

	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	24.55	48.17	42.30/44.38/44.81	46.06/49.64/50.12	40.70/42.13/42.54	40.81/42.48/42.72
	P@5	36.72	58.12	51.25/52.76/53.61	58.71/60.62/62.29	48.06/49.39/49.90	50.12/51.07/51.07
	P@10	43.02	61.27	55.12/56.34/57.55	63.25/64.44/65.16	51.64/52.86/54.29	53.22/54.65/56.09
	P@30	52.11	65.78	60.70/61.99/62.49	67.30/68.26/68.26	57.87/59.30/60.02	59.19/60.62/61.81
ro	P@1	13.89	28.2	18.68/20.33/21.40	19.57/21.24/22.43	18.30/19.94/20.96	19.09/20.53/21.48
	P@5	20.62	35.93	26.34/28.27/29.35	28.40/30.79/31.50	25.46/27.20/28.43	25.54/27.92/28.40
	P@10	23.48	38.51	29.35/30.99/32.14	31.74/32.22/33.17	28.32/30.47/31.70	29.36/31.50/32.46
	P@30	27.7	42.66	33.57/36.08/37.01	35.56/36.75/37.23	32.72/35.79/36.91	33.41/36.28/36.99
eu	P@1	3.01	12.81	5.73/7.16/8.16	6.44/7.16/9.07	5.42/7.16/7.77	4.06/5.25/5.73
	P@5	6.16	17.9	11.38/13.03/13.67	12.89/14.56/15.27	10.74/12.37/12.99	9.07/11.69/11.93
	P@10	7.59	19.83	13.39/15.46/15.96	14.80/16.95/18.14	12.78/14.83/15.03	10.98/14.80/14.80
	P@30	10.67	25.27	17.54/18.90/19.47	19.09/21.24/22.67	16.87/17.89/18.10	15.51/16.23/16.71
fi	P@1	6.87	11.31	3.94/5.80/6.87	4.53/6.68/7.64	3.68/5.42/6.54	5.01/5.97/6.68
	P@5	10.31	16.03	7.59/9.59/10.95	8.59/10.50/11.69	7.16/9.20/10.63	9.07/10.02/10.74
	P@10	11.81	18.04	9.52/12.17/12.81	10.50/12.65/12.41	9.10/11.96/12.99	10.98/11.93/12.65
	P@30	13.24	22.48	13.17/15.39/15.96	13.84/15.27/15.27	12.88/15.44/16.26	14.80/15.51/16.47
ja	P@1	0.07	11.1	3.87/4.80/5.73	7.16/8.59/9.79	2.45/3.17/3.99	2.15/3.58/4.53
	P@5	0.5	19.26	8.45/10.31/11.60	12.89/15.75/15.99	6.54/7.98/9.71	6.68/8.59/10.50
	P@10	0.86	22.69	11.17/13.82/14.60	15.27/19.33/19.57	9.41/11.45/12.47	9.79/12.17/12.65
	P@30	2.29	30.49	18.32/19.90/20.83	22.67/26.25/26.25	16.46/17.18/18.51	16.71/16.47/17.18
th	P@1	0	2.65	0.21/0.21/0.50	0.48/0.48/0.95	0.10/0.10/0.31	0.24/0.24/0.48
	P@5	0	5.37	0.93/1.43/1.65	0.95/2.15/2.86	0.92/1.12/1.12	1.19/1.91/1.19
	P@10	0.14	7.8	1.65/2.79/2.93	2.39/4.06/4.06	1.33/2.25/2.45	1.67/2.86/3.10
	P@30	1.43	12.31	4.58/5.08/5.87	6.44/6.92/7.64	3.78/4.29/5.11	3.82/5.25/5.73

Table 8: Full results for BLOOMZ-1B7 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	22.91	36.94	21.69/25.84/27.27	24.34/30.31/31.26	20.55/23.93/25.56	17.42/21.24/22.91
	P@5	26.63	44.52	30.92/34.29/35.72	33.89/39.38/40.57	29.65/32.11/33.64	26.49/28.64/30.55
	P@10	29.56	47.67	34.50/37.80/39.37	39.14/43.44/45.35	32.52/35.38/36.81	29.83/33.41/34.84
	P@30	35.43	53.19	40.09/42.88/44.09	46.78/51.07/51.31	37.22/39.37/41.00	35.80/38.66/41.29
ro	P@1	14.17	16.68	4.65/6.37/7.30	4.30/6.44/7.64	4.81/6.34/7.16	6.44/7.88/8.59
	P@5	16.61	21.98	9.74/11.67/12.60	10.74/12.17/12.65	9.30/11.45/12.58	11.22/12.89/14.08
	P@10	17.32	24.27	12.03/13.82/14.96	12.89/15.04/14.80	11.66/13.29/15.03	12.89/14.56/16.23
	P@30	19.33	28.85	17.04/18.47/19.69	17.66/18.85/20.29	16.77/18.30/19.43	17.90/20.29/21.00
eu	P@1	11.38	16.68	2.43/4.80/5.73	3.82/6.92/7.16	1.84/3.89/5.11	1.91/4.53/6.21
	P@5	14.46	22.76	7.95/10.45/12.60	10.50/12.65/14.56	6.85/9.51/11.76	7.40/10.74/12.65
	P@10	16.46	24.62	10.16/13.53/14.39	13.60/15.51/17.18	8.79/12.68/13.19	9.55/12.89/14.08
	P@30	18.9	29.85	15.60/17.11/18.75	19.33/19.33/21.48	14.01/16.16/17.59	14.08/16.47/18.14
fi	P@1	4.94	8.09	0.72/1.07/1.65	1.91/2.15/2.15	0.20/0.61/1.43	0.24/0.48/1.19
	P@5	6.37	12.6	2.65/3.51/4.08	4.30/5.25/5.49	1.94/2.76/3.48	1.67/2.39/3.10
	P@10	7.23	14.46	3.87/4.65/5.58	5.97/6.92/7.64	2.97/3.68/4.70	2.63/3.58/4.30
	P@30	8.66	18.75	6.87/6.87/8.30	9.79/10.02/11.69	5.62/5.52/6.85	5.73/5.97/7.16
ja	P@1	0.21	5.87	0.43/0.79/0.72	0.72/0.72/0.48	0.31/0.82/0.82	0.24/0.72/0.48
	P@5	0.93	10.31	1.72/2.00/2.15	2.39/2.63/2.63	1.43/1.74/1.94	0.72/0.95/1.43
	P@10	1.36	13.53	2.79/3.44/3.72	3.58/5.25/5.97	2.45/2.66/2.76	2.39/2.86/2.86
	P@30	2.15	19.11	7.09/7.44/8.38	8.11/10.98/12.41	6.65/5.93/6.65	6.68/6.21/7.64
th	P@1	0	5.01	0.14/0.21/0.29	0.24/0.24/0.48	0.10/0.20/0.20	0.00/0.24/0.24
	P@5	0.29	8.16	0.72/1.07/0.93	0.72/1.43/1.67	0.72/0.92/0.61	0.48/1.19/0.95
	P@10	0.29	10.52	1.15/1.57/1.72	1.19/2.63/2.86	1.12/1.12/1.23	0.95/1.43/1.43
	P@30	0.72	15.39	2.58/3.44/3.79	3.34/6.21/6.21	2.25/2.25/2.76	1.67/2.15/2.39

