

## Open-Ended Multi-Modal Relational Reason for Video Question Answering

**Authors :** Haozheng Luo, Ruiyang Qin

**Abstract :** People with visual impairments urgently need assistance, not only on the fundamental tasks such as guiding and retrieving objects but on the advanced like picturing the new environments. More than a guiding dog, they might want such devices that can provide linguistic interaction. Building on this idea, we aim to study the interaction between the robot agent and visually impaired people. In our research, we are going to develop a robot agent that will be able to analyze the test environment and answer the participants' questions. We also will study the relevant issues regarding the interaction between human beings and the robot agents to figure out which and how the factors will affect the interaction.

**Keywords :** HRI, video question answering, visual question answering, natural language processing

**Conference Title :** ICSRHRI 2021 : International Conference on Social Robotics and Human-Robot Interaction

**Conference Location :** Tokyo, Japan

**Conference Dates :** September 09-10, 2021