Mistral-Nemo-Instruct-2407

		8/8	7/8	6/8	5/8	4/8	3/8	≤2/8
EN	ORG	100%	97.0%	82.4%	68.1%	51.9%	45.5%	22.6%
	SS	100%	94.3%	81.0%	65.0%	61.8%	51.5%	6.7%
	SO	100%	90.5%	77.7%	52.3%	55.6%	36.7%	12.8%
	PSS	100%	93.4%	91.7%	71.4%	60.7%	43.5%	9.6%
	POS	100%	93.1%	90.2%	70.2%	50.0%	72.7%	8.6%
DE	ORG	100%	93.0%	73.0%	63.8%	37.0%	27.3%	16.1%
	SS	100%	91.1%	86.0%	66.7%	50.0%	27.3%	10.0%
	SO	100%	90.5%	79.8%	61.5%	41.7%	26.7%	20.5%
	PSS	100%	91.0%	84.4%	60.3%	35.7%	39.1%	8.2%
	POS	100%	95.7%	84.8%	57.4%	47.2%	31.8%	6.9%
FR	ORG	100%	93.0%	91.9%	70.2%	59.3%	40.9%	14.5%
	SS	100%	89.4%	74.0%	60.0%	44.1%	30.3%	15.0%
	SO	100%	83.8%	70.2%	75.4%	36.1%	43.3%	15.4%
	PSS	100%	91.0%	76.0%	69.8%	28.6%	26.1%	9.6%
	POS	100%	94.0%	83.7%	66.0%	50.0%	27.3%	12.1%
ES	ORG	100%	93.0%	75.7%	51.1%	18.5%	9.1%	3.2%
	SS	100%	96.7%	89.0%	75.0%	58.8%	51.5%	18.3%
	SO	100%	94.3%	87.2%	58.5%	61.1%	40.0%	15.4%
	PSS	100%	96.7%	88.5%	76.2%	71.4%	34.8%	13.7%
	POS	100%	99.1%	93.5%	68.1%	47.2%	36.4%	12.1%
IT	ORG	100%	94.0%	82.4%	51.1%	51.9%	18.2%	11.3%
	SS	100%	96.7%	91.0%	73.3%	58.8%	33.3%	25.0%
	SO	100%	93.3%	84.0%	81.5%	52.8%	40.0%	23.1%
	PSS	100%	97.5%	91.7%	76.2%	78.6%	52.2%	13.7%
	POS	100%	95.7%	88.0%	78.7%	61.1%	27.3%	19.0%
PT	ORG	100%	88.0%	74.3%	53.2%	40.7%	9.1%	3.2%
	SS	100%	99.2%	96.0%	91.7%	58.8%	51.5%	15.0%
	SO	100%	98.1%	92.6%	81.5%	77.8%	53.3%	23.1%
	PSS	100%	95.1%	93.8%	77.8%	64.3%	30.4%	13.7%
	POS	100%	98.3%	98.9%	85.1%	77.8%	45.5%	5.2%
KO	ORG	100%	77.0%	63.5%	66.0%	63.0%	72.7%	38.7%
	SS	100%	67.5%	42.0%	33.3%	35.3%	30.3%	13.3%
	SO	100%	73.3%	54.3%	46.2%	36.1%	36.7%	23.1%
	PSS	100%	60.7%	39.6%	41.3%	25.0%	39.1%	16.4%
	POS	100%	62.1%	32.6%	40.4%	27.8%	27.3%	13.8%
JA	ORG	100%	65.0%	56.8%	76.6%	77.8%	77.3%	29.0%
	SS	100%	65.0%	41.0%	35.0%	32.4%	24.2%	13.3%
	SO	100%	76.2%	54.3%	43.1%	38.9%	23.3%	7.7%
	PSS	100%	74.6%	34.4%	27.0%	35.7%	34.8%	13.7%
	POS	100%	62.1%	28.3%	34.0%	38.9%	31.8%	20.7%

