

Language	EN (P@K: 1000/2000/3000)						
	P@K	Before-Alignment	Ceiling Performance	Eval - Both	Eval - Abstract	Eval-Physical (all)	Eval - Physical (downsample)
fr	p@1	19.97	85.90	13.17/20.04/24.41	14.08/22.20/25.54	12.78/19.12/23.93	11.22/17.66/23.87
	p@5	23.62	89.33	27.06/34.79/38.30	27.92/38.19/42.24	26.69/33.33/36.61	26.97/33.41/36.75
	p@10	26.63	90.41	32.07/39.58/42.59	35.80/44.39/46.78	30.47/37.53/40.80	31.03/36.99/42.48
	p@30	30.57	92.13	39.44/45.88/48.32	44.39/51.55/52.98	37.32/43.46/46.32	38.42/44.39/48.45
ro	p@1	15.18	81.39	4.51/7.52/10.52	3.34/6.44/10.26	5.01/7.98/10.63	3.82/7.64/10.98
	p@5	17.47	85.25	14.39/18.90/21.55	15.04/19.09/21.72	14.11/18.81/21.47	15.04/19.81/23.15
	p@10	18.97	87.33	18.90/24.27/26.63	20.29/25.78/28.88	18.30/23.62/25.66	19.09/24.34/27.21
	p@30	22.55	90.12	26.77/31.21/34.72	29.83/33.89/39.38	25.46/30.06/32.72	26.25/31.98/34.61
eu	p@1	8.02	87.97	1.36/2.43/4.29	1.91/2.86/3.58	1.12/2.25/4.60	0.72/2.39/4.77
	p@5	9.81	91.62	7.37/11.02/13.17	8.35/11.93/15.04	6.95/10.63/12.37	6.92/11.93/12.89
	p@10	11.60	92.63	10.16/14.53/16.39	11.69/16.95/18.62	9.51/13.50/15.44	9.55/15.04/15.99
	p@30	14.46	94.42	16.46/20.69/22.41	19.57/23.87/25.54	15.13/19.33/21.06	14.32/20.76/22.20
fi	p@1	5.65	79.96	1.57/2.51/4.01	2.63/2.86/4.53	1.12/2.35/3.78	0.72/0.95/2.63
	p@5	8.45	85.18	7.52/10.74/13.17	10.02/12.89/15.75	6.44/9.82/12.07	5.97/10.26/12.41
	p@10	10.02	86.61	12.31/15.25/18.04	15.99/17.90/20.76	10.74/14.11/16.87	10.50/15.27/18.62
	p@30	14.53	88.83	19.76/22.69/24.84	24.58/25.78/28.88	17.69/21.37/23.11	17.90/22.67/23.87
ja	p@1	11.81	95.13	3.15/6.16/8.52	3.82/7.64/10.02	2.86/5.52/7.87	3.58/7.16/9.31
	p@5	22.05	96.99	12.46/19.33/23.41	12.65/20.29/25.54	12.37/18.92/22.49	13.37/20.53/23.39
	p@10	25.70	97.28	18.25/25.48/29.63	19.09/26.97/31.98	17.89/24.85/28.63	20.29/26.25/29.36
	p@30	31.78	98.35	27.63/35.43/39.58	30.31/37.23/40.33	26.48/34.66/39.26	28.16/36.28/40.81
th	p@1	0.64	89.91	0.21/0.43/0.64	0.24/0.24/0.48	0.20/0.51/0.72	0.24/1.19/1.19
	p@5	1.57	92.98	2.72/4.58/6.23	2.39/4.77/6.92	2.86/4.50/5.93	3.58/5.25/6.68
	p@10	2.58	93.49	4.72/6.30/8.95	5.49/6.68/10.74	4.40/6.13/8.18	4.53/6.44/8.35
	p@30	4.37	94.56	10.31/13.53/15.39	11.93/15.99/17.18	9.61/12.47/14.62	9.55/12.89/14.80

