

course	temperature	top_k	model	num_quest	avg_answer	bleu	rouge_l	bert_sim
architecture	0	2	gpt	34	1048.647	0.049802	0.187906	0.824987
architecture	0	2	claude	34	1076.412	0.040878	0.169937	0.797502
architecture	0	4	gpt	34	1101.647	0.051793	0.188291	0.829679
architecture	0	4	claude	34	1313.176	0.038719	0.163811	0.797052
architecture	0	6	gpt	34	1180	0.050198	0.188393	0.836251
architecture	0	6	claude	34	1385.206	0.03639	0.159811	0.800874
architecture	0	8	gpt	34	1158.765	0.050553	0.187701	0.841502
architecture	0	8	claude	34	1384.618	0.038378	0.159259	0.797988
architecture	0.2	2	gpt	34	1037.441	0.052046	0.191202	0.823501
architecture	0.2	2	claude	34	1092.559	0.042923	0.171418	0.809896
architecture	0.2	4	gpt	34	1116.559	0.04878	0.187907	0.836487
architecture	0.2	4	claude	34	1316.412	0.035565	0.158194	0.806677
architecture	0.2	6	gpt	34	1203	0.044207	0.186443	0.835354
architecture	0.2	6	claude	34	1443.559	0.035646	0.155787	0.799428
architecture	0.2	8	gpt	34	1178.912	0.047872	0.189487	0.836421
architecture	0.2	8	claude	34	1322.382	0.038152	0.162077	0.795391
architecture	0.5	2	gpt	34	1051.324	0.051302	0.187019	0.83456
architecture	0.5	2	claude	34	1098.147	0.042044	0.168672	0.798504
architecture	0.5	4	gpt	34	1116.029	0.053902	0.186429	0.831267
architecture	0.5	4	claude	34	1294.324	0.040262	0.16155	0.809824
architecture	0.5	6	gpt	34	1175.853	0.04684	0.184925	0.833605
architecture	0.5	6	claude	34	1358.765	0.039423	0.159191	0.797468
architecture	0.5	8	gpt	34	1161.559	0.049949	0.186253	0.82721
architecture	0.5	8	claude	34	1360.853	0.035017	0.159669	0.794477
architecture	0.7	2	gpt	34	1072.029	0.045347	0.179433	0.832951
architecture	0.7	2	claude	34	1113.706	0.043539	0.169466	0.799765
architecture	0.7	4	gpt	34	1101.765	0.044632	0.187784	0.83187
architecture	0.7	4	claude	34	1292.824	0.039822	0.1598	0.803484
architecture	0.7	6	gpt	34	1196.706	0.040877	0.185133	0.829832
architecture	0.7	6	claude	34	1376.706	0.03322	0.15363	0.801844
architecture	0.7	8	gpt	34	1257.176	0.047321	0.181368	0.835078
architecture	0.7	8	claude	34	1356.735	0.036035	0.158106	0.797813
architecture	1	2	gpt	34	1044.824	0.050056	0.181329	0.810512
architecture	1	2	claude	34	1068.824	0.043632	0.168629	0.801452
architecture	1	4	gpt	34	1144	0.045563	0.18545	0.838862
architecture	1	4	claude	34	1256.735	0.039478	0.167233	0.813588
architecture	1	6	gpt	34	1193.324	0.041505	0.180469	0.831949
architecture	1	6	claude	34	1400.765	0.036218	0.161427	0.806081
architecture	1	8	gpt	34	1213.794	0.041731	0.179629	0.829153
architecture	1	8	claude	34	1315	0.038298	0.157531	0.808576
machine_learning	0	2	gpt	34	734.3529	0.031408	0.170127	0.732714
machine_learning	0	2	claude	34	957.8824	0.031798	0.155308	0.721867
machine_learning	0	4	gpt	34	790.7941	0.030619	0.174427	0.769438
machine_learning	0	4	claude	34	1016.706	0.02959	0.156464	0.723951
machine_learning	0	6	gpt	34	819	0.024937	0.163852	0.762651
machine_learning	0	6	claude	34	1167.382	0.020257	0.136875	0.719431

