	Pretrain	Code	Law	Math	Average
Base Model	9.83	2.41	3.86	3.71	3.33
Specialized Models					
Code Dense Expert	15.39	2.18	5.22	4.34	3.91
Law Dense Expert	32.69	6.84	3.09	8.61	6.18
Math Dense Expert	20.32	3.20	5.11	3.20	3.84
Generalist Models					
BTX (Baseline)	10.35	2.40	3.76	3.64	3.27
BAM DM (Expert KV), ours	10.11	2.36	3.66	3.55	3.19
BAM CM (Expert KV), our	10.19	2.37	3.69	3.57	3.21
BAM DM (Shared KV), ours	10.20	2.37	3.69	3.59	3.22
BAM CM (Shared KV), ours	10.28	2.38	3.72	3.61	3.24

Table 1: Updated perplexity evaluation (\downarrow) for large-scale experiments (using a seed model of **2B parameters**).

	Math	Code	Law	Know.	Reason.	MMLU	Average
Seed Model	3.68%	9.41%	73.34%	21.33%	47.73%	34.13%	31.60%
Specialized Models							
Math Dense Expert	4.92%	12.39%	68.21%	13.32%	46.11%	34.29%	29.87%
Code Dense Expert	3.19%	18.80%	21.49%	12.18%	44.29%	31.50%	21.91%
Law Dense Expert	3.05%	0.20%	88.80%	10.41%	44.08%	32.18%	29.79%
Generalist Models							
BTX (Baseline)	3.86%	10.05%	81.85%	19.07%	47.36%	34.07%	32.71%
BAM DM	4 44%	12 83%	85 47%	19 89%	47 11%	34 42%	34 02%
(Expert KV), ours	1.11/0	12.0070	00.4170	10.0070	41.1170	01.1270	04.0270
BAM CM	4 34%	12 48%	82 79%	19 51%	47 43%	34 43%	33 50%
(Expert KV), ours	1.01/0	12.4070	02.1570	10.0170	11.1070	01.1070	00.0070
BAM DM	4.10%	11 76%	86 73%	19.48%	47.27%	34 55%	33 98%
(Shared KV), ours	1.1070	11.1070	00.1070	10.1070	11.2170	01.0070	00.0070
BAM DM	3.65%	11 77%	80.98%	19 22%	47.56%	34 16%	32.89%
(Shared KV), ours	5.0070	11.1170	00.0070	10.2270	11.0070	01.1070	02.0070

Table 2: Updated benchmark evaluations (\uparrow) on large-scale experiments. Highlighted entries indicate models outperform the BTX baseline.

	Pretrain	Code	Law	Math	Average
BTX	26.72	3.78	6.63	5.77	10.72
BAM soft-routing MoA	26.00	3.72	6.51	5.64	10.47
BAM top-2 routing MoA	26.68	3.78	6.69	5.75	10.72
BAM top-1 routing MoA	26.89	3.83	6.69	5.82	10.81

Table 3: Perplexity ablation (\downarrow) of BAM under different attention experts routing methods (compute-matched). Done on small scale experiments using seed model of 590M.