

	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	11.17	26.06	11.38/14.60/15.60	14.32/16.47/17.90	9.41/12.37/12.68	9.31/10.74/11.46
	P@5	17.18	33.93	19.83/22.69/24.41	23.15/27.92/30.31	17.18/20.25/20.65	16.47/19.33/19.09
	P@10	20.11	37.94	23.48/26.91/27.99	27.45/31.50/36.04	21.47/23.52/24.13	21.00/23.15/23.15
	P@30	25.2	44.45	29.06/33.14/34.43	36.04/39.62/42.48	25.97/29.86/31.08	25.06/28.88/30.79
ro	P@1	6.16	14.53	3.58/4.15/4.87	4.53/7.40/6.44	2.86/2.66/3.68	3.10/2.86/3.34
	P@5	11.02	24.12	7.44/9.59/10.81	10.02/14.08/15.51	5.83/6.75/7.77	6.21/7.40/6.92
	P@10	13.6	28.63	10.45/12.88/13.82	15.27/18.38/20.53	8.28/9.41/10.74	9.31/10.50/10.50
	P@30	18.54	36.36	16.54/18.40/20.33	24.34/27.45/29.12	13.50/14.21/15.85	15.51/16.47/16.23
eu	P@1	1.29	7.09	0.64/0.79/1.43	0.72/1.19/1.43	0.41/0.72/1.12	0.24/0.72/0.48
	P@5	2.93	13.96	2.15/3.01/3.65	3.58/4.53/5.01	1.74/2.35/2.86	1.19/2.39/3.10
	P@10	4.15	17.11	3.08/4.29/5.65	4.77/6.21/8.11	2.35/3.68/3.89	2.15/4.53/4.77
	P@30	7.52	23.84	6.66/9.16/10.09	9.79/12.17/12.41	4.91/7.87/7.98	5.25/8.83/9.31
fi	P@1	4.01	9.95	1.72/2.08/2.29	2.63/2.15/2.86	0.92/1.33/1.84	0.95/1.67/2.15
	P@5	6.16	16.68	4.51/4.87/6.94	6.92/6.44/10.26	2.97/3.68/5.01	3.34/4.06/5.97
	P@10	7.52	21.05	6.87/7.44/9.09	9.31/10.50/13.84	5.32/6.13/7.77	5.49/6.68/8.59
	P@30	11.74	28.27	11.45/12.38/14.03	14.80/16.47/18.38	9.30/10.63/11.76	10.26/10.98/11.93
ja	P@1	0.86	7.3	0.72/0.93/1.50	0.72/1.67/1.67	0.61/1.12/0.72	0.72/0.95/0.72
	P@5	1.79	15.03	2.65/3.29/3.36	3.34/3.34/5.01	2.04/2.86/2.15	3.10/3.58/1.19
	P@10	3.15	18.97	4.08/5.08/5.08	5.25/6.92/7.16	3.99/3.78/3.89	5.01/5.01/3.10
	P@30	5.73	26.06	7.80/8.95/9.66	8.59/11.46/13.13	6.75/7.46/8.28	7.88/8.83/7.88
th	P@1	0.43	7.66	0.36/0.64/0.72	0.72/0.24/1.19	0.31/0.72/0.92	0.72/0.72/0.72
	P@5	1.22	15.46	2.43/2.58/2.93	2.86/3.58/3.58	2.45/2.86/2.76	2.63/2.86/2.86
	P@10	1.79	19.9	4.15/5.30/5.30	4.77/6.44/6.92	4.19/4.81/4.29	3.82/4.53/5.25
	P@30	3.51	26.99	8.95/10.31/10.74	10.50/10.26/12.17	8.28/9.51/9.71	7.40/10.26/11.22

Table 14: Full results for mT0-large (1.2B) 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	29.35	52.76	18.83/22.26/24.19	18.14/22.91/25.06	19.12/21.98/23.82	17.42/20.76/23.39
	P@5	34.29	57.77	26.49/29.28/30.64	26.49/29.36/31.03	26.48/29.24/30.47	27.45/30.31/31.50
	P@10	37.01	59.2	28.42/31.85/32.93	28.64/32.22/33.41	28.32/31.70/32.72	29.59/33.65/34.84
	P@30	40.94	61.92	35.36/37.87/38.22	37.71/38.90/39.86	34.36/37.42/37.53	35.80/39.14/39.14
ro	P@1	22.69	52.18	11.88/15.75/17.90	11.22/17.18/18.62	12.17/15.13/17.59	13.60/15.99/17.90
	P@5	27.49	56.41	20.11/23.12/24.91	21.24/24.34/26.97	19.63/22.60/24.03	21.24/22.91/24.34
	P@10	29.56	58.05	22.33/25.77/27.77	24.11/27.92/29.36	21.57/24.85/27.10	22.91/25.54/27.45
	P@30	32.64	60.2	27.20/29.78/31.64	28.88/31.98/32.22	26.48/28.83/31.39	27.45/28.88/31.03
eu	P@1	15.18	54.97	5.44/8.88/10.59	4.30/9.31/10.50	5.93/8.69/10.63	5.97/9.31/11.22
	P@5	19.47	59.2	12.10/14.75/16.46	13.60/16.23/18.62	11.45/14.11/15.54	12.65/15.27/16.47
	P@10	21.55	60.92	14.46/17.54/18.54	16.47/18.62/20.53	13.60/17.08/17.69	14.80/17.42/18.38
	P@30	24.91	62.92	19.97/22.91/23.55	22.67/26.25/26.49	18.81/21.47/22.29	20.29/23.63/24.34
fi	P@1	15.32	56.34	5.08/7.59/9.88	5.49/9.07/11.22	4.91/6.95/9.30	5.25/7.40/9.79
	P@5	20.26	60.06	12.38/15.75/16.82	11.93/17.90/19.33	12.58/14.83/15.75	14.56/16.47/17.66
	P@10	22.41	61.13	15.25/18.47/20.40	16.23/21.24/23.87	14.83/17.28/18.92	16.47/19.33/19.33
	P@30	25.55	63.64	20.11/23.26/24.98	21.72/26.49/28.40	19.43/21.88/23.52	21.24/22.67/23.87
ja	P@1	10.31	56.26	1.86/4.15/6.51	1.91/3.58/5.73	1.84/4.40/6.85	1.91/3.82/6.44
	P@5	18.9	60.2	9.52/14.67/16.68	9.31/14.32/16.95	9.61/14.83/16.56	9.31/14.08/16.71
	P@10	23.34	60.92	14.46/19.69/21.12	15.51/19.33/20.29	14.01/19.84/21.47	14.08/20.29/21.96
	P@30	29.78	62.56	21.69/26.91/28.42	20.29/25.30/25.54	22.29/27.61/29.65	23.39/29.12/30.55
th	P@1	9.81	58.77	2.00/3.58/4.72	2.15/4.53/5.73	1.94/3.17/4.29	1.91/4.30/5.97
	P@5	18.18	63.28	12.17/14.67/17.68	12.17/14.80/16.95	12.17/14.62/18.00	13.13/15.75/19.09
	P@10	22.41	64.92	16.61/20.69/22.62	15.27/19.81/21.24	17.18/21.06/23.21	19.09/21.72/24.34
	P@30	30.92	67.07	26.77/29.85/31.50	24.58/28.88/28.64	27.71/30.27/32.72	27.92/30.55/33.17

