| | | 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	42.09	68.65	36.22/39.44/40.59	42.96/46.30/48.69	33.33/36.50/37.12	33.65/38.19/38.42
	P@5	50.54	76.23	47.67/49.82/51.54	55.13/57.04/59.19	44.48/46.73/48.26	48.69/50.84/52.27
	P@10	53.76	78.6	50.68/53.47/54.69	57.52/60.38/61.10	47.75/50.51/51.94	51.31/54.18/56.09
	P@30	58.2	81.96	56.48/58.12/58.77	62.29/65.16/65.63	53.99/55.11/55.83	57.52/58.95/59.90
ro	P@1	34.86	72.3	27.63/31.42/33.21	31.03/35.56/37.95	26.18/29.65/31.19	28.16/30.79/32.94
	P@5	44.74	77.81	40.23/43.31/45.60	44.87/48.69/50.12	38.24/41.00/43.66	42.24/45.11/46.78
	P@10	47.96	80.89	44.52/47.60/48.89	49.40/53.46/55.13	42.43/45.09/46.22	46.30/47.97/49.16
	P@30	53.11	84.4	49.82/52.97/53.40	55.37/58.23/58.71	47.44/50.72/51.12	51.07/52.51/53.22
eu	P@1	24.62	71.22	15.68/19.76/22.19	15.75/23.39/27.68	15.64/18.20/19.84	16.23/19.09/21.00
	P@5	32.57	77.52	27.92/32.00/33.64	34.61/40.10/42.72	25.05/28.53/29.75	27.45/32.46/31.98
	P@10	36.15	79.81	32.64/35.36/36.86	40.81/44.15/45.82	29.14/31.60/33.03	32.46/34.13/35.80
	P@30	40.09	83.75	38.80/40.87/41.80	47.73/49.88/51.31	34.97/37.01/37.73	38.66/40.57/42.00
fi	P@1	26.27	71.37	15.89/20.90/22.33	17.42/23.87/25.54	15.24/19.63/20.96	14.56/19.81/21.24
	P@5	34.5	78.31	31.07/33.72/35.72	33.65/36.99/40.10	29.96/32.31/33.84	32.70/35.56/37.47
	P@10	37.51	80.17	35.58/37.87/39.44	39.14/40.33/43.44	34.05/36.81/37.73	37.23/40.81/41.29
	P@30	42.73	83.75	41.73/44.31/45.38	46.54/48.21/49.64	39.67/42.64/43.56	42.48/45.11/45.58
ja	P@1	28.49	80.17	13.31/20.11/23.26	11.22/19.33/24.11	14.21/20.45/22.90	17.18/22.20/25.30
	P@5	41.16	88.83	33.50/40.37/43.38	32.46/38.19/40.81	33.95/41.31/44.48	34.84/46.30/49.64
	P@10	46.46	91.2	40.23/45.96/48.39	37.71/42.48/46.06	41.31/47.44/49.39	43.44/52.27/54.18
	P@30	54.62	94.27	50.47/54.69/55.98	48.69/54.42/55.85	51.23/54.81/56.03	55.13/59.43/59.19
th	P@1	25.91	79.1	12.03/16.18/19.90	11.22/15.99/20.29	12.37/16.26/19.73	11.93/18.62/22.20
	P@5	41.95	87.54	36.58/40.52/43.02	36.28/41.29/44.39	36.71/40.18/42.43	39.14/43.91/46.06
	P@10	47.24	90.26	43.66/47.46/49.75	43.68/49.16/51.31	43.66/46.73/49.08	46.30/49.40/51.31
	P@30	56.41	93.27	54.47/57.98/59.70	56.80/61.34/61.58	53.48/56.54/58.90	56.09/58.47/62.29

Table 17: Full results for mT0-xxl (13B) with average 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	29.71	56.62	47.82/49.25/49.96	58.00/59.90/59.90	43.46/44.68/45.71	45.11/45.58/46.78
	P@5	47.96	64.85	58.34/59.34/59.77	68.26/69.69/69.93	54.09/54.91/55.42	55.85/57.04/57.76
	P@10	54.9	67.86	62.49/62.56/62.63	71.12/71.84/71.84	58.79/58.59/58.69	60.14/60.14/60.14
	P@30	62.92	72.73	68.36/68.15/68.36	77.09/77.09/77.80	64.62/64.31/64.31	67.06/67.06/66.83
ro	P@1	9.59	53.4	38.08/40.59/41.16	45.82/48.93/49.88	34.76/37.01/37.42	38.19/40.81/41.29
	P@5	22.48	63.92	51.04/52.40/53.54	59.90/61.81/62.77	47.24/48.36/49.59	50.84/52.27/53.70
	P@10	31.07	67.64	55.76/57.84/58.20	63.72/65.63/65.87	52.35/54.50/54.91	55.37/57.76/59.67
	P@30	45.03	72.37	63.28/64.07/64.64	70.88/71.36/71.84	60.02/60.94/61.55	63.25/64.44/65.39
eu	P@1	1.93	16.68	2.79/4.65/5.73	4.06/6.92/7.88	2.25/3.68/4.81	3.34/5.25/5.73
	P@5	5.73	26.63	8.09/12.10/13.74	11.93/16.23/17.66	6.44/10.33/12.07	8.59/11.93/13.84
	P@10	8.66	30.78	12.46/16.18/18.32	16.47/22.20/23.87	10.74/13.60/15.95	13.13/14.80/16.95
	P@30	15.1	39.08	20.76/24.84/26.49	28.16/34.37/35.32	17.59/20.76/22.70	19.33/22.43/24.82
fi	P@1	5.15	50.47	31.78/33.29/34.93	36.52/37.95/38.90	29.75/31.29/33.23	32.46/35.08/37.23
	P@5	15.46	60.34	45.88/47.46/48.82	53.94/54.65/56.32	42.43/44.38/45.60	46.30/47.97/49.16
	P@10	22.48	63.92	50.11/52.68/53.61	58.95/61.34/61.10	46.32/48.98/50.41	50.60/52.51/53.94
	P@30	36.72	71.73	58.41/59.84/59.99	67.30/68.50/68.74	54.60/56.13/56.24	59.90/59.67/59.90
ja	P@1	6.73	48.25	31.28/32.36/33.36	36.28/36.75/38.42	29.14/30.47/31.19	32.94/35.08/36.28
	P@5	20.33	60.92	48.39/49.46/50.25	52.27/53.46/54.18	46.73/47.75/48.57	51.31/52.51/53.70
	P@10	28.99	65.07	53.47/54.97/55.91	57.28/58.71/59.90	51.84/53.37/54.19	55.85/56.80/57.76
	P@30	43.31	71.58	61.78/62.71/62.42	63.72/65.16/64.44	60.94/61.66/61.55	63.48/64.68/64.68
th	P@1	6.3	48.1	26.70/28.85/29.06	32.22/32.70/34.37	24.34/27.20/26.79	25.06/29.36/28.64
	P@5	16.68	61.56	43.74/46.03/45.96	50.12/51.79/51.31	41.00/43.56/43.66	43.91/47.02/46.78
	P@10	25.05	67.29	50.18/52.83/53.26	56.09/57.52/57.52	47.65/50.82/51.43	50.36/53.22/53.22
	P@30	38.51	75.3	61.99/63.14/63.71	67.06/68.50/68.74	59.82/60.84/61.55	60.86/61.10/61.34

