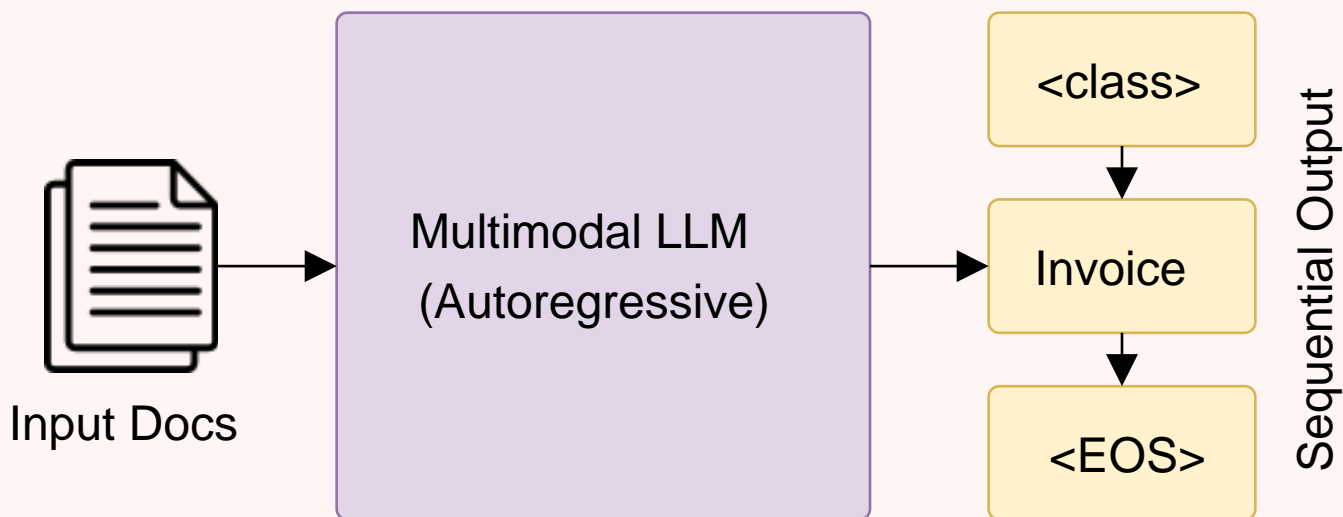


(a) Multimodal Generative Model



(b) CaVL-Doc (Proposed)

