

	EN (P@K: 1000/2000/3000)						
Language	P@K	Before-Alignment	Ceiling Performance	Eval - Both	Eval - Abstract	Eval-Physical (all)	Eval - Physical (downsample)
fr	p@1	60.42	86.97	42.16/48.03/51.40	52.51/57.52/61.58	37.73/43.97/47.03	35.08/43.91/46.30
	p@5	72.23	93.06	66.43/68.29/69.65	73.99/75.89/76.13	63.19/65.03/66.87	65.87/68.26/69.45
	p@10	75.52	95.35	72.01/73.94/74.16	78.04/79.47/80.19	69.43/71.57/71.57	73.51/74.46/74.22
	p@30	79.53	98.00	78.02/79.03/79.67	84.25/84.73/84.96	75.36/76.58/77.40	78.76/80.67/81.62
ro	p@1	43.02	81.53	19.83/29.06/32.71	26.73/37.71/39.62	16.87/25.36/29.75	15.75/24.11/28.64
	p@5	55.05	91.91	47.60/51.40/52.97	57.28/58.95/60.38	43.46/48.16/49.80	44.63/49.16/51.31
	p@10	59.20	94.77	53.04/55.98/56.91	61.81/64.44/64.44	49.28/52.35/53.68	50.84/52.98/55.13
	p@30	63.71	97.64	60.70/62.13/63.28	68.97/71.12/71.84	57.16/58.18/59.61	58.23/59.43/59.67
eu	p@1	15.53	70.65	4.29/8.16/11.60	7.16/10.74/15.99	3.07/7.06/9.71	1.67/6.21/8.83
	p@5	24.77	81.96	16.96/22.91/26.34	24.58/28.88/32.70	13.70/20.35/23.62	12.17/20.29/22.43
	p@10	27.13	84.54	21.83/27.13/30.35	27.92/33.65/36.28	19.22/24.34/27.81	18.38/25.06/27.68
	p@30	30.71	89.33	28.42/32.78/35.08	34.13/38.66/41.29	25.97/30.27/32.41	26.49/31.26/34.13
fi	p@1	25.05	75.81	7.37/11.31/15.10	11.69/15.99/21.72	5.52/9.30/12.27	5.97/9.07/11.46
	p@5	35.22	88.12	26.77/31.50/33.64	35.32/38.42/40.33	23.11/28.53/30.78	25.78/31.26/32.94
	p@10	38.73	91.48	33.57/37.15/39.44	40.33/44.15/47.02	30.67/34.15/36.20	34.37/36.04/38.19
	p@30	44.38	95.13	41.45/45.53/47.03	50.84/55.61/57.52	37.53/41.21/42.54	38.90/43.44/44.15
ja	p@1	26.91	84.18	11.31/16.96/21.83	14.32/20.05/24.34	10.02/15.64/20.76	10.74/17.18/22.67
	p@5	47.75	93.84	38.58/45.74/49.25	47.49/53.70/55.61	34.76/42.33/46.52	37.23/45.11/48.21
	p@10	54.90	96.06	47.67/53.90/56.62	56.09/62.05/64.68	44.07/50.41/53.17	46.78/53.46/56.09
	p@30	62.13	97.85	59.48/64.35/67.07	67.06/71.84/73.75	56.24/61.15/64.21	58.95/64.44/67.30
th	p@1	2.58	38.15	0.57/1.36/1.79	0.48/1.19/2.39	0.61/1.43/1.53	0.72/1.43/1.43
	p@5	6.59	57.19	6.23/8.88/9.38	8.11/13.37/13.13	5.42/6.95/7.77	5.25/7.16/7.40
	p@10	9.74	64.21	10.81/14.17/14.46	13.60/18.14/18.38	9.61/12.47/12.78	9.31/11.22/11.22
	p@30	17.04	75.52	19.83/24.41/25.34	24.11/29.12/30.31	18.00/22.39/23.21	17.90/20.05/21.96