LLaMA-3.1 Instruct-8B

		8/8	7/8	6/8	5/8	4/8	3/8	≤2/8
EN	ORG	100%	87.3%	55.6%	59.4%	54.2%	27.6%	13.0%
	SS	100%	100%	98.9%	93.4%	71.9%	58.9%	30.8%
	SO	100%	97.3%	90.6%	82.3%	69.4%	56.9%	21.1%
	PSS	100%	97.5%	96.4%	88.4%	84.8%	74.3%	27.1%
	POS	100%	94.4%	96.7%	93.0%	89.6%	61.1%	23.9%
DE	ORG	100%	95.5%	82.2%	68.8%	45.8%	48.3%	4.3%
	SS	100%	91.4%	80.7%	42.6%	45.6%	26.8%	8.3%
	SO	100%	91.8%	77.6%	59.7%	40.3%	33.3%	6.8%
	PSS	100%	89.9%	81.1%	63.8%	34.8%	25.7%	5.1%
	POS	100%	93.0%	86.2%	59.2%	45.8%	22.2%	8.8%
FR	ORG	100%	90.0%	82.2%	46.9%	54.2%	17.2%	4.3%
	SS	100%	91.4%	92.0%	73.8%	52.6%	35.7%	8.3%
	SO	100%	94.5%	90.6%	71.0%	54.8%	35.3%	12.0%
	PSS	100%	89.9%	95.5%	69.6%	50.0%	28.6%	9.3%
	POS	100%	95.8%	91.1%	74.6%	52.1%	44.4%	12.4%
ES	ORG	100%	94.5%	64.4%	43.8%	33.3%	6.9%	0.0%
	SS	100%	96.6%	98.9%	91.8%	71.9%	55.4%	17.9%
	SO	100%	100%	91.8%	82.3%	66.1%	49.0%	13.5%
	PSS	100%	97.5%	98.2%	85.5%	71.7%	62.9%	7.6%
	POS	100%	100%	97.6%	87.3%	56.2%	63.9%	15.0%
IT	ORG	100%	91.8%	84.4%	62.5%	37.5%	13.8%	13.0%
	SS	100%	94.8%	93.2%	75.4%	56.1%	37.5%	19.2%
	SO	100%	93.2%	94.1%	77.4%	59.7%	41.2%	9.8%
	PSS	100%	96.2%	93.7%	84.1%	56.5%	28.6%	9.3%
	POS	100%	98.6%	95.1%	77.5%	50.0%	41.7%	13.3%
PT	ORG	100%	85.5%	64.4%	56.2%	33.3%	20.7%	8.7%
	SS	100%	94.8%	94.3%	82.0%	61.4%	55.4%	9.6%
	SO	100%	98.6%	92.9%	87.1%	62.9%	43.1%	15.8%
	PSS	100%	97.5%	95.5%	72.5%	65.2%	54.3%	10.2%
	POS	100%	100%	95.9%	84.5%	68.8%	33.3%	13.3%
KO	ORG	100%	83.6%	84.4%	75.0%	62.5%	82.8%	52.2%
	SS	100%	60.3%	28.4%	16.4%	24.6%	17.9%	4.5%
	SO	100%	71.2%	29.4%	22.6%	25.8%	27.5%	14.3%
	PSS	100%	55.7%	18.9%	23.2%	28.3%	11.4%	5.9%
	POS	100%	69.0%	22.8%	14.1%	22.9%	16.7%	5.3%
JA	ORG	100%	71.8%	82.2%	87.5%	79.2%	82.8%	69.6%
	SS	100%	70.7%	13.6%	24.6%	15.8%	12.5%	7.1%
	SO	100%	53.4%	32.9%	17.7%	21.0%	13.7%	6.8%
	PSS	100%	75.9%	20.7%	13.0%	8.7%	14.3%	5.1%
	POS	100%	49.3%	14.6%	9.9%	14.6%	16.7%	2.7%

