| | | 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	17.61	34.5	16.82/19.83/21.19	21.96/25.30/25.54	14.62/17.48/19.33	13.60/16.71/17.42
	P@5	20.33	43.24	25.98/28.35/30.21	31.03/34.61/36.28	23.82/25.66/27.61	23.39/25.06/26.73
	P@10	21.26	47.03	29.92/32.21/34.86	35.56/38.19/40.33	27.51/29.65/32.52	26.97/28.88/32.22
	P@30	23.77	51.9	37.51/40.80/41.45	43.68/46.54/47.49	34.87/38.34/38.85	35.08/38.90/39.86
ro	P@1	13.24	19.11	3.79/5.65/7.09	4.06/4.77/5.97	3.68/6.03/7.57	4.06/7.16/8.59
	P@5	14.96	25.2	8.38/12.24/13.53	8.35/10.98/11.69	8.38/12.78/14.31	9.07/14.56/15.75
	P@10	16.39	28.49	11.74/14.60/15.53	10.98/13.84/13.60	12.07/14.93/16.36	13.13/15.99/17.42
	P@30	18.18	33.43	16.82/19.90/21.05	16.47/17.90/19.09	16.97/20.76/21.88	17.90/21.48/21.96
eu	P@1	7.87	17.32	3.65/5.30/6.01	3.82/5.97/6.92	3.58/5.01/5.62	4.06/5.25/5.97
	P@5	9.16	24.41	7.59/9.88/10.52	8.35/11.22/11.69	7.26/9.30/10.02	7.16/10.02/10.98
	P@10	10.16	26.77	10.67/12.88/13.53	11.46/14.56/15.04	10.33/12.17/12.88	10.02/12.65/13.84
	P@30	12.24	31.64	16.54/18.68/19.69	17.42/20.76/21.00	16.16/17.79/19.12	14.56/16.47/17.90
fi	P@1	4.22	9.66	0.64/0.86/1.15	0.72/0.72/1.67	0.61/0.92/0.92	0.48/0.48/0.72
	P@5	4.8	14.03	2.15/3.15/3.94	2.86/4.06/5.49	1.84/2.76/3.27	1.19/1.67/2.86
	P@10	5.37	16.03	3.72/4.65/5.30	5.01/5.49/6.92	3.17/4.29/4.60	2.39/3.34/4.30
	P@30	6.73	21.05	6.80/8.09/9.09	7.16/10.26/10.50	6.65/7.16/8.49	6.44/6.92/9.07
ja	P@1	0.29	9.88	0.57/0.64/0.93	0.72/1.43/1.43	0.51/0.31/0.72	0.24/0.24/0.24
	P@5	0.72	16.25	2.08/2.51/2.86	3.58/3.82/4.30	1.43/1.94/2.25	1.19/2.39/1.91
	P@10	1.15	18.75	3.36/4.08/4.94	4.30/5.97/6.92	2.97/3.27/4.09	2.39/2.86/3.58
	P@30	2	24.84	6.59/8.59/8.59	9.07/13.37/13.37	5.52/6.54/6.54	4.53/6.68/6.21
th	P@1	0.07	11.1	0.36/0.29/0.43	0.72/0.48/0.72	0.20/0.20/0.31	0.00/0.24/0.24
	P@5	0.14	16.25	0.93/1.43/1.22	1.43/2.15/1.91	0.72/1.12/0.92	0.72/1.19/0.72
	P@10	0.43	18.9	1.43/1.86/1.86	2.15/3.10/2.86	1.12/1.33/1.43	1.43/1.43/1.19
	P@30	1.36	24.34	3.51/3.08/3.65	4.53/5.01/5.25	3.07/2.25/2.97	3.34/2.15/2.86

Table 11: Full results for BLOOMZ-7B1 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	2.58	61.85	51.11/53.83/55.62	60.86/63.01/64.92	46.93/49.90/51.64	47.49/51.55/52.51
	P@5	7.95	71.58	64.57/66.21/66.86	73.51/74.46/74.70	60.74/62.68/63.50	63.25/65.16/66.35
	P@10	14.96	74.45	67.64/69.08/69.51	75.89/76.13/76.37	64.11/66.05/66.56	66.59/69.69/69.21
	P@30	29.56	79.31	73.09/73.44/73.87	79.95/79.47/79.71	70.14/70.86/71.37	72.55/73.27/74.46
ro	P@1	1.65	36.58	21.76/25.34/28.13	28.64/31.50/33.41	18.81/22.70/25.87	20.05/24.82/27.45
	P@5	3.79	43.88	31.50/34.65/36.15	38.19/40.81/40.81	28.63/32.00/34.15	30.31/34.13/35.80
	P@10	5.3	46.53	35.50/38.01/39.73	41.77/42.96/44.39	32.82/35.89/37.73	34.37/38.42/39.38
	P@30	10.45	50.32	41.16/42.59/43.88	45.82/47.26/49.16	39.16/40.59/41.62	41.29/43.44/43.20
eu	P@1	1.93	30.78	18.54/20.26/21.40	24.11/27.68/27.92	16.16/17.08/18.61	16.95/17.18/18.62
	P@5	4.58	40.52	28.78/30.35/30.78	35.32/37.23/38.19	25.97/27.40/27.61	25.30/26.73/26.97
	P@10	6.59	43.02	32.57/33.00/34.36	39.38/40.57/42.72	29.65/29.75/30.78	28.88/28.16/29.12
	P@30	12.46	49.39	36.94/38.01/39.01	43.91/45.82/47.26	33.95/34.66/35.48	32.22/33.89/35.32
fi	P@1	0.86	15.1	5.65/7.30/7.80	8.59/10.74/10.74	4.40/5.83/6.54	4.77/5.73/6.44
	P@5	2.65	20.26	10.02/12.67/13.46	13.60/15.99/16.23	8.49/11.25/12.27	9.07/11.69/11.93
	P@10	3.44	22.98	12.10/14.60/15.60	15.27/17.66/18.14	10.74/13.29/14.52	11.93/13.37/14.80
	P@30	6.01	27.13	15.60/17.90/18.61	19.09/21.00/21.24	14.11/16.56/17.48	15.27/16.23/16.95
ja	P@1	0.5	14.32	4.01/4.58/5.08	7.40/8.83/10.26	2.56/2.76/2.86	3.34/4.30/4.53
	P@5	2.43	23.34	8.09/10.38/11.60	14.32/18.62/20.05	5.42/6.85/7.98	7.88/8.83/10.02
	P@10	3.44	27.92	11.45/14.60/15.68	18.85/23.39/26.01	8.28/10.84/11.25	9.79/12.89/13.37
	P@30	7.8	35.79	17.04/20.54/22.69	25.78/31.50/33.41	13.29/15.85/18.10	15.27/17.66/19.81
th	P@1	0.14	4.44	0.50/0.50/0.50	0.48/0.95/1.19	0.51/0.31/0.20	0.48/0.00/0.00
	P@5	0.79	10.67	1.65/1.72/2.22	1.91/2.39/3.58	1.53/1.43/1.64	1.91/1.19/0.95
	P@10	1.22	14.03	3.01/2.86/3.72	4.06/4.30/5.73	2.56/2.25/2.86	2.39/1.43/2.15
	P@30	2.65	22.12	6.23/6.87/7.02	7.64/9.79/9.55	5.62/5.62/5.93	5.97/5.73/5.01