machine_1e	0	8 gpt	34	845.5882	0.030023	0.168654	0.762208
machine_1e	0	8 claude	34	1176.059	0.025133	0.14077	0.738498
machine_1e	0.2	2 gpt	34	723.7941	0.033485	0.170701	0.734076
machine_1e	0.2	2 claude	34	916.3529	0.029419	0.153826	0.713805
machine_1e	0.2	4 gpt	34	799.8235	0.029255	0.1696	0.761223
machine_1e	0.2	4 claude	34	1052.5	0.026804	0.151158	0.734792
machine_1e	0.2	6 gpt	34	792.5588	0.032738	0.175744	0.758268
machine_1e	0.2	6 claude	34	1110.265	0.023031	0.143748	0.732768
machine_1e	0.2	8 gpt	34	820.6176	0.030317	0.170351	0.7662
machine_1e	0.2	8 claude	34	1164	0.028027	0.148268	0.742624
machine_1e	0.5	2 gpt	34	720.4706	0.036467	0.171501	0.732315
machine_1e	0.5	2 claude	34	893.5882	0.033894	0.155538	0.720516
machine_1e	0.5	4 gpt	34	819.0882	0.031726	0.17116	0.752347
machine_1e	0.5	4 claude	34	1030.324	0.025833	0.150632	0.733909
machine_1e	0.5	6 gpt	34	814.2941	0.029323	0.174052	0.748533
machine_1e	0.5	6 claude	34	1175.324	0.021783	0.13865	0.727559
machine_1e	0.5	8 gpt	34	782.1765	0.031224	0.177217	0.7559
machine_1e	0.5	8 claude	34	1145.735	0.02312	0.142111	0.729928
machine_1e	0.7	2 gpt	34	758.6471	0.035327	0.170308	0.736123
machine_1e	0.7	2 claude	34	949.1176	0.025165	0.147948	0.721932
machine_1e	0.7	4 gpt	34	829.6471	0.022647	0.162166	0.754709
machine_1e	0.7	4 claude	34	1008	0.031798	0.160551	0.728413
machine_1e	0.7	6 gpt	34	845.8529	0.029254	0.166212	0.753311
machine_1e	0.7	6 claude	34	1106.765	0.025157	0.145583	0.727141
machine_1e	0.7	8 gpt	34	902.6176	0.029762	0.163938	0.757664
machine_1e	0.7	8 claude	34	1035.294	0.029677	0.153247	0.739743
machine_1e	1	2 gpt	34	763.2941	0.031528	0.158092	0.727246
machine_1e	1	2 claude	34	958.8235	0.029356	0.154912	0.731291
machine_1e	1	4 gpt	34	852.8529	0.031362	0.163795	0.751773
machine_1e	1	4 claude	34	1037.471	0.024383	0.15207	0.72643
machine_1e	1	6 gpt	34	825.5294	0.029407	0.17264	0.760578
machine_1e	1	6 claude	34	1127.559	0.025627	0.138155	0.730055
machine_1e	1	8 gpt	34	848.6471	0.025726	0.167779	0.766485
machine_1e	1	8 claude	34	1127.059	0.021307	0.137454	0.731212
networking	0	2 gpt	42	803.3333	0.013887	0.110803	0.667036
networking	0	2 claude	42	985.5952	0.011731	0.108919	0.668151
networking	0	4 gpt	42	906.381	0.015537	0.116393	0.687378
networking	0	4 claude	42	1119.357	0.012458	0.106732	0.684006
networking	0	6 gpt	42	919.7381	0.015878	0.113214	0.690143
networking	0	6 claude	42	1162.667	0.013627	0.106793	0.680166
networking	0	8 gpt	42	937.7381	0.013896	0.112544	0.681964
networking	0	8 claude	42	1146.071	0.011104	0.106253	0.688757
networking	0.2	2 gpt	42	809.119	0.011328	0.106061	0.666036
networking	0.2	2 claude	42	940.2143	0.011111	0.107476	0.665226
networking	0.2	4 gpt	42	911.0238	0.015393	0.115423	0.689407
networking	0.2	4 claude	42	1088.286	0.011176	0.10648	0.680934
networking	0.2	6 gpt	42	926.5	0.01611	0.112676	0.683783

