

Task Category

realtime_responsiveness

----- claude-3-7-sonnet-20250219 (Normalized to 1.0)

technical_fluency

adversarial_resistance

visual_comprehension

cognitive_abilities

0.0

0.2

0.4

0.6

0.8

1.0

Computer-use relative weighted success rate
(claude-3-7-sonnet-20250219=1.0)

