## Imagining the image Learning new information 1.0 -CLIPScore description description R = 0.47\*\*\*R = 0.10(accessibility) (accessibility) caption caption R = 0.32\*\*\* $R = 0.14^*$ (contextualization) (contextualization) 0.0 0.5 0.0 0.5 0.0 Human rating