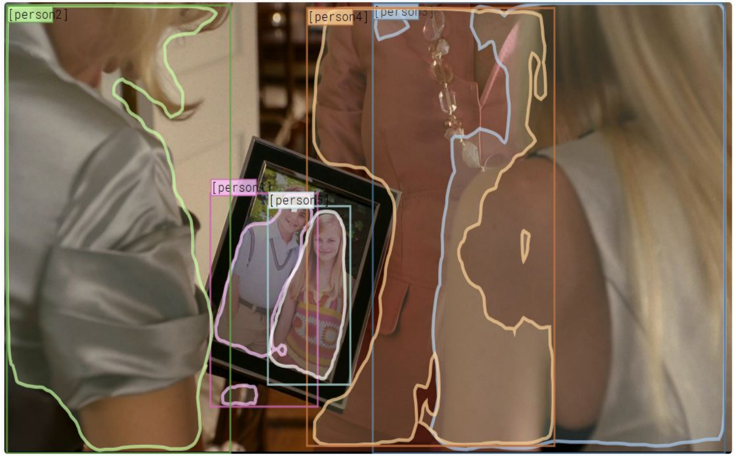
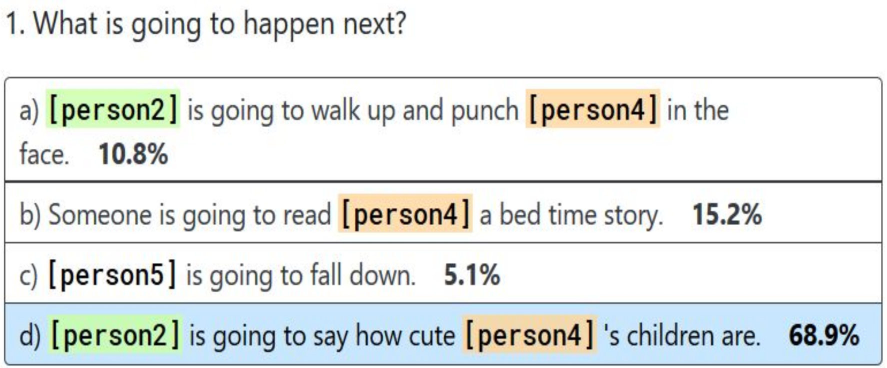
Generating Rationale in VQA

Hammad Abdullah Ayyubi (A53283495)
Ranak Roy Chowdhury (A53317421)

