

Replace the vision encoder with the **PolyBERT**

Then the LM block can just be a decoder? Yes – replace such models like MPT, RedPajama, LLaMA, OPT, GPT-Neo, GPT-J, and Pythia with **ProtGPT2 (with Sigma-GPT variants?)**.