Data type	Model	Func Acc	PN_HR	PN_MR	Args Acc	PR	SR	Args Acc/Func Acc	PR/Func Acc	SR/Func Acc
sQsA	GPT-4o	71.28%	16.47%	4.69%	53.99%	43.43%	34.65%	75.74%	60.92%	48.61%
	Llama-3.1-70B-Instruct-INT4	83.91%	18.99%	1.87%	61.58%	51.91%	39.85%	73.39%	61.87%	47.49%
	Qwen2.5-72B-Instruct-Int4	74.75%	18.62%	13.39%	47.16%	36.82%	22.02%	63.09%	49.25%	29.47%
	Llama-3.1-8B-Instruct	88.36%	27.40%	6.31%	55.98%	44.10%	31.93%	63.35%	49.91%	36.13%
	Qwen2.5-7B-Instruct	72.77%	26.15%	7.55%	44.63%	33.56%	21.78%	61.33%	46.12%	29.93%
	Ministral-8B-Instruct-2410	54.70%	49.93%	5.13%	22.89%	9.67%	6.93%	41.85%	17.67%	12.66%
	xLAM-7b-fc-r	90.84%	63.73%	1.05%	30.28%	11.53%	8.17%	33.33%	12.69%	8.99%
	ToolACE-8B	46.53%	87.29%	0.30%	4.85%	0.25%	0.00%	10.42%	0.53%	0.00%
mQmA	GPT-4o	71.28%	15.89%	3.71%	52.95%	46.62%	40.09%	74.28%	65.40%	56.24%
	Llama-3.1-70B-Instruct-INT4	84.15%	17.46%	2.14%	61.18%	56.47%	46.03%	72.69%	67.10%	54.70%
	Qwen2.5-72B-Instruct-Int4	74.50%	17.87%	4.26%	53.37%	48.87%	42.07%	71.64%	65.60%	56.47%
	Llama-3.1-8B-Instruct	88.36%	26.57%	5.24%	56.01%	50.58%	41.08%	63.38%	57.24%	46.49%
	Qwen2.5-7B-Instruct	73.01%	26.56%	3.77%	45.99%	41.43%	33.66%	62.98%	56.74%	46.10%
	Ministral-8B-Instruct-2410	54.95%	52.68%	3.83%	21.30%	11.22%	8.17%	38.77%	20.42%	14.86%
	xLAM-7b-fc-r	91.08%	62.98%	1.28%	28.37%	14.25%	11.38%	31.15%	15.64%	12.49%
	ToolACE-8B	46.53%	81.00%	0.23%	6.43%	0.45%	0.00%	13.82%	0.98%	0.00%
mQsA	GPT-4o	71.78%	15.86%	6.73%	51.67%	42.65%	35.39%	71.98%	59.42%	49.31%
	Llama-3.1-70B-Instruct-INT4	83.66%	23.28%	3.26%	56.80%	47.47%	34.40%	67.89%	56.74%	41.12%
	Qwen2.5-72B-Instruct-Int4	74.50%	19.10%	9.55%	48.94%	38.65%	27.72%	65.68%	51.88%	37.20%
	Llama-3.1-8B-Instruct	88.36%	30.71%	7.70%	52.49%	41.60%	28.21%	59.40%	47.07%	31.93%
	Qwen2.5-7B-Instruct	73.26%	33.08%	5.53%	41.76%	31.77%	21.78%	57.00%	43.36%	29.72%
	Ministral-8B-Instruct-2410	54.70%	59.69%	6.26%	17.08%	6.93%	3.22%	31.23%	12.67%	5.88%
	xLAM-7b-fc-r	90.84%	72.46%	1.88%	21.39%	8.52%	3.71%	23.55%	9.38%	4.09%
	ToolACE-8B	46.53%	87.59%	0.59%	4.03%	0.25%	0.00%	8.67%	0.53%	0.00%
sQmA	GPT-4o	72.52%	17.87%	3.49%	53.37%	47.18%	40.09%	73.59%	65.05%	55.29%
	Llama-3.1-70B-Instruct-INT4	83.91%	17.00%	2.64%	60.33%	56.18%	45.54%	71.90%	66.96%	54.27%
	Qwen2.5-72B-Instruct-Int4	73.76%	18.31%	5.01%	51.89%	48.37%	41.08%	70.35%	65.58%	55.70%
	Llama-3.1-8B-Instruct	88.36%	27.41%	5.36%	55.48%	50.53%	41.33%	62.79%	57.18%	46.77%
	Qwen2.5-7B-Instruct	73.26%	26.91%	4.49%	45.99%	41.00%	33.16%	62.77%	55.96%	45.27%
