

Input: Image



Input: Image + SoM



Conversation

- User** What items are there near the Marshal speaker?
- LLaVA-1.5** There is a laptop and a cup near the Marshall speaker. ❌
- User** To move the speaker closer to the curtain while keeping the laptop in the current position, which item should I swap it with?
- LLaVA-1.5** You should swap the laptop with the cup. ❌

Conversation

- User** What items are there near the Marshal speaker?
- SoM-LLaVA** There is a laptop tagged with number 7 and a notebook tagged with number 8. ✅
- User** To move the speaker closer to the curtain while keeping the laptop in the current position, which item should I swap it with?
- SoM-LLaVA** You can swap it with the lamp tagged with number 9. ✅