Gemini-2.5 Flash

		8/8	7/8	6/8	5/8	4/8	3/8	≤2/8
EN	ORG	100%	95.1%	80.0%	69.2%	60.0%	46.4%	9.4%
	SS	100%	94.5%	73.2%	74.1%	53.6%	25.0%	9.3%
	SO	100%	85.8%	67.9%	62.5%	50.0%	28.6%	2.9%
	PSS	100%	92.4%	89.8%	47.6%	30.0%	20.0%	10.0%
	POS	100%	89.0%	80.0%	46.2%	66.7%	36.4%	11.1%
DE	ORG	100%	92.7%	73.3%	65.4%	44.0%	21.4%	10.6%
	SS	100%	87.3%	76.8%	55.6%	46.4%	40.0%	7.0%
	SO	100%	90.3%	73.6%	62.5%	45.5%	28.6%	14.3%
	PSS	100%	90.5%	85.7%	66.7%	50.0%	40.0%	10.0%
	POS	100%	90.2%	70.0%	65.4%	33.3%	18.2%	7.4%
FR	ORG	100%	93.9%	88.9%	69.2%	60.0%	46.4%	10.6%
	SS	100%	82.7%	69.6%	48.1%	35.7%	40.0%	4.7%
	SO	100%	80.5%	75.5%	64.6%	31.8%	28.6%	11.4%
	PSS	100%	88.6%	61.2%	66.7%	50.0%	70.0%	3.3%
	POS	100%	86.6%	76.7%	57.7%	50.0%	45.5%	7.4%
ES	ORG	100%	87.8%	82.2%	50.0%	44.0%	28.6%	5.9%
	SS	100%	94.5%	83.9%	59.3%	57.1%	30.0%	14.0%
	SO	100%	93.8%	81.1%	58.3%	31.8%	50.0%	5.7%
	PSS	100%	84.8%	83.7%	47.6%	40.0%	40.0%	0.0%
	POS	100%	91.5%	90.0%	53.8%	50.0%	18.2%	22.2%
IT	ORG	100%	95.1%	77.8%	61.5%	32.0%	50.0%	10.6%
	SS	100%	89.1%	76.8%	74.1%	32.1%	30.0%	11.6%
	SO	100%	89.4%	81.1%	62.5%	50.0%	35.7%	20.0%
	PSS	100%	95.2%	77.6%	71.4%	60.0%	50.0%	10.0%
	POS	100%	93.9%	80.0%	61.5%	50.0%	36.4%	7.4%
PT	ORG	100%	90.2%	73.3%	65.4%	48.0%	42.9%	9.4%
	SS	100%	93.6%	89.3%	55.6%	67.9%	45.0%	16.3%
	SO	100%	94.7%	67.9%	66.7%	54.5%	50.0%	5.7%
	PSS	100%	92.4%	81.6%	61.9%	60.0%	10.0%	6.7%
	POS	100%	89.0%	70.0%	69.2%	33.3%	36.4%	18.5%
KO	ORG	100%	67.1%	62.2%	69.2%	52.0%	32.1%	4.7%
	SS	100%	73.6%	75.0%	74.1%	39.3%	55.0%	25.6%
	SO	100%	87.6%	77.4%	68.8%	72.7%	57.1%	42.9%
	PSS	100%	77.1%	67.3%	90.5%	60.0%	40.0%	20.0%
	POS	100%	89.0%	70.0%	73.1%	50.0%	63.6%	18.5%
JA	ORG	100%	78.0%	62.2%	50.0%	60.0%	32.1%	16.5%
	SS	100%	84.5%	55.4%	59.3%	67.9%	35.0%	18.6%
	SO	100%	77.9%	75.5%	54.2%	63.6%	21.4%	17.1%
	PSS	100%	79.0%	53.1%	47.6%	50.0%	30.0%	6.7%
	POS	100%	70.7%	63.3%	73.1%	66.7%	45.5%	11.1%

