Supplemental: Probing Image-Language Transformers for Verb Understanding

Lisa Anne Hendricks Aida Nematzadeh
DeepMind
{lmh, nematzadeh} @google.com

1 The Annotation Pipeline

We provide figures depicting the interfaces of our three annotation tasks.

☐ This image won't load. No need to answer any more questions; move onto the next hit.
☐ This image is a cartoon. No need to answer any more questions; move onto the next hit

Q1: Can you see the verb **play** in the image?

○ Yes

○ No

○ Unsure

Q2: Can you see a/an **activist** and a/an **guitar** in the image?

○ Yes

○ No

○ Unsure

Figure 1: The interface of our first task for collecting images.

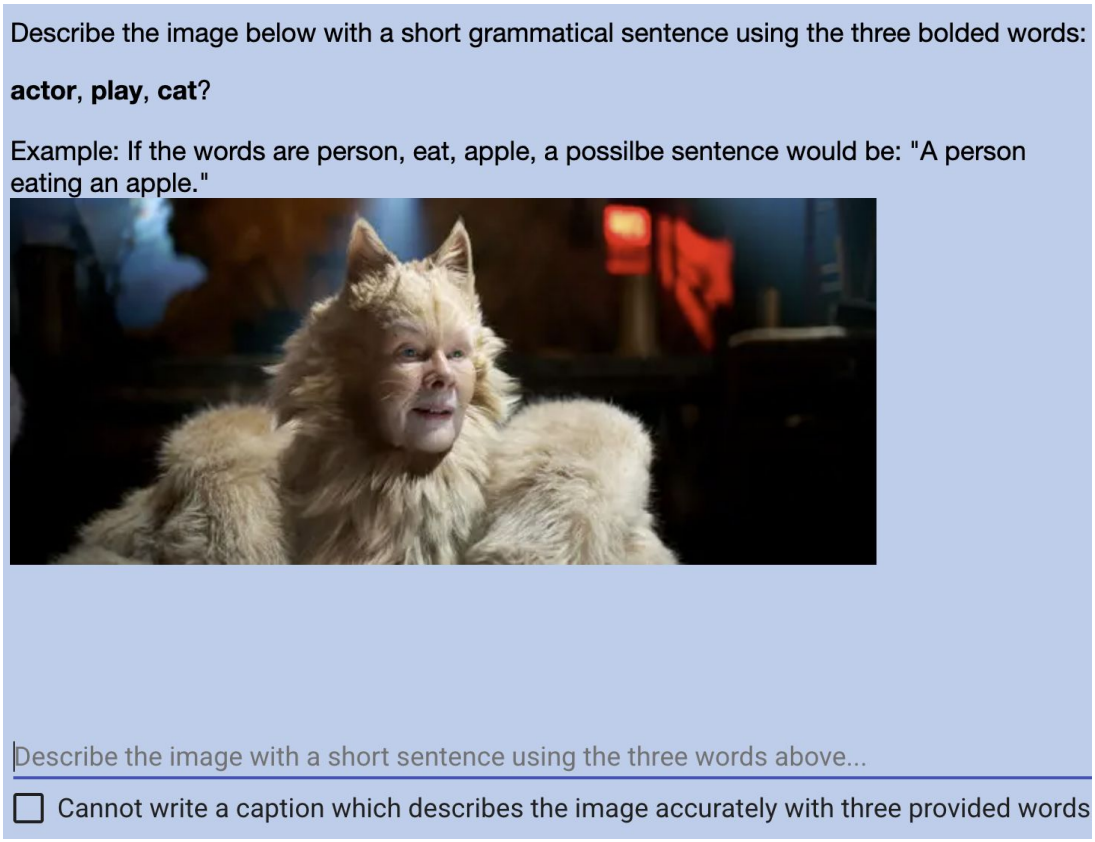

Figure 2: The interface of our second task for writing sentences for images.

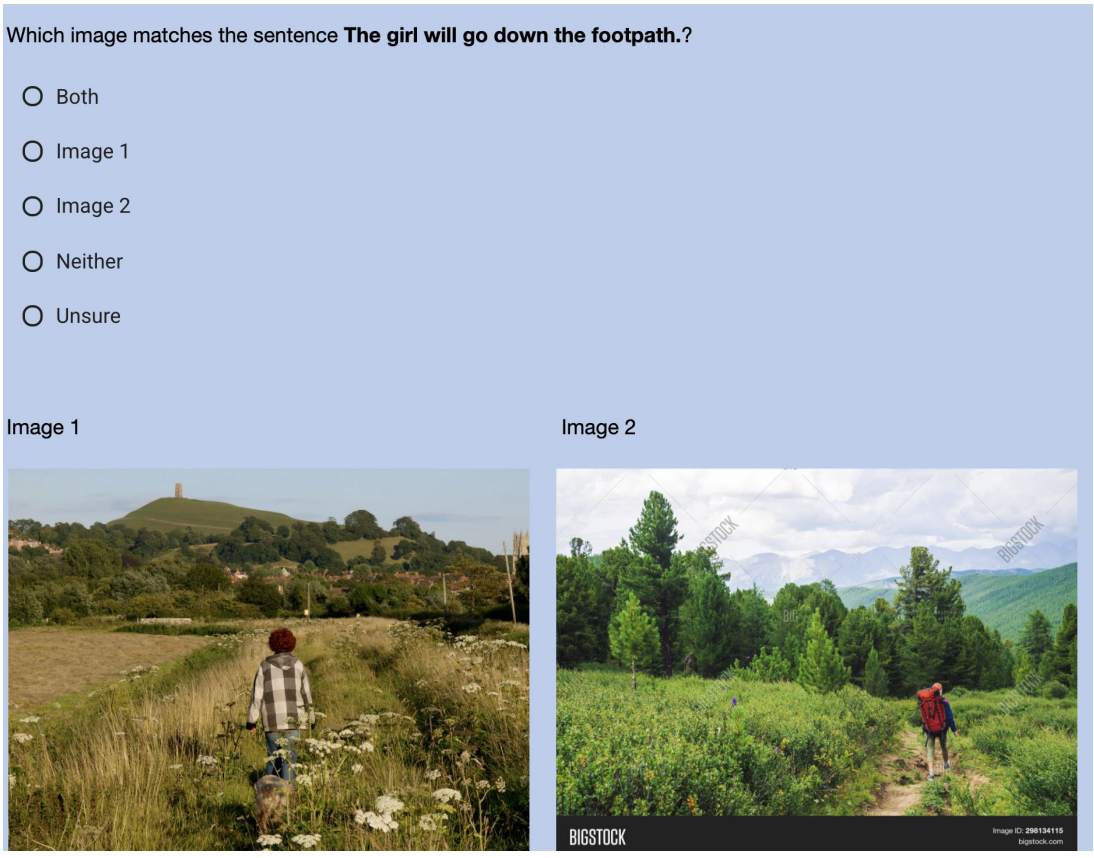

Figure 3: The interface of our third task for confirming images.