Table 20: Full results for Llama2-7B with prompt-based embedding

	EN (P@K: 1000/2000/3000)						
Language	P@K	Before-Alignment	Ceiling Performance	Eval - Both	Eval - Abstract	Eval-Physical (all)	Eval - Physical (downsample)
fr	p@1	30.78	98.50	15.39/24.84/32.00	19.57/27.92/37.71	13.60/23.52/29.55	11.93/21.48/28.40
	p@5	39.73	99.14	34.72/43.74/47.10	40.10/48.45/52.98	32.41/41.72/44.58	31.98/42.24/45.58
	p@10	43.38	99.28	40.23/48.25/51.90	44.63/54.65/58.47	38.34/45.50/49.08	38.66/45.35/51.31
	p@30	50.82	99.64	49.25/55.69/57.55	53.94/61.10/62.29	47.24/53.37/55.52	49.64/54.89/57.04
ro	p@1	21.12	96.13	4.44/10.45/14.60	3.58/10.26/14.32	4.81/10.53/14.72	4.30/9.79/14.56
	p@5	26.70	97.71	18.68/25.27/28.06	21.72/29.36/32.22	17.38/23.52/26.28	17.90/24.82/27.92
	p@10	30.64	97.92	24.05/30.92/32.86	28.88/36.04/36.52	21.98/28.73/31.29	23.63/30.07/32.22
	p@30	35.86	98.28	32.93/39.08/40.87	37.71/42.72/45.82	30.88/37.53/38.75	33.41/40.81/42.72
eu	p@1	13.67	96.99	1.22/3.58/5.65	0.95/3.10/5.49	1.33/3.78/5.73	0.95/3.34/5.73
	p@5	18.90	98.28	9.31/14.46/16.46	10.74/15.27/17.90	8.69/14.11/15.85	8.59/14.08/14.56
	p@10	20.97	99.07	13.39/18.61/20.97	16.47/21.24/23.87	12.07/17.48/19.73	11.69/16.95/19.57
	p@30	24.41	99.43	19.90/25.84/27.63	22.91/29.59/31.26	18.61/24.23/26.07	18.38/25.54/26.73
fi	p@1	9.52	93.70	0.72/2.65/3.94	0.24/1.91/3.10	0.92/2.97/4.29	0.48/2.39/3.40
	p@5	14.10	96.49	7.73/11.31/13.82	8.83/12.65/15.75	7.26/10.74/12.99	7.40/10.98/13.84
	p@10	17.32	97.28	11.17/15.82/18.75	11.46/18.62/22.67	11.04/14.62/17.08	12.41/16.23/18.14
	p@30	22.48	98.14	20.54/25.63/27.99	24.82/30.31/32.22	18.71/23.62/26.18	20.29/25.30/27.92
ja	p@1	24.19	98.14	1.65/4.58/6.37	1.19/5.49/7.88	1.84/4.19/5.73	2.39/4.77/4.77
	p@5	33.07	98.78	12.53/19.54/23.84	16.47/22.67/28.88	10.84/18.20/21.68	10.98/20.53/22.43
	p@10	38.01	99.28	18.61/24.98/30.85	22.91/27.21/33.65	16.77/24.03/29.65	17.42/26.49/31.74
	p@30	46.46	99.43	27.63/36.79/40.23	30.31/39.62/43.91	26.48/35.58/38.65	28.64/37.47/40.81
th	p@1	0.86	90.19	0.00/0.21/0.36	0.00/0.72/1.19	0.00/0.00/0.00	0.00/0.00/0.00
	p@5	1.43	95.06	3.01/4.29/6.37	2.86/4.06/5.97	3.07/4.40/6.54	3.34/4.77/7.88
	p@10	2.08	95.71	5.37/7.23/9.59	5.49/6.92/9.31	5.32/7.36/9.71	6.44/8.11/10.26
	p@30	3.29	97.14	10.16/13.53/16.54	11.46/13.60/17.66	9.61/13.50/16.05	10.02/13.84/16.23