Qwen3-8B

		8/8	7/8	6/8	5/8	4/8	3/8	≤2/8
EN	ORG	100%	94.1%	87.3%	68.9%	50.9%	38.2%	14.6%
	SS	100%	85.8%	62.5%	48.8%	47.2%	29.6%	2.0%
	SO	100%	84.1%	68.7%	48.8%	40.0%	26.9%	5.1%
	PSS	100%	87.0%	81.0%	58.1%	36.0%	20.0%	8.8%
	POS	100%	90.9%	77.3%	54.5%	33.3%	21.4%	13.3%
DE	ORG	100%	92.9%	71.4%	55.6%	39.6%	20.6%	9.2%
	SS	100%	88.7%	81.9%	61.0%	55.6%	33.3%	8.0%
	SO	100%	85.6%	76.1%	48.8%	36.7%	42.3%	12.8%
	PSS	100%	94.0%	79.4%	67.4%	76.0%	35.0%	4.4%
	POS	100%	92.6%	81.8%	60.6%	66.7%	39.3%	13.3%
FR	ORG	100%	95.3%	85.7%	73.3%	62.3%	38.2%	6.2%
	SS	100%	84.0%	72.2%	63.4%	50.0%	18.5%	14.0%
	SO	100%	86.4%	70.1%	68.3%	33.3%	26.9%	7.7%
	PSS	100%	81.0%	74.6%	51.2%	64.0%	25.0%	7.4%
	POS	100%	86.0%	77.3%	78.8%	23.8%	42.9%	6.7%
ES	ORG	100%	96.5%	87.3%	66.7%	41.5%	41.2%	8.5%
	SS	100%	96.2%	80.6%	61.0%	36.1%	40.7%	14.0%
	SO	100%	93.2%	73.1%	65.9%	53.3%	34.6%	7.7%
	PSS	100%	92.0%	84.1%	67.4%	64.0%	55.0%	11.8%
	POS	100%	95.0%	79.5%	57.6%	47.6%	21.4%	11.1%
IT	ORG	100%	92.9%	82.5%	68.9%	47.2%	32.4%	6.9%
	SS	100%	89.6%	77.8%	70.7%	47.2%	55.6%	18.0%
	SO	100%	91.7%	77.6%	63.4%	56.7%	42.3%	17.9%
	PSS	100%	91.0%	81.0%	60.5%	32.0%	35.0%	8.8%
	POS	100%	94.2%	81.8%	60.6%	47.6%	28.6%	17.8%
PT	ORG	100%	94.1%	90.5%	57.8%	41.5%	41.2%	6.9%
	SS	100%	93.4%	81.9%	70.7%	50.0%	33.3%	14.0%
	SO	100%	96.2%	77.6%	65.9%	66.7%	42.3%	5.1%
	PSS	100%	94.0%	87.3%	79.1%	48.0%	25.0%	7.4%
	POS	100%	93.4%	81.8%	66.7%	47.6%	28.6%	8.9%
KO	ORG	100%	68.2%	39.7%	57.8%	56.6%	55.9%	21.5%
	SS	100%	84.0%	73.6%	70.7%	66.7%	44.4%	20.0%
	SO	100%	84.8%	80.6%	82.9%	63.3%	50.0%	41.0%
	PSS	100%	83.0%	52.4%	67.4%	40.0%	55.0%	26.5%
	POS	100%	71.9%	68.2%	66.7%	66.7%	60.7%	26.7%
JA	ORG	100%	65.9%	55.6%	51.1%	60.4%	32.4%	23.8%
	SS	100%	78.3%	69.4%	53.7%	47.2%	44.4%	28.0%
	SO	100%	78.0%	76.1%	56.1%	50.0%	34.6%	30.8%
	PSS	100%	78.0%	60.3%	48.8%	40.0%	50.0%	32.4%
	POS	100%	76.0%	52.3%	54.5%	66.7%	57.1%	24.4%