Table 12: Full results for BLOOMZ-7B1 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	35.86	49.11	33.57/35.22/36.44	38.66/40.81/41.29	31.39/32.82/34.36	31.26/33.89/35.08
	P@5	45.45	60.27	43.45/45.03/46.24	50.12/52.03/52.03	40.59/42.02/43.76	42.96/44.39/46.78
	P@10	48.46	63.49	46.89/49.11/49.82	52.98/55.85/56.32	44.27/46.22/47.03	47.49/49.88/51.79
	P@30	53.33	69.51	52.47/53.83/54.90	59.19/59.90/61.34	49.59/51.23/52.15	53.70/56.09/56.32
ro	P@1	31.35	47.24	27.63/30.57/32.00	31.98/35.32/37.95	25.77/28.53/29.45	26.97/30.07/30.31
	P@5	37.72	59.41	36.22/38.80/39.51	42.00/45.35/45.82	33.74/35.99/36.81	36.52/38.90/40.10
	P@10	40.87	63.56	39.01/41.02/42.02	45.35/47.02/47.97	36.30/38.45/39.47	38.42/41.05/42.72
	P@30	45.17	70.15	44.60/46.39/47.96	50.36/51.55/53.22	42.13/44.17/45.71	43.91/46.78/48.21
eu	P@1	19.47	44.09	15.39/18.11/19.90	16.47/19.33/22.67	14.93/17.59/18.71	15.75/18.85/20.29
	P@5	27.42	56.91	22.91/25.48/27.27	24.58/29.12/31.03	22.19/23.93/25.66	23.15/25.54/27.68
	P@10	30.06	61.85	26.91/29.78/30.49	31.03/33.65/36.04	25.15/28.12/28.12	26.49/29.83/29.59
	P@30	35	69.01	33.36/35.29/36.44	38.66/41.53/41.77	31.08/32.62/34.15	32.46/34.13/35.32
fi	P@1	16.96	42.45	12.53/15.39/17.18	12.41/15.51/16.95	12.58/15.34/17.28	14.08/14.80/17.90
	P@5	24.12	54.04	20.90/23.84/25.13	21.48/24.58/26.25	20.65/23.52/24.64	22.91/25.54/26.97
	P@10	27.34	57.62	23.84/27.42/28.63	25.54/29.83/30.07	23.11/26.38/28.02	26.25/28.64/31.03
	P@30	33.07	65.21	30.64/33.29/35.00	33.41/35.80/37.23	29.45/32.21/34.05	32.46/34.61/37.23
ja	P@1	14.39	53.4	9.59/13.96/15.60	8.11/11.46/12.89	10.22/15.03/16.77	12.41/19.33/18.85
	P@5	25.41	66.43	21.26/27.56/29.42	19.33/24.82/26.01	22.09/28.73/30.88	25.54/32.70/36.28
	P@10	30.06	71.08	27.42/33.57/35.79	25.30/31.74/33.17	28.32/34.36/36.91	31.50/39.38/42.24
	P@30	39.51	77.24	37.80/43.66/45.17	36.99/42.48/44.39	38.14/44.17/45.50	42.96/50.12/50.84
th	P@1	11.88	52.97	8.38/10.52/13.10	8.83/11.93/13.60	8.18/9.92/12.88	8.11/10.74/13.84
	P@5	22.33	66.93	20.40/24.48/27.70	20.53/23.63/25.54	20.35/24.85/28.63	21.96/27.21/30.55
	P@10	27.49	70.79	25.98/31.21/33.50	26.73/30.79/33.17	25.66/31.39/33.64	26.97/33.41/37.47
	P@30	36.44	80.31	36.86/41.52/43.24	35.56/40.81/43.91	37.42/41.82/42.94	39.62/43.68/45.58

Table 13: Full results for mT0-large (1.2B) with average embedding