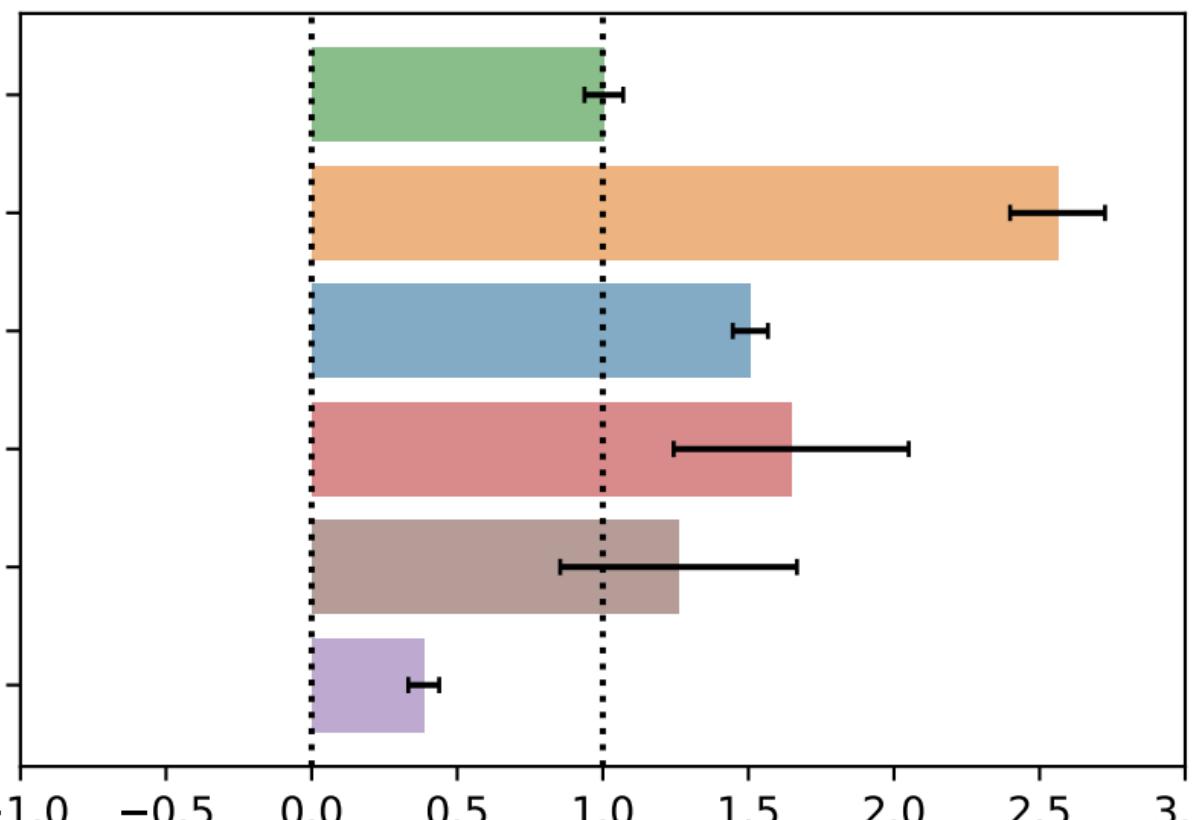


gpt-4

Probabilistic Reasoning



claude-2

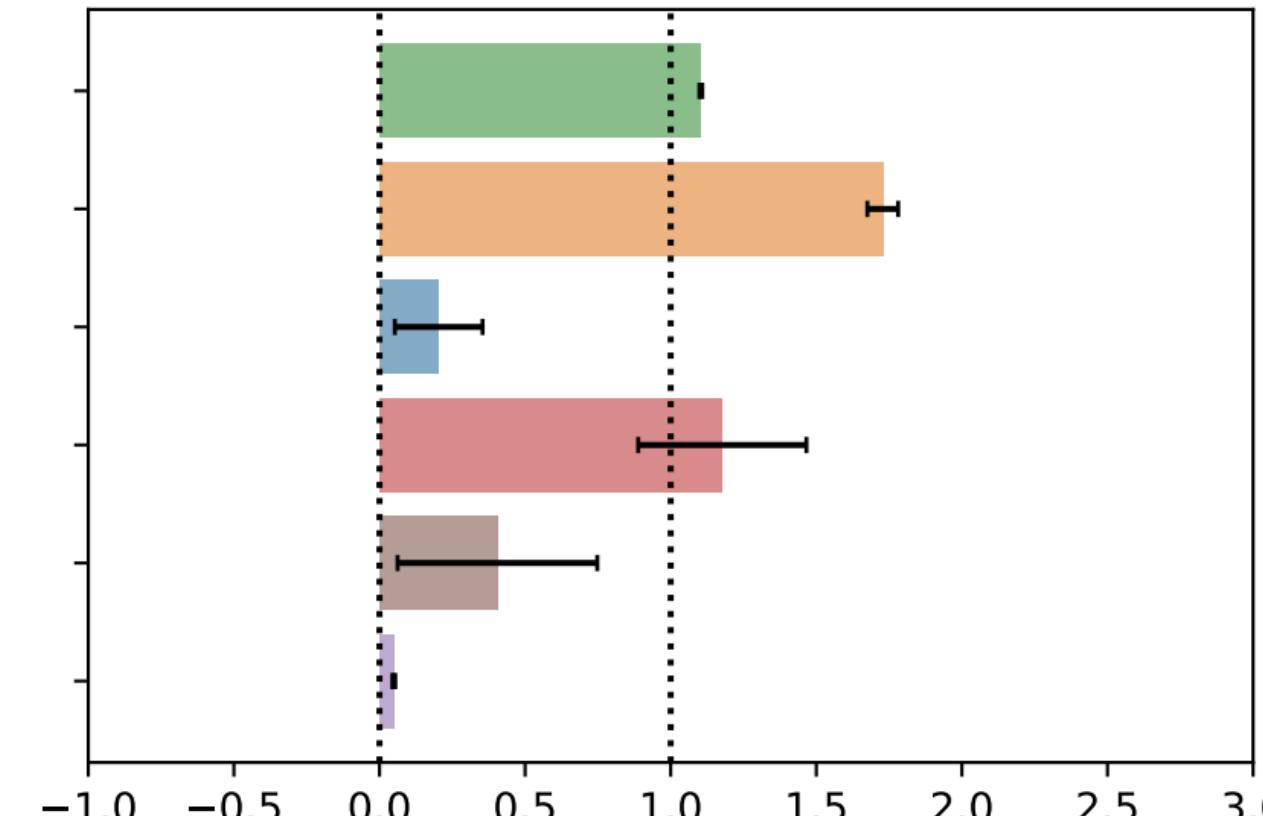
Horizon Task

Restless Bandit

Instrumental Learning

Two Step Task

BART



Values (Normalized: Random=0, Human average=1)