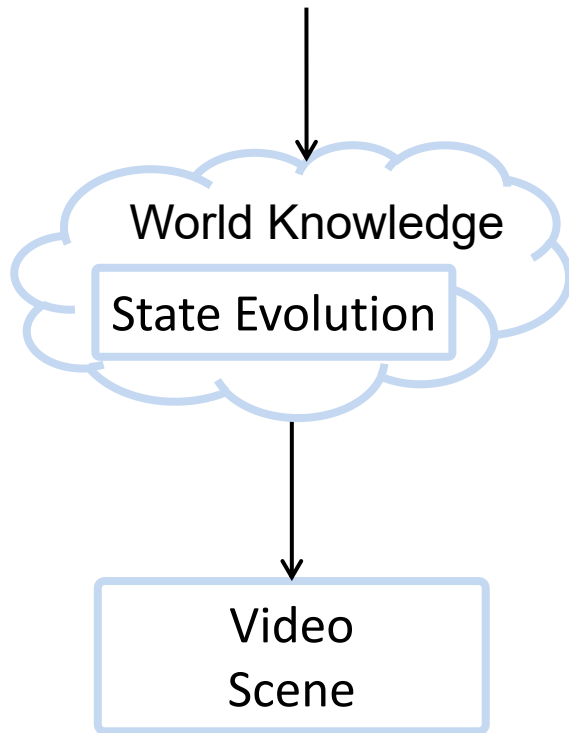


Physical World Model

Text, Vision, Audio, Spatial
Condition, Navigation Condition



Mental World Model

Text, Vision,
Audio