	Ministral-8B-Instruct-2410	54.95%	53.06%	4.02%	21.09%	11.09%	8.17%	38.39%	20.19%	14.86%
	xLAM-7b-fc-r	90.84%	63.94%	1.40%	28.16%	14.04%	11.13%	31.00%	15.46%	12.26%
	ToolACE-8B	46.53%	81.00%	0.23%	6.33%	0.45%	0.00%	13.60%	0.98%	0.00%
w/o SO	GPT-4o	74.15%	15.62%	5.30%	55.74%	48.91%	41.16%	75.18%	65.88%	55.40%
	Llama-3.1-70B-Instruct-INT4	85.77%	21.38%	2.08%	60.82%	52.27%	43.36%	70.93%	60.95%	50.54%
	Qwen2.5-72B-Instruct-Int4	79.31%	18.51%	10.38%	51.81%	45.37%	34.22%	65.29%	57.06%	42.90%
	Llama-3.1-8B-Instruct	89.00%	28.32%	5.07%	56.26%	46.56%	37.73%	63.22%	52.31%	42.38%
	Qwen2.5-7B-Instruct	76.08%	26.00%	6.91%	47.03%	39.06%	29.89%	61.81%	51.26%	39.14%
	Ministral-8B-Instruct-2410	60.82%	49.87%	4.70%	24.92%	12.34%	9.83%	41.01%	20.20%	16.02%
	xLAM-7b-fc-r	90.10%	62.67%	1.00%	30.72%	13.63%	11.20%	34.09%	15.14%	12.44%
	ToolACE-8B	52.64%	83.67%	0.23%	7.21%	0.14%	0.00%	13.56%	0.27%	0.00%
w/ SO	GPT-4o	74.91%	16.37%	4.26%	55.08%	48.62%	40.68%	73.56%	64.82%	54.19%
	Llama-3.1-70B-Instruct-INT4	85.63%	24.89%	1.71%	56.74%	47.64%	34.98%	66.29%	55.63%	40.82%
	Qwen2.5-72B-Instruct-Int4	79.10%	19.11%	9.81%	51.40%	44.61%	33.67%	64.93%	56.16%	42.25%
	Llama-3.1-8B-Instruct	89.00%	30.72%	3.69%	52.46%	42.45%	30.51%	58.96%	47.69%	34.28%
	Qwen2.5-7B-Instruct	76.35%	29.11%	4.58%	43.35%	35.59%	23.50%	56.82%	46.55%	30.70%
	Ministral-8B-Instruct-2410	60.82%	53.40%	2.93%	23.31%	11.46%	8.25%	38.40%	18.71%	13.36%
	xLAM-7b-fc-r	90.10%	67.51%	0.89%	24.62%	11.89%	8.25%	27.32%	13.22%	9.17%
	ToolACE-8B	52.71%	84.45%	0.16%	6.63%	0.12%	0.00%	12.40%	0.24%	0.00%
w/o mSv	GPT-4o	75.25%	15.27%	5.64%	56.37%	51.01%	43.65%	74.91%	67.78%	57.99%
	Llama-3.1-70B-Instruct-INT4	86.45%	22.34%	2.14%	60.60%	52.46%	44.77%	70.10%	60.68%	51.79%
	Qwen2.5-72B-Instruct-Int4	80.80%	18.50%	9.12%	53.70%	48.68%	39.04%	66.45%	60.25%	48.31%
	Llama-3.1-8B-Instruct	89.08%	28.71%	4.58%	56.23%	47.27%	39.69%	63.12%	53.06%	44.56%
	Qwen2.5-7B-Instruct	77.51%	25.73%	6.73%	48.31%	41.38%	33.20%	62.33%	53.38%	42.83%
	Ministral-8B-Instruct-2410	63.21%	49.82%	4.48%	25.87%	13.41%	11.00%	40.93%	21.21%	17.41%
	xLAM-7b-fc-r	89.74%	62.38%	0.96%	30.82%	14.33%	12.22%	34.34%	15.97%	13.62%
	ToolACE-8B	55.22%	82.40%	0.20%	8.06%	0.09%	0.00%	14.59%	0.17%	0.00%
w/ mSv	GPT-4o	76.01%	14.36%	6.26%	48.14%	41.25%	16.74%	63.34%	54.28%	22.02%
	Llama-3.1-70B-Instruct-INT4	86.45%	22.82%	4.52%	55.53%	44.59%	26.05%	64.23%	51.58%	30.14%