Task





1. What is going to happen next?

a) [person2] is going to walk up and punch [person4] in the face. **10.8%**

b) Someone is going to read [person4] a bed time story. **15.2%**

c) [person5] is going to fall down. **5.1%**

d) [person2] is going to say how cute [person4] 's children are. **68.9%**

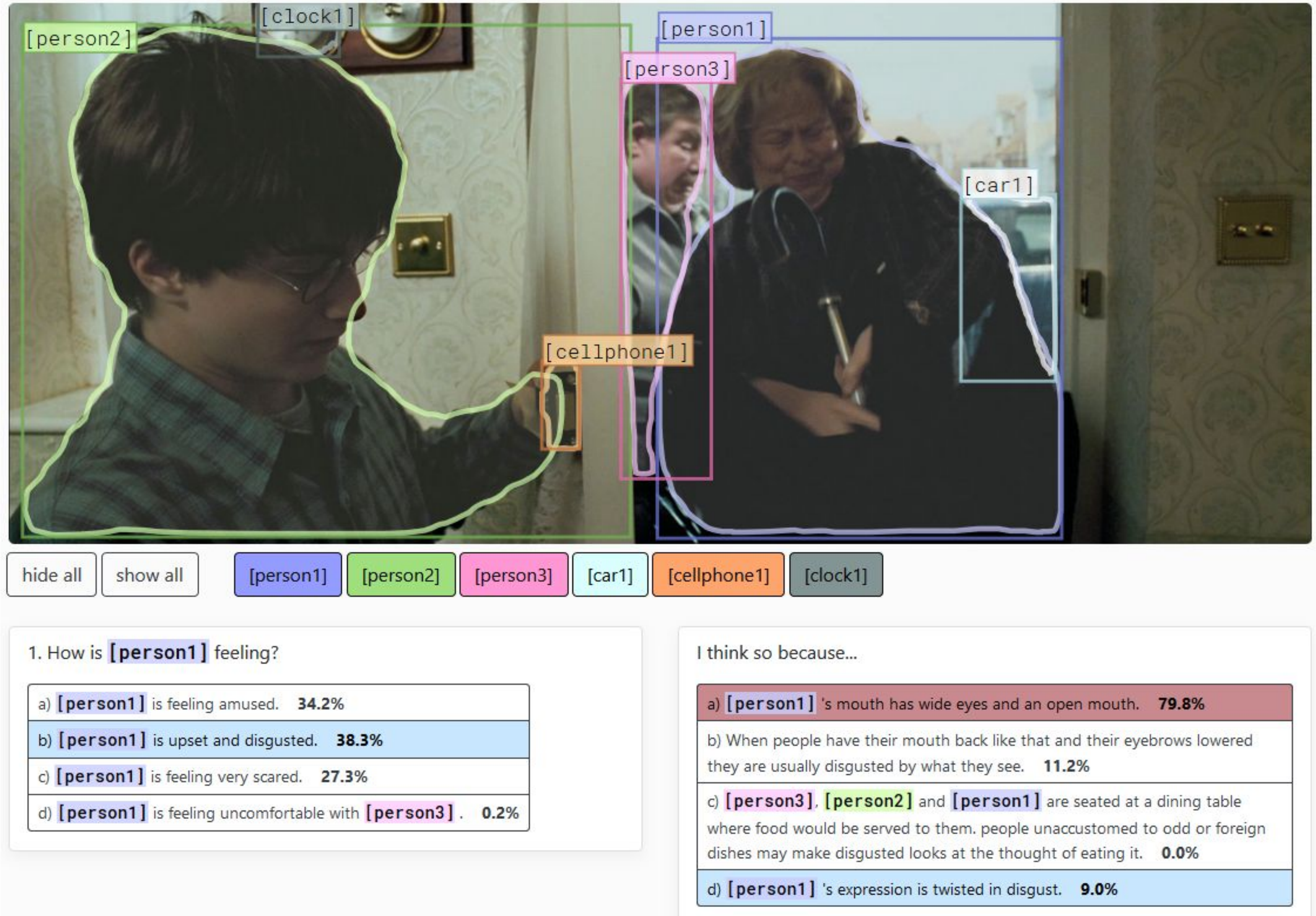
- State-of-the-art VQA Models: DFAF (Peng et al): 70.22% accuracy
- But how well do they understand the answer they are predicting?
- We propose the novel task of generating rationale to address this.

Gao Peng, Hongsheng Li, Haoxuan You, Zhengkai Jiang, Pan Lu, Steven Hoi, and Xiaogang Wang.Dynamic fusion with intra-and inter-modality attention flow for visual question answering.arXiv preprintarXiv:1812.05252, 2018

Task

- The objective of the task is to predict answer given a question and image.
- The novelty of task lies in the fact that we further ask the model to provide a rationale to justify the answer it predicted.
- This makes the task more interesting and concurrently more difficult.
- Dataset for rationale generation: Visual Commonsense Reasoning (VCR) Dataset. (Zellers et. al.)

Visual Commonsense Reasoning (VCR) Dataset (Zellers et. al.)



