

Table 1: Performance comparison of different model methods under various configurations (1.45B parameters). AVG represents average perplexity (lower is better) or average correct score rate.

Method	Pa	Loss	PPL ↓			Accuracy ↑					
			Wiki	LMB	AVG	BOQA	PIQA	Hella	Wino	ARC-e	AI
<i>Vector decay</i>											
Mamba2	1.45	2.513	22.8	25.1	24.0	61.7	70.0	47.7	52.8	67.1	3
GLA	1.45	2.530	23.4	29.4	26.4	57.3	69.3	47.3	54.1	66.5	3
Hgrn2	1.45	2.526	23.2	24.3	23.8	59.7	70.1	47.2	52.5	65.9	3
LightNet	1.45	2.561	25.2	34.8	30.0	58.9	69.4	43.3	53.8	64.6	3
<i>Mamba ablation</i>											
Mamba2 w/o A	1.45	2.513	22.8	24.3	23.5	62.4	69.9	47.4	55.6	66.6	3
Mamba2 w/o Δ	1.45	2.585	25.3	31.5	28.4	58.5	68.9	44.7	50.8	64.9	3
Mamba2 w/o A , Δ	1.45	2.526	23.4	25.8	24.6	61.0	69.6	47.7	53.2	67.5	3
<i>Parameter share</i>											
Mamba2	1.45	2.517	22.8	24.6	23.7	60.7	69.9	47.4	54.1	66.8	3
GLA	1.45	2.583	25.5	35.9	30.7	61.7	69.4	45.5	50.8	65.5	3
Hgrn2	1.45	2.529	23.3	24.2	23.7	58.0	70.2	47.2	51.1	67.0	3
LightNet	1.45	2.620	26.0	49.1	37.6	60.9	68.8	42.7	50.9	61.5	3
<i>Scalar decay</i>											
Mamba2	1.45	2.529	23.4	28.3	25.8	56.6	69.3	47.0	51.7	66.7	3
GLA	1.45	2.550	23.8	28.9	26.3	60.6	70.0	46.3	52.6	65.9	3
Hgrn2	1.45	2.541	24.2	32.0	28.1	60.0	69.3	45.9	53.5	66.0	3
LightNet	1.45	2.574	24.3	33.3	28.8	62.0	69.3	45.1	51.3	65.3	2
TNL	1.45	2.552	24.3	29.4	26.9	61.3	69.9	45.9	53.8	66.6	3
TNL-L	1.45	2.545	23.7	29.0	26.4	59.6	70.7	46.1	51.4	64.1	3
<i>Rope</i>											
Mamba2	1.45	2.531	23.5	28.2	25.9	60.7	69.4	46.6	53.7	65.7	3
GLA	1.45	2.580	25.5	35.0	30.2	60.1	69.0	45.3	54.2	65.2	3
Hgrn2	1.45	2.560	24.6	29.3	27.0	59.1	69.2	45.6	51.5	66.0	3
LightNet	1.45	2.570	24.5	30.1	27.3	61.4	69.4	45.5	52.4	64.9	2
TNL	1.45	2.547	24.2	26.7	25.5	60.9	70.2	46.1	53.7	66.1	3
TNL-L	1.45	2.553	24.0	31.8	27.9	61.6	69.8	46.1	53.7	66.0	3
<i>Tpe</i>											
Mamba2	1.45	2.531	23.4	28.9	26.2	61.7	70.8	47.0	54.1	67.0	3
GLA	1.45	2.569	25.1	36.0	30.5	61.8	68.8	45.5	53.2	65.6	3
Hgrn2	1.45	2.554	24.3	31.0	27.7	61.7	69.5	46.3	52.6	65.5	3
LightNet	1.45	2.567	24.4	31.1	27.8	61.1	69.4	45.3	52.8	64.9	3
TNL	1.45	2.556	24.3	29.6	27.0	61.1	70.5	46.2	52.3	65.9	3
TNL-L	1.45	2.550	24.0	30.8	27.4	61.7	69.9	45.9	51.9	67.3	3
<i>Baseline</i>											
LLaMA	1.44	2.520	22.3	25.1	23.7	61.7	69.4	46.9	53.2	65.8	3