networking	0.2	6 claude	42	1159.976	0.012993	0.106378	0.680879
networking	0.2	8 gpt	42	997.0952	0.016727	0.116832	0.702783
networking	0.2	8 claude	42	1132.619	0.011628	0.107452	0.693856
networking	0.5	2 gpt	42	787.8571	0.012851	0.111223	0.654497
networking	0.5	2 claude	42	987.7381	0.00734	0.105961	0.660251
networking	0.5	4 gpt	42	903.9524	0.01575	0.112308	0.692262
networking	0.5	4 claude	42	1092.69	0.01175	0.111023	0.679602
networking	0.5	6 gpt	42	945.1429	0.015798	0.113242	0.681107
networking	0.5	6 claude	42	1157.19	0.011668	0.107316	0.685567
networking	0.5	8 gpt	42	1014.881	0.015099	0.11572	0.702379
networking	0.5	8 claude	42	1151.167	0.013409	0.109449	0.688549
networking	0.7	2 gpt	42	803.0238	0.012536	0.110413	0.648458
networking	0.7	2 claude	42	938.7619	0.013654	0.105167	0.658689
networking	0.7	4 gpt	42	897.0952	0.01448	0.115496	0.680624
networking	0.7	4 claude	42	1090.024	0.009578	0.105782	0.680381
networking	0.7	6 gpt	42	999.5238	0.011271	0.112644	0.679931
networking	0.7	6 claude	42	1148.405	0.013184	0.108317	0.690986
networking	0.7	8 gpt	42	1034	0.013147	0.118303	0.710496
networking	0.7	8 claude	42	1146.619	0.013081	0.105165	0.687915
networking	1	2 gpt	42	780.3333	0.015046	0.105286	0.647757
networking	1	2 claude	42	971.8333	0.013057	0.110027	0.659984
networking	1	4 gpt	42	917.7143	0.013412	0.108234	0.683708
networking	1	4 claude	42	1078.19	0.010059	0.106058	0.687156
networking	1	6 gpt	42	998.9762	0.009131	0.105543	0.677978
networking	1	6 claude	42	1194.524	0.010137	0.103701	0.687766
networking	1	8 gpt	42	1016.524	0.015255	0.114994	0.716642
networking	1	8 claude	42	1198.548	0.009761	0.099269	0.682395

avg_judge_score

0.870588

0.770588

0.885294

0.811765

0.894118

0.814706

0.885294

0.785294

0.885294

0.797059

0.879412

0.832353

0.897059

0.811765

0.879412

0.773529

0.879412

0.8

0.891176

0.832353

0.917647

0.808824

0.932353

0.75

0.882353

0.797059

0.891176

0.805882

0.905882

0.764706

0.923529

0.811765

0.854412

0.770588

0.882353

0.811765

0.914706

0.861765

0.908824

0.808824

0.908824

0.85

0.927941

0.920588

0.939706

0.861765

0.960294
0.942647
0.883824
0.854412
0.927941
0.935294
0.955882
0.908824
0.970588
0.948529
0.911765
0.851471
0.955882
0.941176
0.964706
0.867647
0.967647
0.925
0.914706
0.838235
0.95
0.920588
0.938235
0.920588
0.958824
0.942647
0.920588
0.885294
0.938235
0.935294
0.964706
0.923529
0.976471
0.920588
0.738095
0.633333
0.764286
0.728571
0.809524
0.747619
0.811905
0.778571
0.72619
0.638095
0.745238
0.738095
0.804762

0.764286
0.833333
0.761905
0.728571
0.640476
0.797619
0.735714
0.804762
0.740476
0.816667
0.778571
0.716667
0.659524
0.785714
0.740476
0.786905
0.761905
0.827381
0.761905
0.72619
0.686905
0.792857
0.741667
0.794048
0.752381
0.821429
0.740476