1. How is [person1] feeling?

a) [person1] is feeling amused. **34.2%**

b) [person1] is upset and disgusted. **38.3%**

c) [person1] is feeling very scared. **27.3%**

d) [person1] is feeling uncomfortable with [person3] . **0.2%**

I think so because...

a) [person1] 's mouth has wide eyes and an open mouth. **79.8%**

b) When people have their mouth back like that and their eyebrows lowered they are usually disgusted by what they see. **11.2%**

c) [person3], [person2] and [person1] are seated at a dining table where food would be served to them. people unaccustomed to odd or foreign dishes may make disgusted looks at the thought of eating it. **0.0%**

d) [person1] 's expression is twisted in disgust. **9.0%**

Investigating Rationale Generation for VCR leader (ViLBERT)

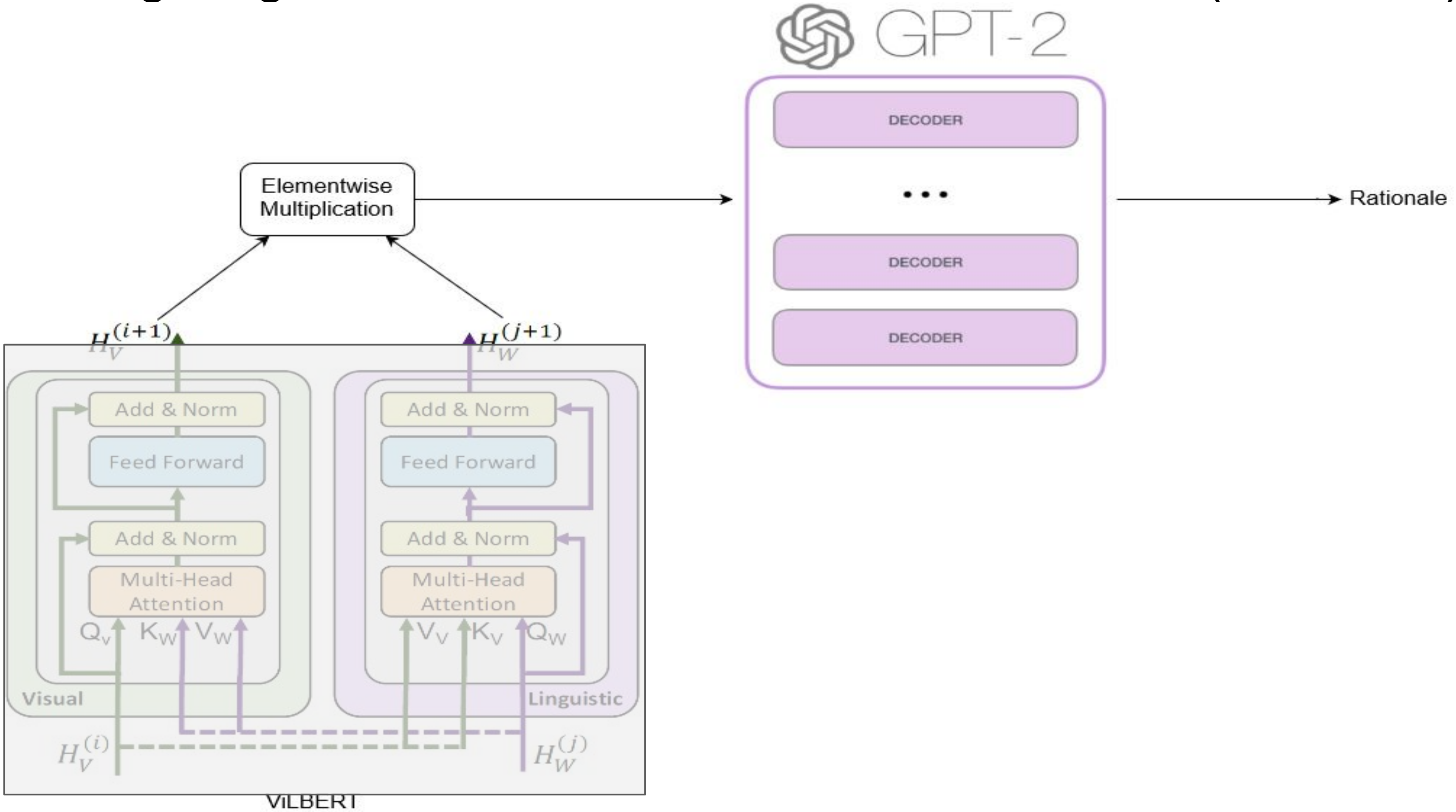


Fig: Model Architecture

Proposed Model

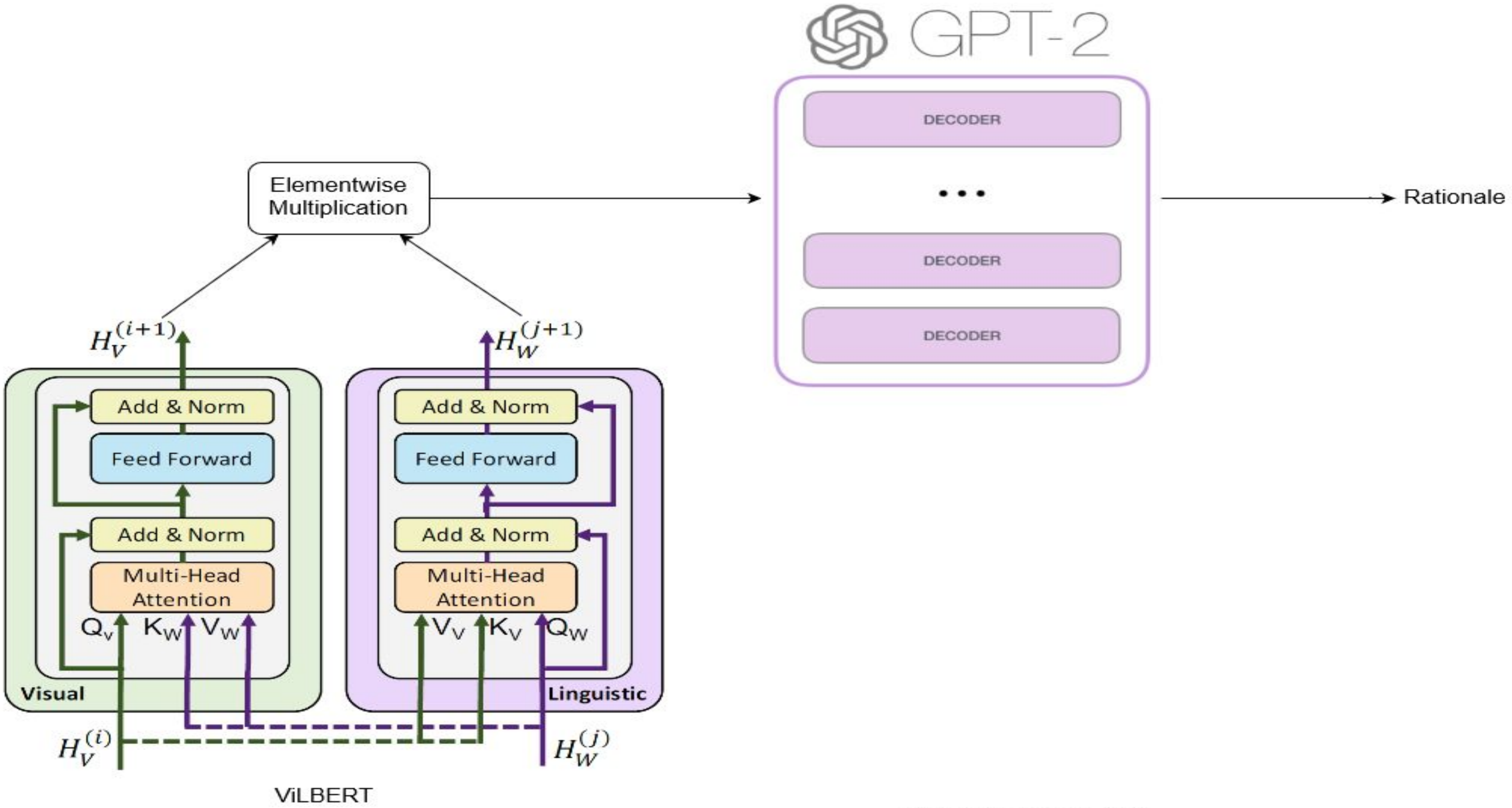
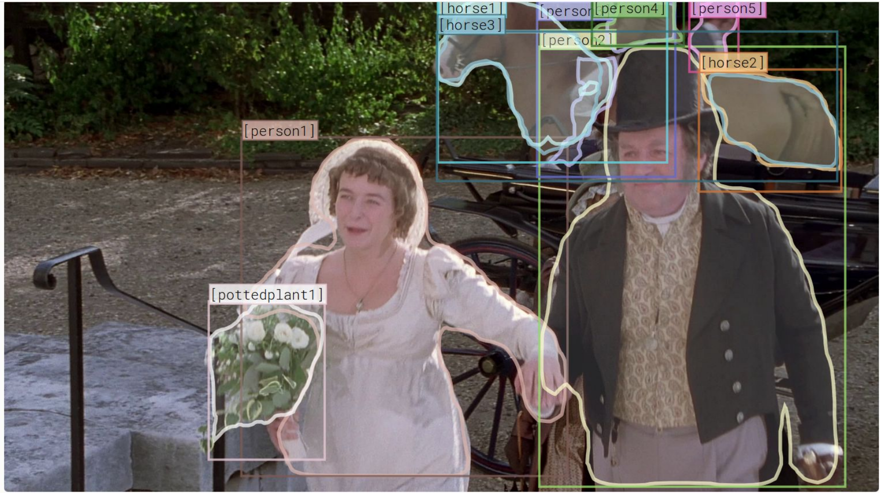


