Table 1: RL experimental results with the added Finance Domain.

Method	**Generalist Domain**		**Health Domain**		**Finance Domain**	
	Filtered Set	**LMArena**	**Medical-o1**		**Finance**	
	Win%	**Win%**	**Win%**	**Score**	**Win %**	**Score**
Base Policy	5.2	4.1	10.8	0.1721	5.84	0.1738
SFT	35.9	29.6	25.8	0.2999	26.04	0.2218
Initial, Prompt only	31.3	29.7	21.7	0.3004	37.23	0.2683
1 Good Pair	33.5	32.8	22.4	0.2912	39.07	0.2694
1 Great Pair	**36.8**	**42.24**	**26.5**	**0.3163**	**39.23**	**0.2838**
4 Great Pairs	38.7	34.7	31.4	0.3348	48.91	0.2961
4 Great & Diverse Pairs	**39.7**	**35.1**	**34.4**	**0.3513**	**49.58**	**0.3018**