

Figure 1: **Correlation of architecture scales.** Dataset rankings are consistent when training 3B models (green) compared to 7B models (just as they were for 400M and 1B models).

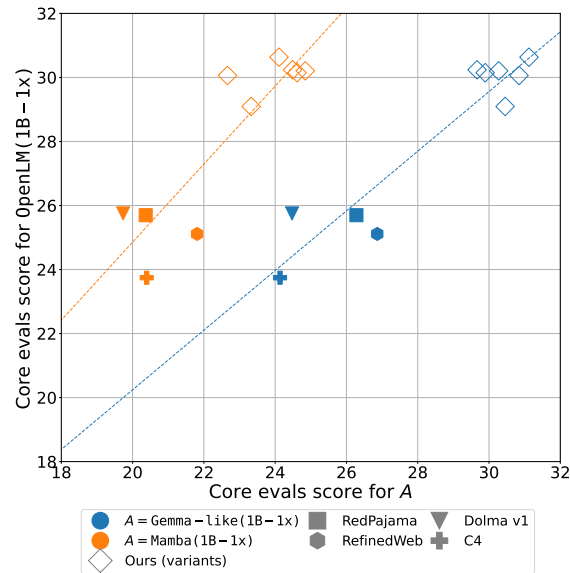


Figure 2: **Correlation of architecture variants.** Dataset rankings remain consistent when switching from our existing architecture to ones inspired by Gemma and Mamba. Our datasets still perform considerably better than existing baselines.