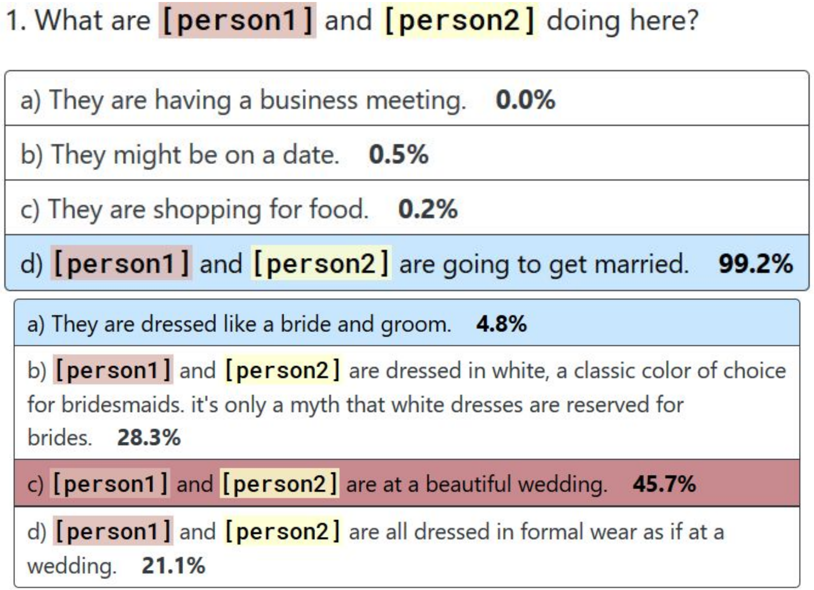
Fig: Model Architecture

Results

	Frozen ViLBERT Model	Our proposed Model
BLEU Score	38.8	67.8
Rouge Score	11.7	14.6

