Perception

(a) Perception of Text-based LLM

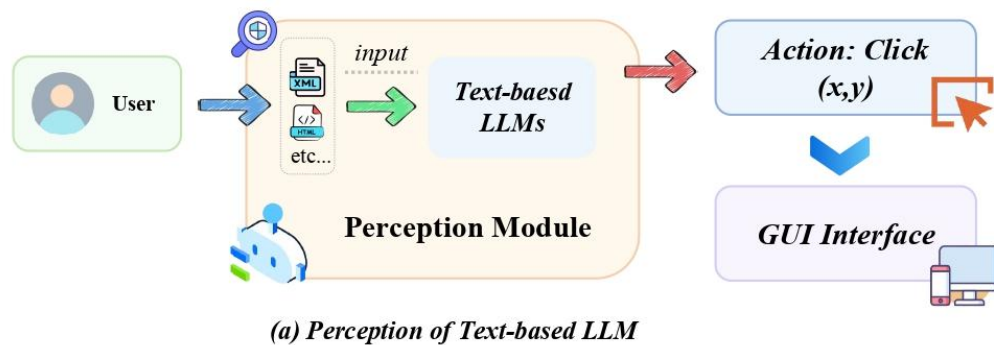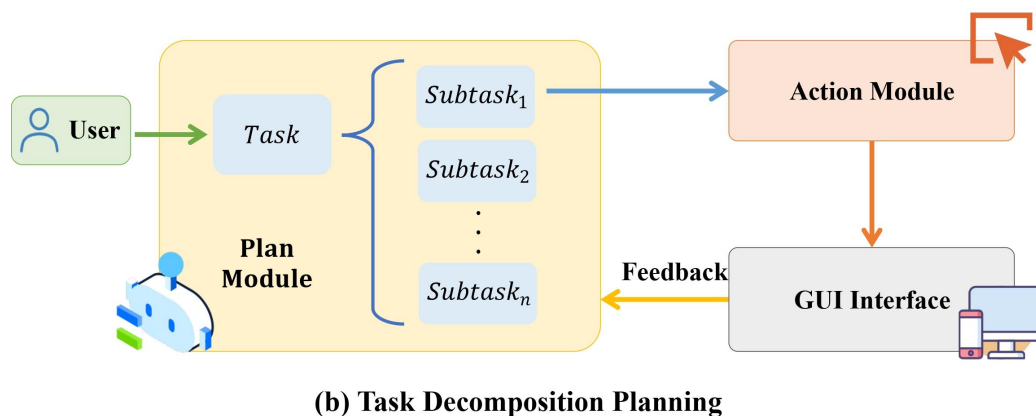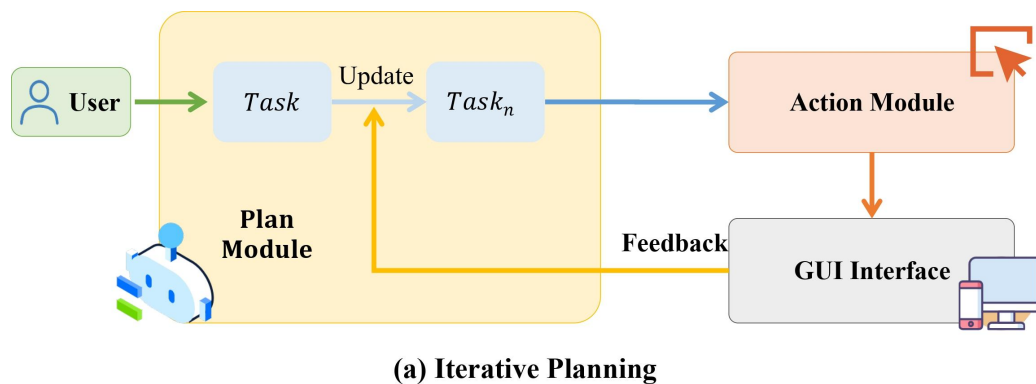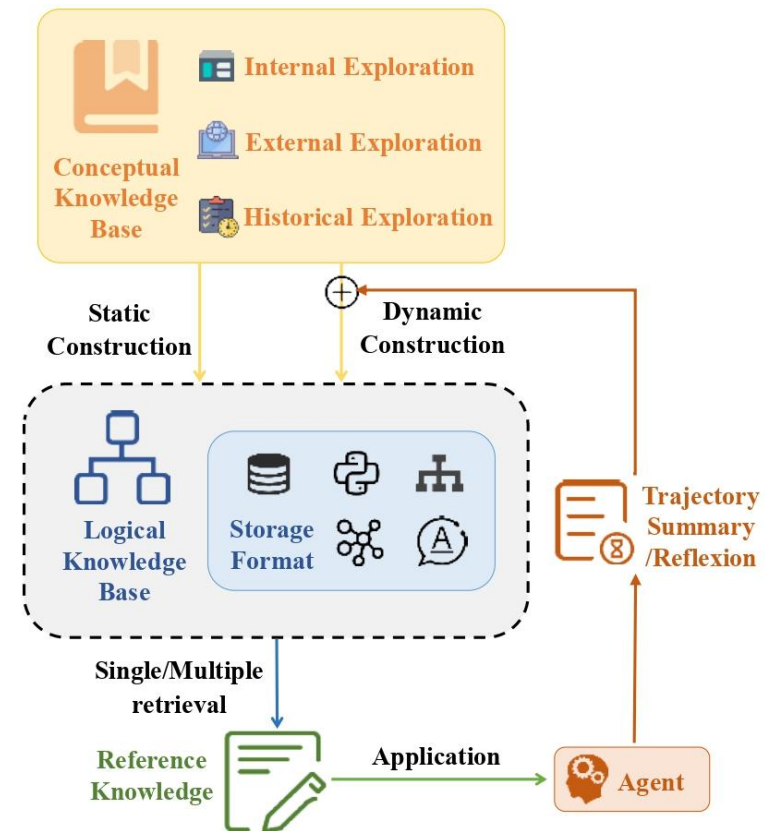*(b) Perception of Multimodal LLM*

Planning

(a) Iterative Planning

(b) Task Decomposition Planning

Exploration

Interaction

(a) Simulated Action

(b) API Call