

Table 1. *pss* scores by task, grouped by project and sorted by magnitude of the developer improvement.

Project	Hash	Dev	o4-mini				Gemini 2.5 Pro				DeepSeek V3				DeepSeek R1			
			NH	Bench	Prob	Both	NH	Bench	Prob	Both	NH	Bench	Prob	Both	NH	Bench	Prob	Both
RoaringBlimp	34bce1b	4.99	2.98	2.88	2.98	4.59	4.07	4.01	3.88	3.62	-	1.17	5.01	1.16	1.72	4.28	5.16	2.39
	53ba593	2.14	1.18	1.00	-	1.90	1.30	1.34	1.00	1.97	-	1.91	-	-	1.00	1.34	0.93	0.93
	5b54af6	1.94	-	-	-	-	1.93	1.32	1.14	1.92	-	-	-	-	0.99	-	1.00	0.89
	abd8df	1.14	1.13	1.10	1.12	1.08	0.99	1.07	1.01	1.06	1.00	1.01	1.15	0.99	1.17	1.06	0.99	1.07
	1de6825	1.11	-	1.10	1.10	1.11	1.11	1.11	1.11	1.11	-	1.11	-	-	-	1.11	-	-
	Avg	1.92	1.32	1.28	1.30	1.60	1.62	1.53	1.38	1.74	1.00	1.20	1.42	1.03	1.15	1.47	1.36	1.16
Kafka	b25c96a	1401.47	0.78	1954.51	1910.43	1941.48	1898.94	1905.23	1879.81	1901.39	-	1986.78	1991.64	2019.21	-	1932.48	1946.33	1918.99
	9bb2f78	654.58	647.51	673.51	642.90	631.33	647.27	549.45	654.25	632.65	673.32	0.99	539.97	1.00	534.61	663.91	664.91	634.21
	4fa5bdc	4.54	0.71	1.02	3.63	1.04	3.95	4.34	4.10	3.79	1.04	3.68	1.09	3.83	1.10	-	1.03	1.05
	cf34ca	3.95	-	-	-	2.87	-	3.68	-	1.83	-	-	-	-	-	-	-	-
	e3ccf20	2.81	0.96	0.98	2.59	0.97	0.96	0.93	3.45	4.12	0.98	-	1.96	2.59	-	-	0.39	0.91
	c8af740	2.10	2.18	2.25	2.24	2.27	2.14	2.19	2.18	2.19	-	-	1.03	0.98	-	1.13	2.07	-
	e2e2c62	2.04	1.25	-	-	-	1.66	-	-	1.06	0.01	-	1.53	4.38	-	0.97	-	-
	c05403f	1.55	1.03	1.24	1.27	1.31	1.24	1.24	1.52	1.24	-	1.23	-	1.25	1.30	1.23	1.29	1.54
	922a95a	1.28	0.81	0.89	1.27	1.46	1.14	8.81	1.26	1.26	-	-	-	1.37	-	-	-	1.20
	aecd47b	1.05	2.25	2.60	1.06	1.05	2.42	2.30	1.06	1.06	2.17	2.27	1.05	-	2.58	2.43	1.05	1.05
	Avg	7.27	2.11	4.93	5.81	5.26	5.96	7.81	6.14	6.49	1.26	2.70	4.55	3.29	2.14	4.60	4.13	4.32
Nerdy	2ed95c9	6.30	3.26	-	1.00	1.01	1.05	1.89	1.12	0.81	1.00	2.03	-	-	0.99	-	2.92	-
	87ec2f8	3.00	0.99	1.00	1.00	1.00	1.00	1.01	2.50	2.52	1.13	-	1.13	1.14	-	-	1.11	2.46
	b2caab0	1.80	1.01	1.00	1.00	1.01	1.00	1.00	1.00	1.52	1.04	1.05	1.04	1.04	1.03	1.10	1.05	1.04
	83a19d5	1.71	0.93	1.27	1.20	1.15	1.35	0.83	1.28	1.24	0.69	0.60	0.64	0.64	0.68	0.61	0.59	0.75
	2791f0f	1.45	1.00	0.97	1.29	1.29	1.01	1.03	1.37	1.42	0.62	1.19	-	1.39	-	-	1.18	1.47
	67d3a78	1.25	0.99	0.99	-	1.32	0.99	1.07	1.22	1.02	1.04	1.05	1.28	1.18	-	0.91	1.28	1.25