Correct Rationale: Example 1





1. What are [person1] and [person2] doing here?

a) They are having a business meeting. **0.0%**

b) They might be on a date. **0.5%**

c) They are shopping for food. **0.2%**

d) [person1] and [person2] are going to get married. **99.2%**

a) They are dressed like a bride and groom. **4.8%**

b) [person1] and [person2] are dressed in white, a classic color of choice for bridesmaids. it's only a myth that white dresses are reserved for brides. **28.3%**

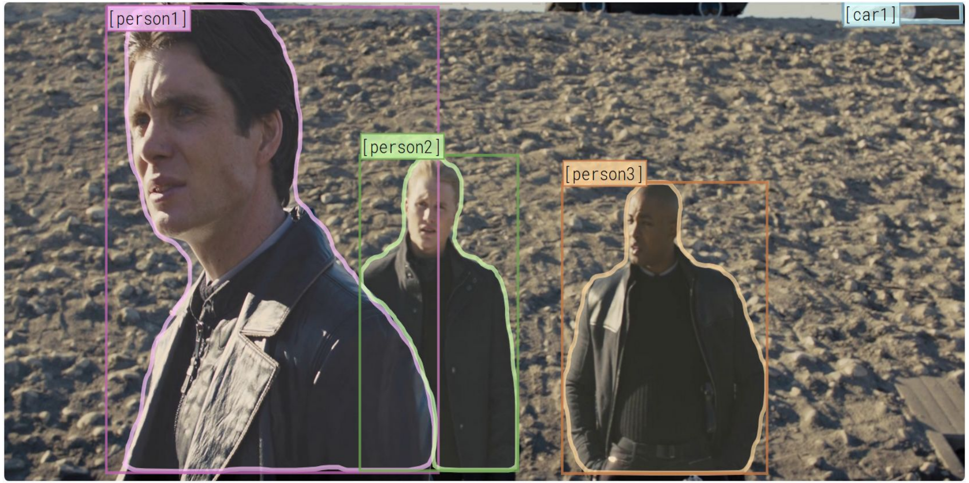
c) [person1] and [person2] are at a beautiful wedding. **45.7%**

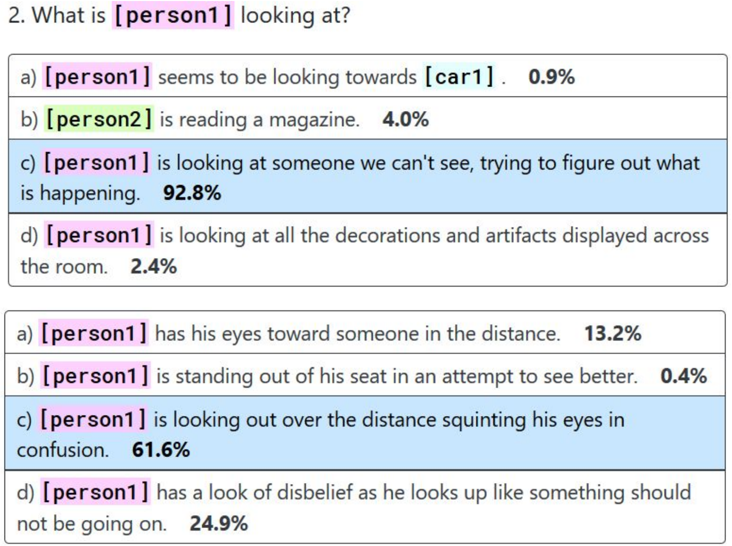
d) [person1] and [person2] are all dressed in formal wear as if at a wedding. **21.1%**

Predicted Rationale by Frozen ViLBERT: He is facing the train .

Predicted Rationale by proposed model: They look very pleased with each other and are happily dancing .

