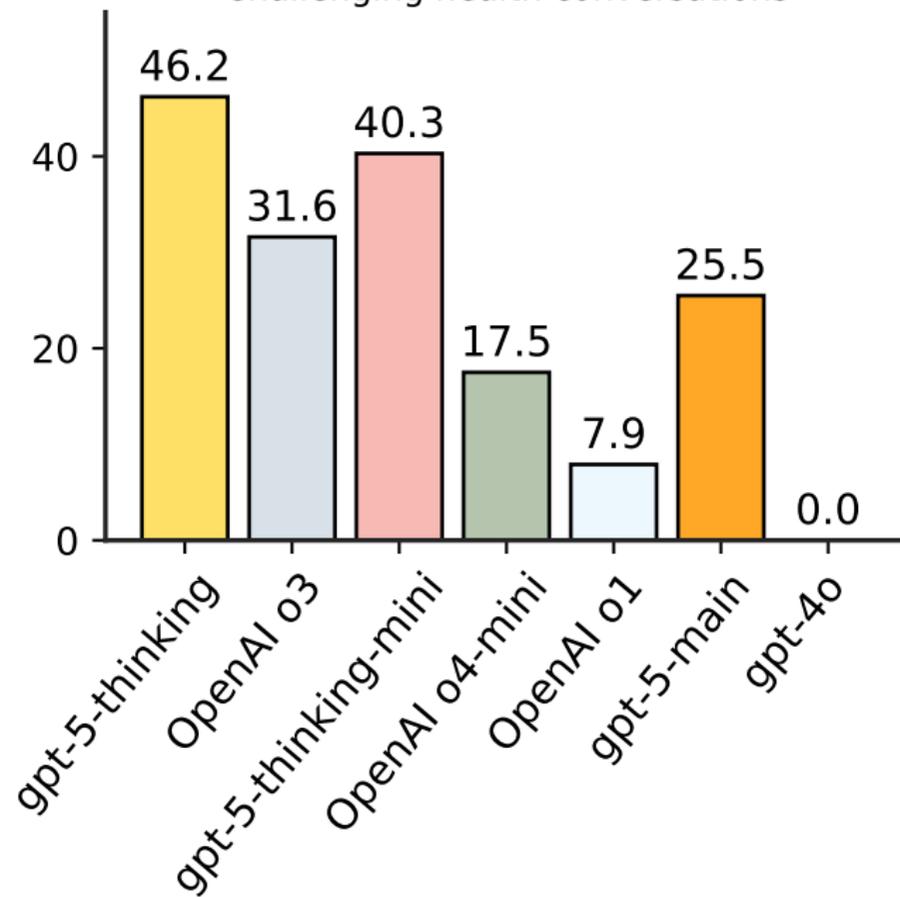
### HealthBench

Realistic health conversations



### HealthBench Hard

Challenging health conversations



### HealthBench Consensus

Validated by 2+ physicians

