(a) Guided Tour (b) Egocentric RGB-D observations + pose (c) Episodic Memory Representation (d) Question Answering