Correct Rationale: Example 2





2. What is [person1] looking at?

a) [person1] seems to be looking towards [car1] . 0.9%

b) [person2] is reading a magazine. 4.0%

c) [person1] is looking at someone we can't see, trying to figure out what is happening. 92.8%

d) [person1] is looking at all the decorations and artifacts displayed across the room. 2.4%

a) [person1] has his eyes toward someone in the distance. 13.2%

b) [person1] is standing out of his seat in an attempt to see better. 0.4%

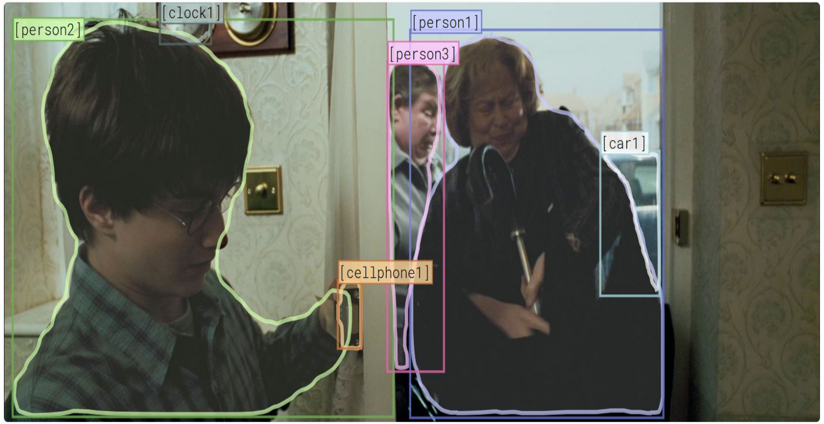
c) [person1] is looking out over the distance squinting his eyes in confusion. 61.6%

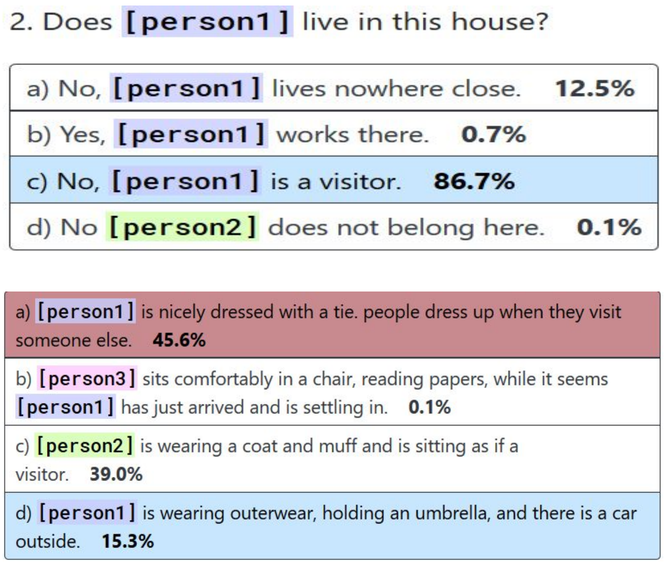
d) [person1] has a look of disbelief as he looks up like something should not be going on. 24.9%

Predicted Rationale by Frozen ViLBERT: Anay is looking at Britain with determination , his eyes are looking down and he is unimpressed .

Predicted Rationale by proposed model: Kaylor is looking in another direction like someone is or has said something .

Incorrect Rationale: Example 1




2. Does [person1] live in this house?

a) No, [person1] lives nowhere close. **12.5%**

b) Yes, [person1] works there. **0.7%**

c) No, [person1] is a visitor. **86.7%**

d) No [person2] does not belong here. **0.1%**

a) [person1] is nicely dressed with a tie. people dress up when they visit someone else. **45.6%**

b) [person3] sits comfortably in a chair, reading papers, while it seems [person1] has just arrived and is settling in. **0.1%**

