

A vs B	language	Prefers A	Neutral	Prefers B	Avg Score A	Avg Score B
human vs gpt-3.5-turbo	ja	16.7%	2.4%	81.0%	0.77	0.88
	th	11.4%	3.6%	84.9%	0.73	0.85
	vn	13.5%	2.7%	83.8%	0.77	0.87
human vs wangchanglm-7.5B-sft-en	ja	57.1%	2.4%	40.5%	0.73	0.62
	th	38.0%	1.2%	60.8%	0.73	0.78
	vn	64.9%	5.4%	29.7%	0.74	0.67
human vs wangchanglm-7.5B-sft-enth	ja	58.5%	7.3%	34.1%	0.70	0.60
	th	34.5%	3.0%	62.4%	0.70	0.75
	vn	64.9%	8.1%	27.0%	0.68	0.79
human vs openthaigpt	th	49.7%	3.0%	47.3%	0.67	0.68
gpt-3.5-turbo vs wangchanglm-7.5B-sft-en	ja	71.4%	2.4%	26.2%	0.76	0.60
	th	81.3%	1.2%	17.5%	0.79	0.67
	vn	83.8%	0.0%	16.2%	0.81	0.65
gpt-3.5-turbo vs wangchanglm-7.5B-sft-enth	ja	82.1%	5.1%	12.8%	0.78	0.57
	th	80.0%	3.0%	17.0%	0.78	0.67
	vn	75.7%	10.8%	13.5%	0.71	0.52
gpt-3.5-turbo vs openthaigpt	th	86.0%	3.0%	11.0%	0.78	0.60
openthaigpt vs wangchanglm-7.5B-sft-en	th	42.8%	0.6%	56.6%	0.67	0.72
openthaigpt vs wangchanglm-7.5B-sft-enth	th	48.5%	0.6%	50.9%	0.69	0.71