Table 15: Full results for mT0-xl (3.7B) 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	34.93	49.25	40.37/41.30/42.02	50.60/51.31/52.03	35.99/37.01/37.73	32.70/35.32/35.32
	P@5	46.17	59.56	52.18/52.90/53.90	63.48/63.25/64.92	47.34/48.57/49.18	46.78/48.21/48.45
	P@10	50.89	63.56	56.69/57.55/58.12	66.59/67.30/67.54	52.45/53.37/54.09	52.03/52.98/53.94
	P@30	58.05	69.79	63.64/64.14/64.42	72.32/73.99/73.51	59.82/59.92/60.53	60.14/60.38/60.86
ro	P@1	29.85	41.8	30.64/33.21/34.29	37.47/41.77/42.24	27.71/29.65/30.88	29.59/32.22/33.65
	P@5	40.01	51.83	42.02/43.88/45.03	53.94/55.37/57.04	36.91/38.96/39.88	40.33/42.00/43.20
	P@10	44.31	56.41	46.89/48.53/48.60	59.19/60.14/60.14	41.62/43.56/43.66	45.58/46.78/46.78
	P@30	52.11	63.56	54.19/54.83/55.12	63.72/64.44/65.16	50.10/50.61/50.82	52.03/52.74/53.46
eu	P@1	12.17	26.41	15.03/16.46/17.39	18.14/19.57/21.24	13.70/15.13/15.75	14.80/15.75/15.75
	P@5	19.26	36.22	23.84/25.41/25.55	29.59/31.74/32.22	21.57/22.70/22.60	21.96/23.39/22.91
	P@10	21.83	39.58	27.06/28.99/29.78	33.41/36.99/37.23	24.34/25.56/26.58	24.34/26.49/27.21
	P@30	26.41	45.1	33.00/34.65/34.72	40.33/42.72/43.68	29.96/31.19/30.88	31.03/32.22/32.46
fi	P@1	20.54	35.15	22.91/25.34/25.98	27.92/30.79/32.22	20.65/23.01/23.31	22.20/24.82/24.34
	P@5	31.64	47.39	35.86/37.87/38.51	43.20/46.30/45.82	32.92/34.25/35.38	34.13/36.52/37.95
	P@10	36.65	51.61	40.23/41.52/42.09	48.93/49.88/50.84	36.50/37.83/38.45	38.19/40.81/41.05
	P@30	43.95	59.7	48.03/49.11/49.89	58.00/58.95/58.47	43.66/44.79/46.22	45.58/47.26/48.45
ja	P@1	19.26	37.15	20.69/21.76/22.69	23.15/25.78/24.34	19.53/20.04/21.98	22.20/22.91/24.82
	P@5	32.86	52.83	36.86/37.80/39.80	41.53/42.48/44.63	34.87/35.69/37.73	38.42/38.66/40.10
	P@10	39.23	58.12	43.81/44.45/45.81	49.40/48.69/50.60	41.41/42.84/43.76	45.11/45.35/45.82
	P@30	47.67	64.92	53.61/54.47/55.69	59.19/60.38/61.81	51.02/51.94/53.07	52.74/53.94/54.89
th	P@1	15.32	38.3	20.62/22.41/21.62	26.73/29.83/28.40	17.89/19.22/18.71	19.33/21.48/20.05
	P@5	27.56	54.26	38.37/40.52/39.94	44.39/47.73/48.93	35.79/37.32/36.09	37.47/39.14/38.66
	P@10	34.14	59.2	45.10/47.53/47.10	53.46/56.56/56.09	41.51/43.76/43.25	43.20/44.87/45.35
	P@30	45.1	68.79	55.48/56.76/56.62	62.29/64.44/64.92	52.56/53.48/53.07	54.18/54.65/55.37

Table 16: Full results for mT0-xl (3.7B) with prompt-based embedding