Average Factual Correctness Score by Model and Question (Q1-Q10)

	Claude 3.7 Sonnet	1.000	0.994	0.991	0.449	1.000	1.000	1.000	1.000	0.997	0.061	
	Llama 3.1 8B	0.030	0.180	0.100	0.100	0.367	0.163	0.064	0.047	0.003	0.052	
Models	Llama 3.3 70B	0.494	0.299	0.700	0.479	0.273	0.471	0.293	0.192	0.600	0.266	
	Ministral 8B	0.146	0.446	0.000	0.101	0.000	0.326	0.030	0.103	0.139	0.354	
	Qwen 2.5 72B	0.900	0.796	1.000	0.649	0.927	0.477	0.580	0.214	0.200	0.383	
		Q1 Q2 Q3 Q4 Q5 Q6 Q1 Q8 Q9 Q10 Questions										