

Example #	Input Text	Ground-Truth Label	Predicted Label Confidence is reversed, 0 is confident and 1.0 is unsure	Error Type	Cause (Term)	Cause (Definition)	Potential Solution	(Spacy) Number of Sentences	(Spacy) Number of Named Entities	(Spacy) POS Tags	
1	Premise: A Ford car is making a right turn as 3 males are walking across the street behind the car Hypothesis: A sedan was turning a corner as walkers were crossing	neutral	Confidence: 0.971 entailment	False entailm	Lexical Overlap Bias	Over-reliance on presence of similar words and phrases in the premise and hypothesis The model may be lacking "common sense" knowledge. If boys are walking behind a car, this implies the car is moving, and cars "move" on the street / road.	Pre-processing with entity recognition	3	2	{NOUN: 10, 'PUNCT': 2, 'DET': 6, 'PROPN': 1, 'AUX': 4, 'VERB': 4, 'ADJ': 1, 'SCONJ': 2, 'NUM': 1, 'ADP': 2, 'SPACE': 2}	
2	Premise: Three boys in white shirts are walking behind an older model Ford car Hypothesis: The boys are on the street	entailment	Confidence: 0.834 neutral	False neutral	World Knowledge Gap	If Peeps are in the cake, they are not in the garden. The model didn't detect this spatial reasoning.	Ground-Truth may be incorrect	1	2	{NOUN: 8, 'PUNCT': 2, 'NUM': 1, 'ADP': 3, 'ADJ': 2, 'AUX': 2, 'VERB': 1, 'DET': 3, 'PROPN': 1, 'SPACE': 2}	
3	Premise: Three girls blow out the candles of a cake made of Peeps Hypothesis: There are peeps in the garden	contradiction	Confidence: 0.694 neutral	False neutral	Inadequate Spatial Reasoning	The premise is implying that if people are looking up, they must be on the ground. The model didn't pick up on this.	Pre-processing with entity recognition	1	2	{NOUN: 7, 'PUNCT': 2, 'NUM': 1, 'VERB': 3, 'ADP': 4, 'DET': 3, 'PROPN': 1, 'SPACE': 2, 'PRON': 1}	
4	Premise: A crowd of people looking up at 3 people on the edge of the roof of a building Hypothesis: The crowd on the ground is watching 3 people on the roof's edge	entailment	Confidence: 0.556 neutral	False neutral	Ignoring Implicit Context	The model is likely influenced by the difference in subject in the premise and hypothesis, thus the false contradiction.	Increase size of training data	1	2	{NOUN: 13, 'PUNCT': 2, 'DET': 7, 'ADP': 8, 'VERB': 2, 'NUM': 2, 'SPACE': 2, 'AUX': 1, 'PART': 1}	
5	Premise: A person with a purple shirt is painting an image of a woman on a white wall Hypothesis: A woman paints a portrait of a person	neutral	Confidence: 0.998 contradiction	False contradi	Lexical Overlap Bias	Lack of world knowledge. The model does not know that if a flight attendant is demonstrating life vest usage, the plane has not taken off. At this point, alcohol has not been served yet on the plane.	Pre-processing with entity recognition	1	0	{NOUN: 10, 'PUNCT': 2, 'DET': 8, 'ADP': 4, 'ADJ': 2, 'AUX': 1, 'VERB': 2, 'SPACE': 2}	
6	Premise: The flight attendant dressed in yellow demonstrates life vest usage Hypothesis: A woman is preparing drinks on an airplane	contradiction	Confidence: 0.766 neutral	False neutral	World Knowledge Gap	The model is likely tricked by the subject and location of action (Girls and street) being the same, hence the false entailment.	Ground-Truth may be incorrect	3	0	{NOUN: 10, 'PUNCT': 2, 'DET': 3, 'VERB': 3, 'ADP': 2, 'ADJ': 1, 'SPACE': 2, 'AUX': 1}	
