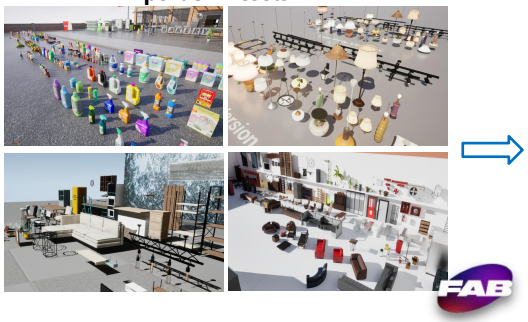
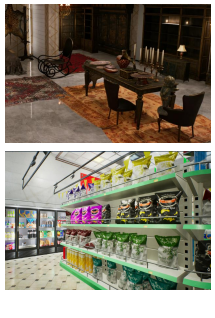
(a) Situated 3D Environment Construction

Import 3D Assets



Scene Generation



Scene Info.
Name:
Location:
Product List:

Product Info.
Asset:
Name:
Attribute Tags:
Price:



UNREAL ENGINE

(b) User Simulation

User Modeling.

Profile: Basic Info, Life Context, Personality, Interaction Style, ⋯
Shopping Goal: I need a small treat for lunchboxes, It should be easy to pack, ⋯

Virtual Embodiment.
Character Asset:
Action Execution: Blueprint Control, Skeletal Animation

UNREAL ENGINE



(c) Interactive Evaluation

① **Spawn User**



② **Start Conversation**

User

Assistant

③ **Termination Decision**

Moderator

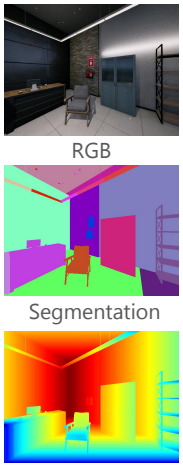
④ **Continue or End**

- Item in the shopping cart?
- or reached max T turns?
- or rejected recommendation ≥3 times?
- or failed moves in a row ≥3 turns?

Common Observation
(at t-th turn)



Set-of-Mark (SoM) Images
(Egocentric View, Role-specific Labels)

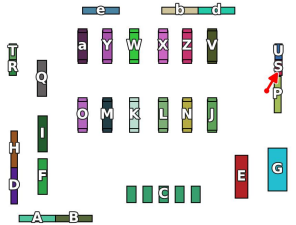


RGB / Segmentation / Depth

Assistant
(at t-th turn)

Input:
- Common Observation
- Dialogue History
- Scene Info.
- Product Info.
- Location Map

Action Space:
- **Dialogue Action**
  Ask Preference, Recommend, Navigate, ⋯



Location Map

User
(at t-th turn)

Input:
- User Info.
- Common Observation
- Dialogue History

Action Space:
- **Dialogue Action**
  Response
- **Physical Action**
  Move, Pick Up Item, Add to Cart, ⋯