Table 21: Full results for Llama2-13B with last-token embedding

	EN (P@K: 1000/2000/3000)						
Language	P@K	Before-Alignment	Ceiling Performance	Eval - Both	Eval - Abstract	Eval-Physical (all)	Eval - Physical (downsample)
fr	p@1	59.27	87.47	44.95/50.68/53.61	52.74/60.62/63.48	41.62/46.42/49.39	41.29/47.49/50.12
	p@5	72.01	95.71	68.50/71.08/71.87	77.09/78.04/79.00	64.83/68.10/68.81	67.06/70.17/69.45
	p@10	74.95	97.35	72.58/74.30/75.38	79.95/81.38/82.82	69.43/71.27/72.19	71.12/72.55/73.51
	p@30	79.24	98.35	78.45/79.17/79.53	84.25/84.96/84.96	75.97/76.69/77.20	77.33/78.04/79.24
ro	p@1	38.87	82.39	24.98/34.65/36.65	31.50/44.39/46.06	22.19/30.47/32.62	22.91/29.12/33.41
	p@5	52.97	93.41	51.40/54.76/56.76	60.38/63.25/63.96	47.55/51.12/53.68	49.40/52.51/54.65
	p@10	56.55	95.35	56.41/58.98/60.34	64.92/67.30/67.54	52.76/55.42/57.26	55.61/55.85/58.23
	p@30	62.99	97.35	63.06/65.50/66.14	70.17/73.03/73.27	60.02/62.27/63.09	62.29/63.72/64.68
eu	p@1	20.26	75.30	6.08/10.95/15.75	6.21/11.22/17.42	6.03/10.84/15.03	5.73/10.74/14.08
	p@5	29.49	86.83	20.40/28.06/31.78	23.15/31.26/33.89	19.22/26.69/30.88	17.90/26.25/29.59
	p@10	31.64	89.84	24.41/31.93/35.15	27.21/34.84/36.75	23.21/30.67/34.46	22.67/30.55/33.89
	p@30	35.72	92.84	30.42/37.37/39.44	31.74/38.66/40.33	29.86/36.81/39.06	30.55/38.19/39.14
fi	p@1	23.91	79.96	9.81/15.03/18.68	12.65/17.66/21.00	8.59/13.91/17.69	7.88/14.56/18.62
	p@5	33.86	90.91	30.06/34.22/37.01	33.41/39.86/40.81	28.63/31.80/35.48	28.64/32.70/36.99
	p@10	36.79	92.84	34.86/39.66/41.88	40.81/45.58/48.45	32.31/37.12/39.06	33.89/36.52/39.62
	p@30	41.45	95.85	41.80/46.39/47.89	49.16/53.46/55.13	38.65/43.35/44.79	39.62/43.20/45.35
ja	p@1	15.60	88.05	12.74/19.69/23.48	13.13/20.76/26.01	12.58/19.22/22.39	12.65/19.81/23.39
	p@5	28.35	96.99	43.38/51.40/53.47	47.49/58.71/60.62	41.62/48.26/50.41	43.91/51.07/54.42
	p@10	36.29	98.21	51.61/58.34/61.42	57.52/65.87/68.26	49.08/55.11/58.49	52.03/58.00/61.10
	p@30	47.60	99.28	62.71/68.29/70.29	70.41/76.61/77.09	59.41/64.72/67.38	63.01/66.59/69.45
th	p@1	1.36	46.39	0.79/1.57/1.93	1.43/1.43/2.15	0.51/1.64/1.84	0.24/1.91/1.91
	p@5	3.51	63.21	5.65/7.73/9.52	7.40/10.74/12.65	4.91/6.44/8.18	4.77/5.25/7.40
	p@10	5.44	70.01	8.88/12.53/13.60	10.26/15.99/16.95	8.28/11.04/12.17	7.88/8.83/10.74
	p@30	8.52	77.88	15.75/19.40/21.83	18.14/21.00/24.34	14.72/18.71/20.76	14.56/17.90/19.09

Table 22: Full results for Llama2-13B with prompt-based embedding