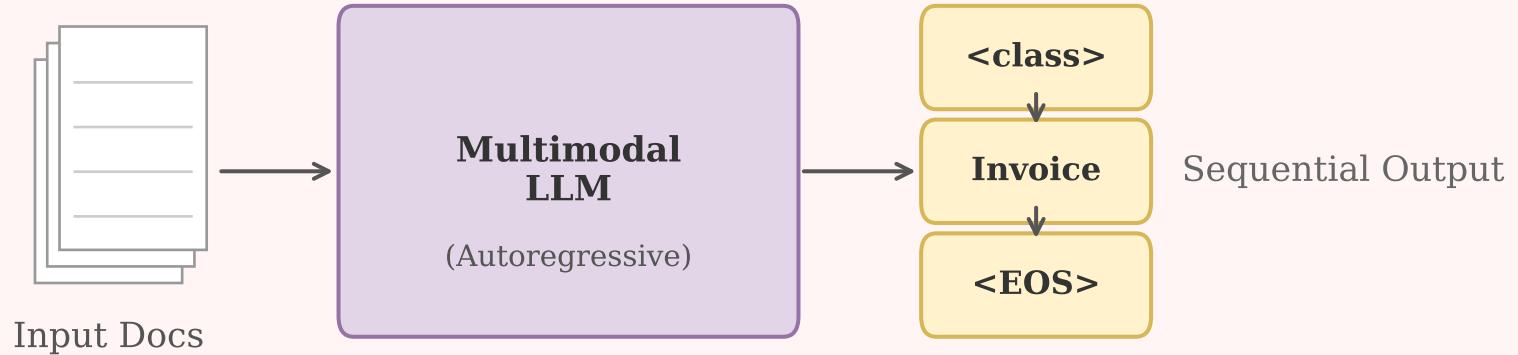


(A) Multimodal Generative Model



(B) CaVL-Doc (Ours)

