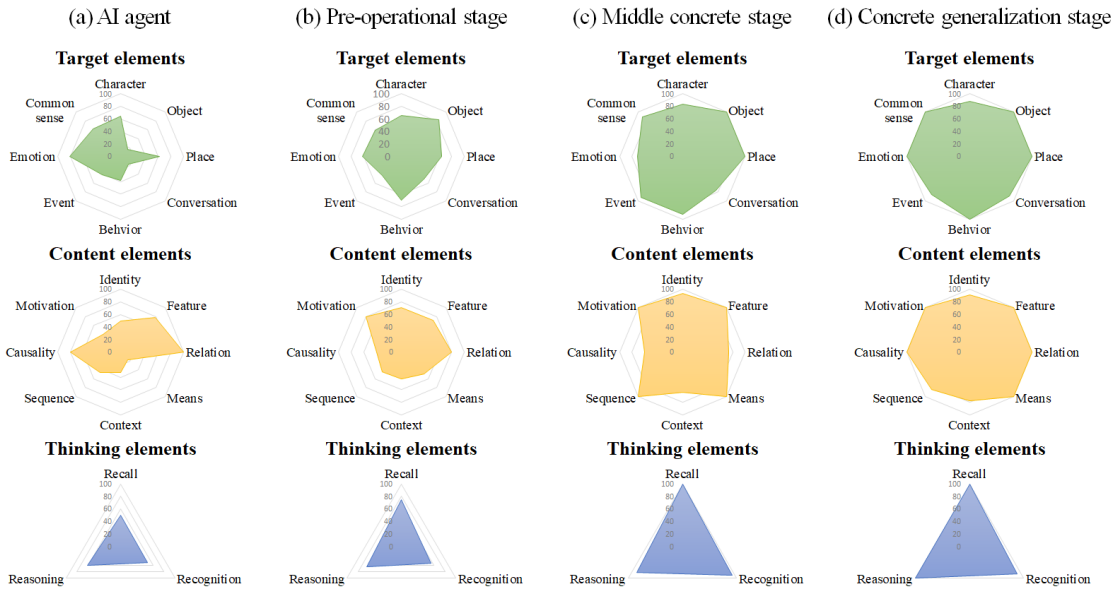

Figure 2: The accuracy-based profiles for each player participated in the case study of the Video Turing Test (VTT). To evaluate each player's video story intelligence, we apply CogME based on cognitive process of human.

Analysis on cognitive modules

We analyze each player's performance based on CogME for specifying their video understanding intelligence. We calculate the correct answer rate (accuracy) for each story element associated with the questions used in the case study. The accuracy-based profiles for each player on story elements in the three cognitive modules (i.e., target, content, thinking) are shown in Figure 2. Here, we highlight two key observations as follows. First, as the developmental stages of human progress, the accuracy of QA across all three modules is gradually improved. This means that some cognitive processes mature in a way that is reflected in the chosen metrics of this test. This result validates the use of these measures to evaluate video understanding intelligence. Second, when comparing the profiles between the AI agent and the human players, we notice that the performance of each component revealed more evenly in human players than the AI agent. Quantitatively, the standard deviation of the accuracy of each story element is 13.78 for a child in the pre-operational stage and 23.13 for the AI agent. In particular, the AI shows deficient performance in Object and Conversation in Target elements, Means and Motivation in Context elements, and Recall in Thinking elements. It is expected for two reasons that the AI agent's profile differs from the previous paper (**?**). This study considers two types of QA, multiple-choice and open-ended, while the previous work use only the former. Furthermore, we employ two distinct video QA algorithms as our AI agents to answer multiple-choice and open-ended QA.

Discussion and Conclusion

In this paper, we introduced the Video Turing Test (VTT), a novel measurement to evaluate a human-likeness of video understanding intelligence. We defined a general format and procedure of VTT and conducted a case study to confirm the feasibility and effectiveness of the proposed test. The case study provided new insight into the association of video understanding intelligence between the AI agent and human players from the different developmental stages. While the case study suggested a new perspective of measurement for video understanding intelligence, we still need to discuss several aspects. First, a 4-years child was included in the test as a human player for comparing the video understanding ability between the AI agent and a human from the pre-operational stage. However, several requirements (e.g., writing as a full sentence, choosing an answer among five answer candidates) were not familiar for the child so that it is hard to interpret the answers as totally reflecting the child's video understanding ability. Furthermore, we did not strictly specify the criteria to pass the VTT. The passing criteria can be defined by considering the detailed design of the VTT such as the composition of human players, juries, and the selected question set. As future work, we expect that the additional case study and its interpretation validates the clearness of VTT. For more objective and analytical evaluation for VTT, we plan to conduct various case studies and suggest appropriate guidelines including the composition of participants and the arrangement of question set.

Acknowledgements

This work was partly supported by the IITP (2015-0-00310-SW.StarLab/20%, 2017-0-01772-VTT/20%, 2018-0-00622-RMI/15%, 2019-0-01371-BabyMind/15%, 2021-0-02068-AIHub/15%) grants and the NRF of Korea (2021R1A2C1010970/15%) grant funded by the Korean government.

Consequuntur dignissimos fuga quis expedita nostrum

consectetur quam voluptas nemo rerum, corrupti deleniti repellat numquam ratione