

HumanEval	Original	A1	A2	A3	A4	D1	D2	D3	D4	E1	E2	E3	E4	E5	E6	S1	S2	P1	P2	C1	C2	C3	Avg.	K of Pass@K
InCoder	16.1	15.9	13.4	16.0	15.2	15.7	16.3	16.0	14.6	13.6	16.2	15.7	15.4	15.7	15.6	15.1	15.5	13.9	15.3	14.9	15.1	14.6	15.2	1
	26.8	24.9	20.3	24.6	23.0	23.2	26.4	24.8	23.3	21.9	24.5	24.8	23.5	25.9	24.2	22.8	24.9	22.6	26.8	24.6	24.2	22.1	24.0	10
Code Llama	30.5	30.0	28.5	29.1	29.4	30.3	28.7	28.8	29.3	29.6	29.2	29.7	30.9	29.6	29.2	29.7	29.3	28.6	29.8	30.0	29.4	28.8	29.4	1
	45.2	48.6	41.5	43.2	45.4	45.9	42.8	44.5	43.6	46.6	43.1	42.6	46.9	45.4	45.1	44.5	42.3	47.8	46.8	44.0	42.3	42.8	44.6	10
CodeGeeX2	33.2	31.8	30.6	31.6	32.1	31.4	32.6	32.2	31.0	31.4	32.3	32.8	32.5	31.9	32.2	31.9	32.3	30.9	31.4	32.8	32.6	31.7	31.9	1
	48.0	48.1	44.8	46.8	45.4	46.8	44.8	45.4	45.9	45.7	49.2	48.2	47.8	45.8	47.2	45.4	46.5	44.7	47.8	47.0	46.1	47.3	46.5	10
CodeGen2.5	33.4	33.4	32.5	32.3	31.0	31.3	32.5	32.9	32.6	32.2	32.3	33.2	33.1	32.8	32.9	31.5	33.4	32.1	30.2	32.4	31.7	31.1	32.2	1
	48.3	47.8	49.5	46.8	45.6	47.0	47.8	46.6	46.0	46.8	48.2	49.5	49.1	49.4	49.0	47.4	48.0	47.7	45.0	48.8	44.4	45.8	47.4	10
StarCoder	33.9	33.0	33.2	33.9	33.5	32.9	32.9	33.4	33.7	32.0	34.4	33.1	34.2	33.4	33.9	34.0	33.2	33.2	31.3	33.7	33.6	33.2	33.3	1
	51.4	49.8	46.8	48.6	51.1	49.5	49.6	51.8	48.9	48.4	51.3	49.6	48.7	48.6	49.3	49.1	48.6	48.0	45.8	50.7	48.0	48.2	48.1	10
WizardCoder	43.9	43.5	42.1	42.8	43.8	42.9	43.2	43.9	42.0	41.7	44.4	43.5	43.8	44.3	43.3	42.8	43.3	41.1	44.0	44.7	43.1	41.5	43.1	1
	54.8	56.8	55.0	54.6	55.2	54.3	57.0	55.0	52.3	53.2	56.2	55.2	55.1	57.8	55.6	54.4	54.6	54.5	57.2	58.2	56.7	52.3	55.3	10
P@1 Avg.	31.8	31.3	30.0	31.0	30.9	30.7	31.0	31.2	30.5	30.1	31.5	31.3	31.7	31.3	31.2	30.8	31.2	30.0	30.3	31.4	30.9	30.2	30.9	1
P@10 Avg.	45.7	46.0	43.0	44.1	44.3	44.4	44.7	43.3	43.8	43.8	45.4	45.0	45.2	45.5	45.1	43.9	44.2	44.2	44.9	45.6	43.6	43.1	44.5	10
MBPP	Original	A1	A2	A3	A4	D1	D2	D3	D4	E1	E2	E3	E4	E5	E6	S1	S2	P1	P2	C1	C2	C3	Avg.	
InCoder	19.7	17.5	19.0	20.3	19.7	18.3	19.2	20.0	19.0	18.4	19.6	20.7	19.7	19.4	19.9	19.1	19.9	19.8	15.6	19.3	19.5	19.1	19.2	1
	41.4	37.8	40.7	40.6	39.7	39.3	40.0	40.8	39.7	39.3	39.1	40.8	40.6	40.0	40.4	39.8	41.2	41.0	36.4	39.7	39.7	39.3	39.8	10
Code Llama	49.6	50.2	48.9	49.0	48.0	48.4	48.2	49.9	48.8	47.8	50.7	48.1	48.0	47.3	49.0	48.4	49.4	48.5	46.0	48.5	46.7	47.9	48.5	1
	68.0	68.2	67.6	69.5	68.1	69.5	67.2	69.3	69.4	70.1	70.7	70.7	66.8	68.9	68.8	68.3	68.7	69.1	69.0	68.1	68.4	69.4	68.8	10
CodeGeeX2	45.8	46.0	45.3	46.0	45.5	45.6	44.8	45.6	46.1	43.7	47.0	43.6	41.8	44.5	44.2	45.7	43.1	45.2	41.7	44.5	44.8	43.6	44.7	1
	64.0	64.4	62.7	63.8	64.2	63.5	64.6	63.6	64.4	62.6	65.0	65.0	63.0	64.2	63.4	63.8	62.5	63.3	62.8	63.4	62.8	61.8	63.6	10
CodeGen2.5	54.3	54.6	53.9	54.4	53.8	54.2	52.6	54.1	54.0	53.8	53.1	52.6	51.7	53.1	53.0	54.2	54.5	53.9	51.6	52.9	52.9	52.3	53.4	1
	69.4	68.7	68.2	68.8	68.6	68.8	68.5	68.5	68.0	69.1	67.7	69.4	68.5	69.6	69.2	68.0	68.4	68.3	68.9	68.7	68.3	67.8	68.6	10
StarCoder	56.0	55.5	55.5	55.7	55.6	54.6	54.9	56.0	56.0	54.3	55.5	53.9	53.9	55.8	54.8	55.6	56.5	55.8	54.5	55.6	54.2	55.4	55.2	1
	71.6	70.5	70.7	71.4	71.1	70.4	70.7	70.3	70.4	68.6	70.6	70.4	69.9	71.2	68.9	70.7	72.1	71.4	70.9	70.4	68.6	70.0	70.4	10
WizardCoder	57.1	56.2	56.9	56.6	56.1	56.4	55.8	56.5	57.0	56.3	56.8	56.9	56.6	57.2	56.6	57.4	55.5	55.0	56.1	56.2	56.1	54.6	56.3	1
	67.0	66.7	67.3	67.3	65.4	66.0	66.7	66.9	66.7	66.4	66.7	67.3	63.8	67.3	67.4	66.7	66.3	64.2	67.0	66.1	65.7	64.8	66.3	10
P@1 Avg.	47.1	46.7	46.6	47.0	46.4	46.3	45.9	47.0	46.8	45.7	47.1	46.0	45.3	46.2	46.2	46.7	46.5	46.4	44.2	46.2	45.7	45.5	46.2	1
P@10 Avg.	63.6	62.7	62.9	63.6	62.8	62.9	62.9	63.2	63.1	62.7	63.3	63.9	62.1	63.5	63.0	62.9	63.2	62.9	62.5	62.7	62.3	62.2	62.9	10