Table 23: Full results for Llama2-70B with last-token embedding

Language	EN (P@K: 1000/2000/3000)						
	P@K	Before-Alignment	Ceiling Performance	Eval - Both	Eval - Abstract	Eval-Physical (all)	Eval - Physical (downsample)
fr	p@1	61.92	96.06	34.00/44.31/48.96	42.00/53.22/56.56	30.57/40.49/45.71	30.31/41.29/45.35
	p@5	72.15	99.36	66.43/69.86/71.44	74.22/77.80/79.47	63.09/66.46/68.00	65.63/68.26/69.69
	p@10	74.66	99.50	71.73/74.16/75.02	78.52/81.38/81.38	68.81/71.06/72.29	71.84/73.03/74.22
	p@30	78.53	99.93	77.17/78.95/79.96	83.05/84.25/84.49	74.64/76.69/78.02	77.57/79.00/80.19
ro	p@1	45.67	92.20	16.96/26.56/30.49	22.43/36.52/37.95	14.62/22.29/27.30	13.84/21.48/26.01
	p@5	55.19	97.71	47.96/53.97/55.33	55.13/60.62/61.81	44.89/51.12/52.56	45.35/51.07/52.98
	p@10	57.84	98.57	54.76/58.91/60.63	61.58/66.11/68.74	51.84/55.83/57.16	52.74/55.85/57.04
	p@30	61.27	99.50	63.42/65.43/66.86	73.27/73.51/74.46	59.20/61.96/63.60	59.90/63.25/65.16
eu	p@1	20.19	87.54	3.65/6.87/10.09	7.16/10.50/13.84	2.15/5.32/8.49	2.86/6.92/9.79
	p@5	27.49	95.28	19.76/26.34/29.13	23.63/31.50/33.41	18.10/24.13/27.30	19.57/25.06/27.45
	p@10	29.78	96.85	24.34/30.06/33.21	26.73/34.61/36.52	23.31/28.12/31.80	25.54/29.12/32.94
	p@30	31.93	98.14	30.49/36.08/38.58	33.65/41.77/42.96	29.14/33.64/36.71	30.55/35.08/38.19
fi	p@1	22.91	85.61	6.01/10.52/13.46	9.07/15.99/21.00	4.70/8.18/10.22	4.53/8.59/10.26
	p@5	29.21	92.70	27.77/33.86/36.36	35.32/42.00/43.91	24.54/30.37/33.13	26.01/31.26/34.13
	p@10	31.85	94.56	33.50/39.16/41.59	41.77/47.26/49.40	29.96/35.69/38.24	32.22/37.23/39.62
	p@30	34.72	96.49	42.09/45.74/47.17	49.40/53.70/54.65	38.96/42.33/43.97	42.24/44.87/45.82
ja	p@1	29.42	97.21	12.67/18.18/21.83	14.56/21.00/23.87	11.86/16.97/20.96	13.13/17.90/21.72
	p@5	44.74	99.36	44.02/51.83/55.69	52.51/58.71/62.77	40.39/48.88/52.66	44.87/53.94/57.28
	p@10	51.11	99.43	53.33/60.77/63.85	61.34/68.02/71.60	49.90/57.67/60.53	53.70/61.34/65.16
	p@30	60.49	99.86	65.00/70.87/73.30	72.55/78.28/80.19	61.76/67.69/70.35	64.68/69.69/72.79
th	p@1	2.43	58.98	1.07/1.43/2.15	0.48/1.19/2.63	1.33/1.53/1.94	1.67/1.67/1.91
	p@5	5.44	74.80	7.16/9.88/11.52	8.59/11.69/14.56	6.54/9.10/10.22	7.16/8.59/9.31
	p@10	7.66	79.24	11.81/16.25/17.32	13.60/20.05/20.76	11.04/14.62/15.85	11.22/14.80/15.51
	p@30	12.03	86.54	20.69/25.05/27.27	26.25/30.31/30.55	18.30/22.80/25.87	16.95/21.72/24.34

Table 24: Full results for Llama2-70B with prompt-based embedding