Qwen3-14B

		8/8	7/8	6/8	5/8	4/8	3/8	≤2/8
EN	ORG	100%	100%	97.5%	83.0%	70.6%	78.9%	20.5%
	SS	100%	79.1%	66.7%	42.9%	40.0%	36.8%	0.0%
	SO	100%	83.0%	59.6%	53.1%	53.3%	18.2%	5.0%
	PSS	100%	82.2%	75.7%	58.3%	30.8%	14.3%	0.0%
	POS	100%	83.6%	69.7%	75.0%	30.8%	8.3%	0.0%
DE	ORG	100%	93.3%	90.0%	59.6%	41.2%	31.6%	11.3%
	SS	100%	79.1%	64.4%	57.1%	50.0%	31.6%	5.9%
	SO	100%	87.5%	68.1%	56.2%	60.0%	18.2%	5.0%
	PSS	100%	84.4%	73.0%	58.3%	61.5%	35.7%	3.4%
	POS	100%	87.7%	75.8%	50.0%	38.5%	50.0%	13.0%
FR	ORG	100%	96.6%	91.2%	78.7%	50.0%	39.5%	12.6%
	SS	100%	81.4%	73.3%	57.1%	40.0%	15.8%	5.9%
	SO	100%	81.2%	78.7%	59.4%	53.3%	18.2%	25.0%
	PSS	100%	80.0%	73.0%	58.3%	61.5%	28.6%	6.9%
	POS	100%	82.2%	69.7%	56.2%	38.5%	16.7%	8.7%
ES	ORG	100%	100%	92.5%	83.0%	55.9%	31.6%	7.3%
	SS	100%	90.7%	71.1%	61.9%	50.0%	26.3%	5.9%
	SO	100%	89.3%	76.6%	37.5%	46.7%	54.5%	10.0%
	PSS	100%	90.0%	67.6%	50.0%	30.8%	28.6%	13.8%
	POS	100%	95.9%	75.8%	50.0%	53.8%	33.3%	4.3%
IT	ORG	100%	96.6%	95.0%	66.0%	58.8%	50.0%	9.3%
	SS	100%	90.7%	84.4%	61.9%	40.0%	31.6%	23.5%
	SO	100%	87.5%	74.5%	68.8%	40.0%	18.2%	20.0%
	PSS	100%	94.4%	75.7%	79.2%	23.1%	28.6%	6.9%
	POS	100%	84.9%	72.7%	50.0%	53.8%	16.7%	13.0%
PT	ORG	100%	95.5%	93.8%	83.0%	73.5%	39.5%	6.0%
	SS	100%	96.5%	82.2%	66.7%	40.0%	47.4%	5.9%
	SO	100%	92.0%	80.9%	56.2%	46.7%	63.6%	5.0%
	PSS	100%	92.2%	75.7%	58.3%	38.5%	35.7%	10.3%
	POS	100%	91.8%	81.8%	56.2%	61.5%	16.7%	0.0%
KO	ORG	100%	62.9%	15.0%	17.0%	29.4%	10.5%	6.6%
	SS	100%	95.3%	82.2%	76.2%	80.0%	52.6%	23.5%
	SO	100%	94.6%	87.2%	90.6%	46.7%	45.5%	50.0%
	PSS	100%	88.9%	75.7%	79.2%	69.2%	64.3%	44.8%
	POS	100%	90.4%	81.8%	87.5%	84.6%	91.7%	34.8%
JA	ORG	100%	55.1%	25.0%	29.8%	20.6%	18.4%	7.3%
	SS	100%	87.2%	75.6%	76.2%	60.0%	57.9%	23.5%
	SO	100%	84.8%	74.5%	78.1%	53.3%	63.6%	5.0%
	PSS	100%	87.8%	83.8%	58.3%	84.6%	64.3%	41.4%
	POS	100%	83.6%	72.7%	75.0%	38.5%	66.7%	43.5%