Table 9: Full results for BLOOMZ-3B 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	19.33	58.48	52.18/53.33/53.69	59.90/60.38/60.38	48.88/50.31/50.82	48.93/50.36/50.60
	P@5	34.93	69.29	63.28/64.07/64.71	71.36/72.55/73.03	59.82/60.43/61.15	61.34/62.29/62.77
	P@10	41.95	72.15	66.28/67.43/67.86	73.75/75.42/76.13	63.09/64.01/64.31	65.87/66.11/66.59
	P@30	55.48	77.38	71.22/72.08/73.09	78.76/79.24/79.95	68.00/69.02/70.14	70.41/71.36/72.79
ro	P@1	2.43	33.43	21.33/23.91/26.27	24.58/26.97/28.64	19.94/22.60/25.26	21.72/22.91/25.06
	P@5	6.59	41.95	30.99/33.07/35.36	34.84/36.99/37.47	29.35/31.39/34.46	31.26/32.46/35.56
	P@10	9.66	45.24	34.22/36.29/37.65	37.71/38.90/39.86	32.72/35.17/36.71	35.32/37.71/37.23
	P@30	15.39	49.75	40.73/43.02/43.52	44.15/44.15/46.06	39.26/42.54/42.43	40.81/43.91/43.68
eu	P@1	1.07	18.75	8.95/11.17/11.24	10.02/14.32/14.32	8.49/9.82/9.92	8.11/10.02/10.26
	P@5	3.79	26.34	15.25/18.25/19.61	18.62/23.63/25.06	13.80/15.95/17.28	14.56/16.47/17.66
	P@10	5.08	29.06	18.18/21.62/22.55	21.72/26.49/27.45	16.67/19.53/20.45	18.14/18.85/20.76
	P@30	9.23	34.86	23.62/26.41/27.27	28.16/30.31/31.26	21.68/24.74/25.56	23.63/24.82/25.78
fi	P@1	1.22	13.67	5.58/7.09/8.23	7.40/9.31/11.22	4.81/6.13/6.95	5.25/6.21/7.16
	P@5	2.51	19.18	9.81/11.31/12.53	11.69/14.08/14.56	9.00/10.12/11.66	10.02/10.26/11.69
	P@10	3.36	21.69	11.38/13.60/14.82	12.89/15.99/16.71	10.74/12.58/14.01	10.98/12.89/14.56
	P@30	6.37	25.34	13.96/16.68/18.32	15.99/18.38/20.05	13.09/15.95/17.59	13.13/16.95/19.57
ja	P@1	0.79	13.17	3.29/4.65/5.73	6.21/9.07/11.22	2.04/2.76/3.37	2.39/2.86/3.10
	P@5	2.86	22.98	11.31/12.38/14.10	17.90/19.81/22.43	8.49/9.20/10.53	9.07/9.07/9.79
	P@10	4.44	26.7	14.46/16.03/17.68	21.24/23.87/27.68	11.55/12.68/13.39	11.22/12.41/12.89
	P@30	9.09	34.22	20.76/23.19/24.19	28.88/34.37/35.08	17.28/18.40/19.53	16.71/16.71/19.33
th	P@1	0.14	2.43	0.29/0.07/0.29	0.24/0.00/0.72	0.31/0.10/0.10	0.24/0.24/0.24
	P@5	0.43	6.37	1.29/1.22/0.86	1.67/1.43/0.95	1.12/1.12/0.82	0.95/0.95/0.48
	P@10	0.64	9.74	1.57/1.72/1.65	1.91/2.63/2.39	1.43/1.33/1.33	1.43/1.19/0.95
	P@30	1.86	15.32	2.93/4.44/5.08	3.34/7.16/7.40	2.76/3.27/4.09	3.34/3.34/4.06

Table 10: Full results for BLOOMZ-3B with prompt-based embedding