Table 18: Full results for mT0-xxl (13B) 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	25.41	87.62	19.18/27.56/30.64	22.20/32.22/36.04	17.89/25.56/28.32	17.66/25.06/28.40
	p@5	32.28	92.13	33.00/39.30/42.30	37.71/46.78/49.16	30.98/36.09/39.37	32.94/36.75/40.10
	p@10	35.65	93.92	36.72/43.52/45.96	42.48/49.64/53.46	34.25/40.90/42.74	36.04/42.00/44.39
	p@30	42.95	95.71	44.45/49.53/51.40	51.07/56.56/57.52	41.62/46.52/48.77	43.44/47.97/50.84
ro	p@1	16.39	84.40	6.94/12.53/15.53	6.44/11.46/15.04	7.16/12.99/15.75	7.40/13.13/16.47
	p@5	22.05	88.40	19.11/23.98/26.84	21.24/27.21/30.79	18.20/22.60/25.15	19.57/25.30/27.68
	p@10	24.70	89.41	23.62/27.49/30.21	26.97/30.79/34.61	22.19/26.07/28.32	24.58/29.36/30.79
	p@30	28.85	91.34	30.21/34.50/36.29	36.52/40.10/41.53	27.51/32.11/34.05	31.03/34.37/36.52
eu	p@1	10.45	80.89	1.93/4.08/6.73	2.86/4.30/7.40	1.53/3.99/6.44	0.24/2.15/5.49
	p@5	13.82	87.19	9.74/13.39/15.32	11.46/13.60/16.23	9.00/13.29/14.93	7.64/12.17/14.08
	p@10	15.75	88.98	12.81/16.61/18.83	15.04/17.90/20.76	11.86/16.05/18.00	11.46/15.27/17.18
	p@30	18.83	91.41	17.61/22.41/22.91	20.53/23.87/24.11	16.36/21.78/22.39	16.71/21.96/21.48
fi	p@1	7.73	74.37	0.93/2.65/4.29	0.24/1.91/5.01	1.23/2.97/3.99	1.43/3.82/4.77
	p@5	10.67	82.18	7.37/10.24/12.03	9.55/13.13/15.51	6.44/9.00/10.53	6.68/10.26/12.17
	p@10	12.60	84.32	11.31/14.75/16.39	13.37/18.62/21.24	10.43/13.09/14.31	12.17/14.08/15.51
	p@30	16.96	87.40	18.11/21.33/23.19	24.11/26.25/28.16	15.54/19.22/21.06	17.42/20.05/22.43
ja	p@1	10.09	89.62	1.93/3.51/5.37	3.10/5.73/7.88	1.43/2.56/4.29	1.43/2.63/5.01
	p@5	19.90	94.27	8.80/14.60/17.04	11.69/18.14/20.05	7.57/13.09/15.75	8.35/14.80/16.23
	p@10	24.12	95.78	12.88/19.90/23.48	16.23/23.63/29.36	11.45/18.30/20.96	12.89/20.05/23.15
	p@30	31.28	96.78	22.41/29.78/33.64	27.92/36.52/38.42	20.04/26.89/31.60	21.72/29.83/33.65
th	p@1	0.79	71.37	0.21/0.43/0.29	0.24/0.72/0.48	0.20/0.31/0.20	0.48/0.48/0.24
	p@5	1.72	79.46	1.50/2.51/3.44	2.39/2.86/4.30	1.12/2.35/3.07	1.43/3.10/3.10
	p@10	2.51	81.10	3.51/5.08/6.59	3.82/5.25/8.11	3.37/5.01/5.93	3.58/5.97/6.44
	p@30	4.08	84.54	8.23/10.67/12.31	8.83/12.17/14.80	7.98/10.02/11.25	8.11/10.02/11.46

Table 19: Full results for Llama2-7B with last-token embedding