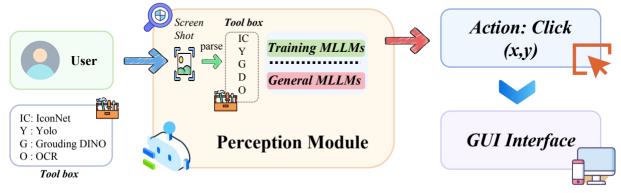


(a) Perception of Text-based LLM



(b) Perception of Multimodal LLM