

-0.148

X

-0.146

Agent Score Expected (ASE)

1.22

-0.150

_Aclaude-3-5-sonnet-20241022_Jgpt-4o-mini-2024-07-18

-0.142

-0.144

	ASD vs ASE for Debate_t0_n4 (brier)									
1	1.45						_/	Adeepseek-cha	at_Jgpt-4o-min	i-2024
nce (ASD)	1.40				_	Aclaude-3-5-s	onnet-202410	22-math_eval_	Jgpt-4o-mini-2	2024-0
Agent Score Difference (ASD)	1.35		_Ac	laude-3-5-hail	ku-20241022-r			1022_Jgpt-4o- 4-07-18_Bclaud		
Agent										
1	1.30 t-40-mini-2024		de-3-5-haiku-2 -0.2							
	-0.30	Score Expe			225	-0.2	200			

ASD vs ASE for Consultancy_t0_n4 (brier) _Adeepseek-chat_Jgpt-4o-mini-2024-07-18 0.35 ıiku-20241022-math_eval_Jgpt-4o-mini-2024-07-18 _Aclaude-3-5-haiku-20241022_Jgpt-4o-mini-2024-07-18 Agent Score Difference (ASD) 0.30 _Aclaude-3-5-sonnet-20241022_Jgpt-4o-mini-2024-07-18 0.25 0.20 _Aclaude-3-5-sonnet-20241022-math_ev 0.15-0.065 -0.075-0.070-0.060 -0.055 Agent Score Expected (ASE)

			ASD \	vs ASE for C	Consultancy	_t0_n2 (brie	er)		
						_Adeepse	ek- ch at_Jgpt-4o-mi	ni-2024-07-18	
	0.5								
(ASD)									
Agent Score Difference (ASD)	0.4				5-sonnet-20241022				
e Diff	าล	aiku- 2 0241022_Jgpt-4	4o-mini-2024-07-18.	_Aclaude-3-5-haiku	i-20241022-math_	eval_Jgpt-4o-mini-2	2024-07-18		
nt Scor									
Ager	0.3								
	0.2						laude-3-5-sonnet-20	024 1 022-math_e\	
	-0.09 -0.08 -0.07 Agent Score Expected (ASE)								

			,	ASD vs AS	SE for Deb	ate_t1_n2	(brier)			
1.5	5		_Adeepseek-chat_Jgpt-4o-m							
						_Aclaude-	-3-5-sonnet-202	41022_Jgpt-4o-r	mini-2024-07-18_	
(OSV										
Agent Score Difference (ASD)		net-20241022-m	nath_eval_Jgpt-4d	o-m <mark>ip</mark> i-2024-07-	18_Bclaude-3-5-	sonnet-2024102	22-math_eval			
Score Diff	4o-m	ipi-2024-07-18_	Bclaude-3-5-hail	ku-20241022-m	ath_eval					
Agent 9	3									
				_Aclaude-3-5-h	aiku-20241022_J	gpt-4o-mini-202	24-07-18_Bclaud	e-3-5-haiku-202	41022	
	-0.35 -0.30 -0.25 -0.20 Agent Score Expected (ASE)									

ASD vs ASE for Consultancy_t1_n4 (brier) Adeepseek-chat_Jgpt-4o-mini-2 0.40 Aclaude-3-5-haiku-20241022_Jgpt-4o-mini-2024-07-18 Agent Score Difference (ASD) 0.35 _Aclaude-3-5-haiku-20241022-math_eval_Jgpt-4o _Aclaude-3-5-sonnet-20241022_Jg 0.30 0.25_Aclaude-3-5_zsonnet-20241022-math_eval_Jgpt-4o-mini-2024-07-18

-0.0575

Agent Score Expected (ASE)

-0.0550

-0.0525

-0.0625

-0.0600