

A Qualitative Results

While the experimental results demonstrate the benefit of the proposed method quantitatively, we perform some additional analysis to gain a better insight into the operation of the models.

In particular, we analyze the outputs of the proposed model GPT-speaker and the baseline Up-Down with different personality traits. In particular, given an image from the development data, we use both models to generate captions for every personality trait in the dataset (i.e., the 215 personality traits). By comparing a sample of outputs of the two models over different input images and personality traits, we find that UpDown’s generated captions tend to be less engaging and less consistent with the input images and personality traits than those from the proposed model. Tables 4 and 5 shows some examples for the two models on some personality traits. On the one hand, we attribute the better engagement of the generated captions from GPT-Speaker to the employment of the GPT2 model that has high language modeling capacity to produce more interesting captions for PIC. On the other hand, the better consistencies with the input images and traits for the generated captions of GPT-Speaker can be explained by the introduction of the rewards R_{img} and R_{trait} in the listener of the model. The goal of these rewards is to improve the informativeness of the generated captions about the input images and personality traits (via the identification of the input images and traits among the distractors), thereby boosting such consistencies as the direct outcome. Overall, these qualitative results demonstrate the effectiveness of the proposed model GPT-Speaker for PIC in this work.

Image	Trait	Generated Caption
	Adventurous	GPT-Speaker: That is a cool building, i would love to climb to the top. UpDown: I know that this place is about to UNK it's so much fun .
	Cultured	GPT-Speaker: This building reminds me of the type of architecture around the european town. UpDown: This is this a beautiful place for my summer and put the UNK next
	Eloquent	GPT-Speaker: They look like nice, very well-built for a wall and for a window ornament, so vintage. UpDown: The rug is nothing taken of a chinese building
	Extraordinary	GPT-Speaker: The architecture of this building's design is grand! UpDown: That's a nice place
	Logical	GPT-Speaker: More concrete must be needed to stand around this building. UpDown: I love looking at this type of pink design
	Emotional	GPT-Speaker: This reminds me of the train ride that i had with my grandmother on our trip to the city. UpDown: I love to have everyone in this picture
	Sensual	GPT-Speaker: Those trains reminds me of how i would blow in the sun. UpDown: Great picture of an amazing classic
	Aloof	GPT-Speaker: These trains make no sense to me. why do you have to travel on them? UpDown: The train is on a regular room
	Arrogant	GPT-Speaker: I can make that train better than that train. UpDown: Old buildings are they like me
	Confused	GPT-Speaker: What are the ferry on the train? UpDown: When i really don't feel this area
	Money-minded	GPT-Speaker: If i paid someone these flowers for all that could be worth the cost, should everyone get in the way? UpDown: This is a pretty nice picture
	Doubtful	GPT-Speaker: This looks photoshopped in color UpDown: Do they really have to assume these bushes
	Critical	GPT-Speaker: Those flowers are too black UpDown: That seems like a big tower
	Extreme	GPT-Speaker: I will fall off this flower UpDown: I love this photo
	Hostile	GPT-Speaker: This is just an basic flower. UpDown: I think they need to want a job this good
	Stiff	GPT-Speaker: Statues are just a religious emblem, not unique. UpDown: Doesn't that good
	Morbid	GPT-Speaker: For all these statues, this is no way to reflect all the time has passed. UpDown: What 's going this at
	Gloomy	GPT-Speaker: This statue reminds me of the bust of our history. UpDown: This looks like a statue in this picture
	Fearful	GPT-Speaker: Scary statue that is. UpDown: I wonder what this is
	Excitable	GPT-Speaker: That statue is soo cute! UpDown: Very dark picture

Table 4: Output examples of the proposed model GPT-Speaker and the UpDown model (Shuster et al., 2019) for the different personality traits with images on the development set.

Image	Trait	Generated Caption
	Procrastinating	GPT-Speaker: I'd like to sweat all the cooking in that stove. UpDown: You would not have a lovely life
	Stubborn	GPT-Speaker: She is doing love cooking on the kitchen table. Don't even try to get a meat for him, can she cook that pot? UpDown: This is a good picture i have to buy this
	Pretentious	GPT-Speaker: I could make better cooking than these people UpDown: I have no idea what this is
	Moody	GPT-Speaker: Those pots look disturbing UpDown: Looks like a fun day
	Fawning	GPT-Speaker: He looks like he is cooking something nice. UpDown: This is a very beautiful family
	Shy	GPT-Speaker: I would be embarrassed to compete in those swimsuits. UpDown: I need to know what this is
	Enthusiastic	GPT-Speaker: I hope to see the swimmer on the bottom!? UpDown: This is beautiful
	Humorous	GPT-Speaker: These girls look like babies swimming in the pool! Yuck! UpDown: Look i'd love to go
	Objective	GPT-Speaker: They can win race this year UpDown: Being on your day
	Formal	GPT-Speaker: They seem like they are taking a swim. the guys in the swimming attire looks nice to them. UpDown: I love the game of people
	Stupid	GPT-Speaker: What are these animals? UpDown: I do not care about that animal
	Extreme	GPT-Speaker: These animals are unbelievable at feeding on this verdant grass! UpDown: I wonder what a beautiful animal
	Argumentative	GPT-Speaker: These animals look really far from the fence to those people or should get along. UpDown: Well you have a bird
	Formal	GPT-Speaker: These animals are in a proper grazing area. UpDown: The young family
	Emotional	GPT-Speaker: This reminds me of the animals that I grew up in the field and the vibrant colors of these horns! UpDown: I bet these kids aren't waiting for those babies
	Imaginative	GPT-Speaker: This is such a great view of the lake. have you ever seen the ghost? UpDown: I bet this is a cool resort
	Exciting	GPT-Speaker: These clouds are awesome! I wish i was there! UpDown: I love the sun in the sky
	Deep	GPT-Speaker: The view of the lake is like a mirror of time. UpDown: Stunning yet to work
	Playful	GPT-Speaker: I would love to have a fun jumping in the clouds - to see this? UpDown: This photo is the only water i can see
	Sweet	GPT-Speaker: What a scenic view from the side of the mountain UpDown: What a nice little view

Table 5: Additional output examples of the proposed model GPT-Speaker and the UpDown model (Shuster et al., 2019) for the different personality traits with images on the development set.