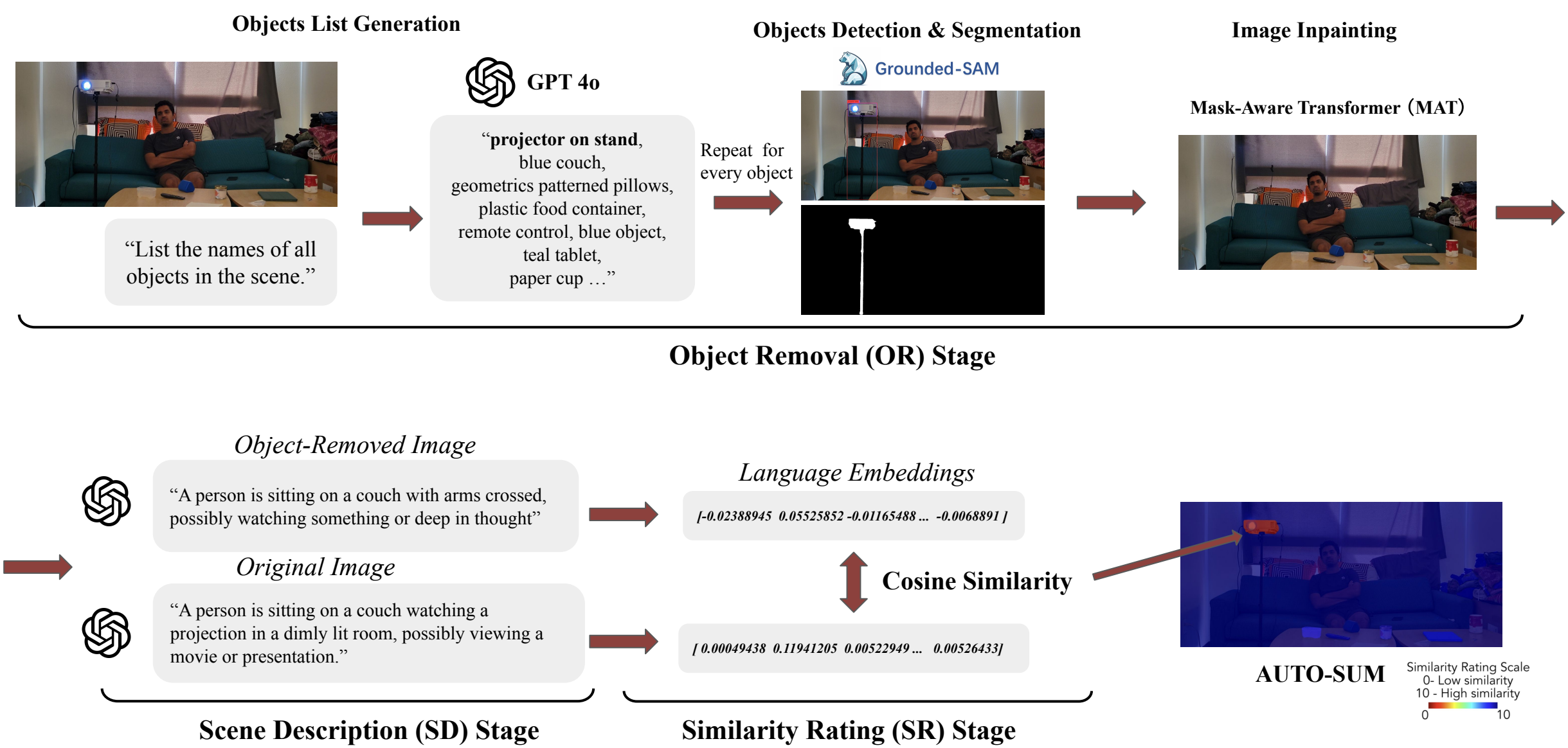
Objects List Generation
GPT 4o
Objects Detection & Segmentation
Grounded-SAM
Image Inpainting
Mask-Aware Transformer（MAT）
"List the names of all objects in the scene."
"projector on stand, blue couch, geometrics patterned pillows, plastic food container, remote control, blue object, teal tablet, paper cup …"
Repeat for every object
Object Removal (OR) Stage
Object-Removed Image
"A person is sitting on a couch with arms crossed, possibly watching something or deep in thought"
Original Image
"A person is sitting on a couch watching a projection in a dimly lit room, possibly viewing a movie or presentation."
Language Embeddings
[-0.02388945 0.05525852 -0.01165488 ... -0.0068891]
[0.00049438 0.11941205 0.00522949 ... 0.00526433]
Cosine Similarity
AUTO-SUM
Similarity Rating Scale
0- Low similarity
10 - High similarity
0
10
Scene Description (SD) Stage
Similarity Rating (SR) Stage