

Table 1: Model configurations for different parameter sizes.

<b>Params(B)</b>	<b>Layers</b>	<b>Hidden Dim</b>	<b>Num Heads</b>	<b>L.R.</b>	<b>Batch Size</b>	<b>SeqLen</b>
0.16	12	768	12	3E-04	32	2048
0.41	24	1024	16	3E-04	32	2048
1.45	24	2048	32	3E-04	32	2048