	deea51e	1.25	0.99	1.13	1.33	1.24	1.00	1.13	1.11	1.25	-	-	-	-	-	-	-	-
	5a9b28a	1.21	1.15	1.24	1.33	1.27	0.95	1.14	1.09	1.08	0.43	0.49	0.33	0.35	0.38	0.58	0.34	0.80
	8a8337e	1.19	1.01	1.02	-	1.26	1.06	1.04	1.11	1.01	1.06	1.06	1.28	1.27	1.04	1.14	1.30	1.21
	feb804d	1.10	1.08	1.08	1.09	1.05	1.07	1.09	1.08	1.11	-	0.65	0.67	0.59	0.67	-	0.65	0.69
	Avg	1.71	1.14	1.07	1.08	1.15	1.04	1.10	1.24	1.24	0.87	0.94	0.88	0.89	0.84	0.91	0.98	1.09
Presto	bf618d6	145169.44	1.01	1.04	1.02	103640.35	1.03	1.03	1.03	117653.53	1.01	1.01	-	1.02	1.03	1.03	1.04	-
	20dcaed	2142.71	2134.32	2141.47	2162.46	2136.73	2119.50	2126.92	2129.77	2136.70	-	-	-	-	-	-	-	-
	239f726	1033.72	1.04	-	1080.35	-	1.00	1087.83	1.00	1.03	1.02	-	1.08	1.06	-	1.02	1060.24	1087.03
	8d7f587	136.32	-	1.00	134.64	131.43	-	1.01	130.84	132.86	-	-	-	-	-	-	-	-
	ce74e4b	38.54	1.01	0.99	-	1.00	1.00	1.00	1.00	-	0.99	1.00	1.00	0.98	-	0.99	0.99	0.99
	8abba9e	30.27	-	-	-	-	-	30.05	-	-	-	-	-	-	-	-	-	-
	3f63ade	12.86	0.97	7.87	7.48	7.50	0.98	0.93	7.27	8.41	0.99	-	-	0.83	1.03	-	-	-
	46d219f	10.12	0.97	0.99	9.94	9.97	0.97	1.01	10.03	10.02	-	-	-	-	-	-	-	-
	a26f440	9.66	0.99	1.00	0.99	-	0.98	0.89	1.00	0.99	0.96	0.96	0.96	-	0.96	0.96	-	0.95
	4c71b83	2.45	0.99	1.00	1.04	1.09	1.75	1.83	2.27	2.33	0.94	0.98	0.98	1.92	1.19	1.02	-	1.94
	3a4f2c3	1.89	1.00	1.00	1.07	1.08	1.00	1.00	1.10	1.11	1.01	1.00	1.09	1.08	1.21	1.02	1.05	1.07
	6908344	1.87	0.94	0.33	2.24	1.00	0.97	-	1.86	2.22	-	-	-	-	-	-	-	-
	9e010a7	1.80	0.99	0.96	0.96	0.96	0.98	0.92	0.87	0.87	0.59	0.58	0.93	0.92	0.90	0.74	0.74	0.77
	53aa7f6	1.40	0.99	1.01	1.01	1.01	1.01	1.00	1.00	1.00	0.98	1.00	0.95	1.00	1.00	1.00	1.01	0.99
	b54d248	1.29	1.26	1.22	1.18	1.23	1.20	1.26	1.26	1.23	-	-	-	-	1.21	1.23	1.09	-
	efd0834	1.18	-	-	1.16	1.16	-	-	1.00	1.00	-	-	-	-	-	-	-	0.98
	c3cb7a9	1.10	1.01	1.01	1.08	1.10	1.00	1.10	1.01	1.00	0.96	0.99	0.99	0.98	0.96	0.99	1.09	0.98
	297b089	1.07	1.03	1.06	1.00	1.00	1.04	1.00	1.02	1.04	1.00	1.00	1.01	1.01	1.06	1.00	1.05	1.07
	659c3ee	1.06	1.05	1.07	1.01	1.08	1.08	1.01	-	-	1.06	1.06	0.99	1.01	1.05	1.05	0.99	1.07
	Avg	13.42	1.51	1.60	3.76	4.62	1.56	2.69	2.65	4.99	0.97	0.97	1.00	1.03	1.03	1.00	1.44	1.48
Overall		5.86	1.51	1.84	2.77	3.08	1.94	2.62	2.51	3.42	1.01	1.25	1.42	1.30	1.18	1.45	1.67	1.72