

Hyperparameter	Pythia-1B and OLMo-1B	BLOOM-1B
hidden_dim	2048	1536
num_layers	16	24
mid_layer	8	12

Table 1: **Model-specific Hyperparameters.** Note that the HF model names we used to load these models are `pythia-1b`, `OLMo-1B-0724-hf`, `bloom-1b1-intermediate`.