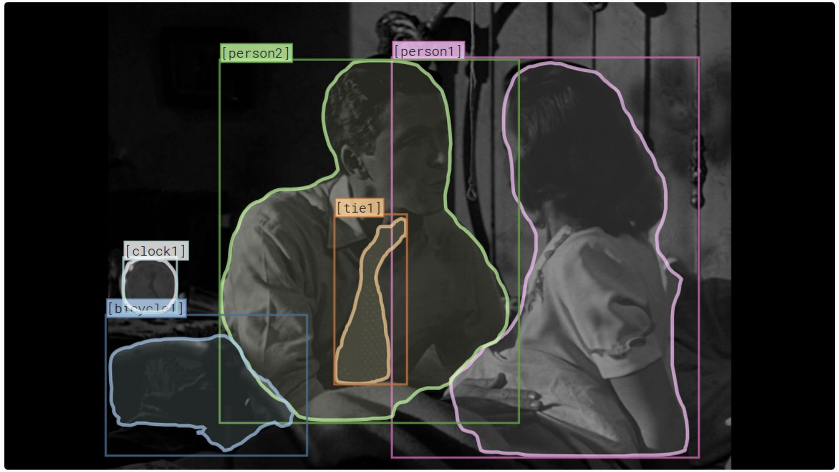
c) [person2] is wearing a coat and muff and is sitting as if a visitor. **39.0%**

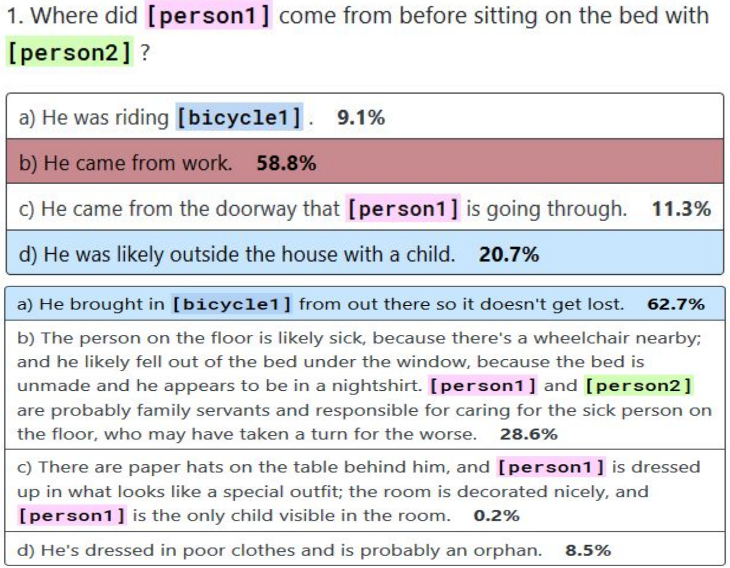
d) [person1] is wearing outerwear, holding an umbrella, and there is a car outside. **15.3%**

Predicted Rationale by Frozen ViLBERT: Yi has a caucasian outfit . Yacine Camil look identical . the ornate style of that era probably had a luxurious atmosphere .

Predicted Rationale by proposed model: Justice is recouping money for a lost child .

Incorrect Rationale: Example 2





1. Where did [person1] come from before sitting on the bed with [person2] ?

a) He was riding [bicycle1] . 9.1%

b) He came from work. 58.8%

c) He came from the doorway that [person1] is going through. 11.3%

d) He was likely outside the house with a child. 20.7%

a) He brought in [bicycle1] from out there so it doesn't get lost. 62.7%

b) The person on the floor is likely sick, because there's a wheelchair nearby; and he likely fell out of the bed under the window, because the bed is unmade and he appears to be in a nightshirt. [person1] and [person2] are probably family servants and responsible for caring for the sick person on the floor, who may have taken a turn for the worse. 28.6%

c) There are paper hats on the table behind him, and [person1] is dressed up in what looks like a special outfit; the room is decorated nicely, and [person1] is the only child visible in the room. 0.2%

d) He's dressed in poor clothes and is probably an orphan. 8.5%

Predicted Rationale by Frozen ViLBERT: Kery is looking directly at Marchan as she lies on a bed in Sonnie Jailin hiding from them .

Predicted Rationale by proposed model: He is wearing a nightgown that most typically only goes worn on after getting into bed .