	Qwen2.5-72B-Instruct-Int4	80.52%	17.68%	8.32%	51.93%	43.12%	28.59%	64.49%	53.55%	35.51%
	Llama-3.1-8B-Instruct	89.08%	29.06%	4.72%	50.60%	39.49%	20.79%	56.79%	44.33%	23.33%
	Qwen2.5-7B-Instruct	77.79%	26.26%	5.64%	42.71%	33.25%	14.01%	54.90%	42.74%	18.01%
	Ministral-8B-Instruct-2410	62.93%	46.33%	3.47%	22.71%	10.16%	2.92%	36.09%	16.14%	4.63%
	xLAM-7b-fc-r	89.74%	57.47%	2.29%	26.10%	10.72%	2.82%	29.09%	11.95%	3.14%
	ToolACE-8B	55.31%	75.31%	0.15%	8.46%	0.07%	0.00%	15.29%	0.13%	0.00%
w/o External	GPT-4o	71.42%	8.92%	3.57%	59.64%	54.62%	48.57%	83.50%	76.47%	67.99%
	Llama-3.1-70B-Instruct-INT4	85.30%	15.41%	1.57%	66.33%	57.17%	49.59%	77.76%	67.02%	58.13%
	Qwen2.5-72B-Instruct-Int4	77.95%	13.41%	9.90%	54.57%	50.34%	41.02%	70.00%	64.57%	52.61%
	Llama-3.1-8B-Instruct	88.97%	20.44%	4.04%	63.73%	54.29%	45.71%	71.62%	61.02%	51.37%
	Qwen2.5-7B-Instruct	74.28%	16.95%	6.26%	52.04%	45.02%	36.73%	70.05%	60.61%	49.45%
	Ministral-8B-Instruct-2410	59.18%	42.71%	3.95%	28.23%	14.41%	11.83%	47.71%	24.35%	19.99%
	xLAM-7b-fc-r	90.20%	54.29%	0.86%	38.23%	20.93%	17.55%	42.39%	23.21%	19.45%
	ToolACE-8B	55.10%	78.67%	0.27%	9.65%	0.14%	0.00%	17.50%	0.25%	0.00%
w/ External	GPT-4o	73.87%	11.14%	8.91%	52.81%	50.78%	38.57%	71.49%	68.73%	52.20%
	Llama-3.1-70B-Instruct-INT4	84.89%	20.39%	3.16%	55.77%	49.28%	34.48%	65.69%	58.05%	40.62%
	Qwen2.5-72B-Instruct-Int4	78.16%	15.86%	12.82%	48.73%	46.52%	33.46%	62.34%	59.51%	42.81%
	Llama-3.1-8B-Instruct	88.97%	24.96%	9.59%	49.22%	43.96%	25.51%	55.32%	49.40%	28.66%
	Qwen2.5-7B-Instruct	74.48%	23.91%	7.68%	42.18%	38.95%	24.69%	56.62%	52.29%	33.15%
	Ministral-8B-Instruct-2410	58.77%	46.15%	9.06%	20.42%	11.68%	6.12%	34.74%	19.88%	10.41%
	xLAM-7b-fc-r	90.20%	60.21%	1.87%	27.81%	17.50%	10.81%	30.83%	19.40%	11.99%
	ToolACE-8B	55.30%	76.98%	2.69%	6.76%	0.14%	0.00%	12.22%	0.25%	0.00%
Overall	GPT-4o	73.25%	14.78%	5.26%	53.98%	47.51%	37.96%	73.76%	64.88%	51.93%
	Llama-3.1-70B-Instruct-INT4	85.01%	20.40%	2.51%	59.57%	51.55%	39.91%	70.09%	60.66%	46.96%
	Qwen2.5-72B-Instruct-Int4	77.34%	17.70%	9.26%	51.35%	45.13%	34.29%	66.43%	58.34%	44.32%
	Llama-3.1-8B-Instruct	88.76%	27.43%	5.63%	54.85%	46.08%	34.25%	61.79%	51.92%	38.59%
	Qwen2.5-7B-Instruct	74.88%	26.07%	5.92%	45.40%	38.10%	27.24%	60.66%	50.90%	36.43%
	Ministral-8B-Instruct-2410	58.50%	50.36%	4.78%	22.78%	11.24%	7.64%	38.91%	19.14%	13.01%
	xLAM-7b-fc-r	90.37%	62.76%	1.35%	28.65%	13.73%	9.73%	31.71%	15.20%	10.76%
	ToolACE-8B	51.24%	81.84%	0.51%	6.84%	0.21%	0.00%	13.21%	0.43%	0.00%