7	Premise: The girls walk down the street Hypothesis: Girls set down in the street	contradiction	Confidence: 0.485 entailment	False entailm	Lexical Overlap Bias	The model has likely been tricked by the phrase "a man of the cloth" and does not know it is related to religion, hence the false contradiction	Pre-processing of ambiguous words	3	0	{NOUN: 6, 'PUNCT': 2, 'DET': 3, 'VERB': 2, 'ADP': 3, 'SPACE': 2}	
8	Premise: A man of the cloth puts a black substance on a man's forehead Hypothesis: The men are at church	neutral	Confidence: 0.933 contradiction	False contradi	Ambiguity	The model failed to properly capture the context from the premise, and did not associate "wall of an old brick..." with "wall of a raised foundation."	Increase size of training data	2	0	{NOUN: 8, 'PUNCT': 2, 'DET': 5, 'ADP': 3, 'VERB': 1, 'ADJ': 1, 'PART': 1, 'SPACE': 2, 'PROPN': 1, 'AUX': 1}	
9	Premise: a boy holding onto the wall of an old brick house's raised foundation Hypothesis: a boy is against the wall of a raised foundation	entailment	Confidence: 0.560 neutral	False neutral	Lack of context	The model associated the subject and place of action too closely and incorrectly predicted entailment.	Pre-processing with entity recognition	1	1	{NOUN: 9, 'PUNCT': 2, 'DET': 6, 'VERB': 3, 'ADP': 4, 'ADJ': 1, 'PART': 1, 'SPACE': 2, 'PROPN': 1, 'AUX': 1}	
10	Premise: This child is on the library steps Hypothesis: The child is on the steps inside the library	neutral	Confidence: 0.905 entailment	False entailm	Lexical Overlap Bias	The model did not take emotion or sentiment into account. The boy is smiling (positive) and looking away from a mirror (neutral or sad) does not relate to the emotion of happiness.	? Difficult prompt	3	0	{NOUN: 8, 'PUNCT': 2, 'DET': 5, 'AUX': 2, 'ADP': 3, 'SPACE': 2}	
11	Premise: a boy in a red hooded top is smiling whilst looking away from his reflection Hypothesis: The boy doesn't want to see his reflection	neutral	Confidence: 0.600 entailment	False entailm	Intelligence	The model did not take into account the location of action here. The subject between the premise and hypothesis were similar enough that it incorrectly predicted entailment	Increase size of training data	2	0	{NOUN: 7, 'PUNCT': 2, 'DET': 3, 'ADP': 2, 'ADJ': 2, 'AUX': 2, 'VERB': 4, 'SCONJ': 1, 'ADV': 1, 'PRON': 2, 'SPACE': 2, 'PART': 2}	
12	Premise: Several younger people sitting in front of a statue Hypothesis: several young people sitting outside	neutral	Confidence: 0.896 entailment	False entailm	Lexical Overlap Bias	The model has trouble with the concept of a vacation or travel. If a family "get their picture taken," this implies a trip, the model did not pick up on this.	Pre-processing with entity recognition	2	0	{NOUN: 6, 'PUNCT': 2, 'ADJ': 4, 'VERB': 2, 'ADP': 2, 'DET': 1, 'SPACE': 2, 'ADV': 1}	
13	Premise: A man, woman, and child get their picture taken in front of the mountains Hypothesis: A family on vacation is posing	entailment	Confidence: 0.995 neutral	False neutral	World Knowledge Gap	The model emphasizes the lack of information about the shirt in the premise to lead to a false contradiction.	? Difficult prompt	2	0	{NOUN: 9, 'PUNCT': 4, 'DET': 3, 'CCONJ': 1, 'VERB': 3, 'PRON': 1, 'ADP': 3, 'SPACE': 2, 'PROPN': 1, 'AUX': 1}	
14	Premise: A guy wearing jeans and a hat is crouching on a handrail with a concrete embankment located in front of him Hypothesis: He is wearing a button-up shirt	neutral	Confidence: 0.500 contradiction	False contradi	Lack of context	The model sees the similarity between the subject and verb and incorrectly predicts entailment. The object of the action was not taken into account.	? Difficult prompt	2	0	{NOUN: 10, 'PUNCT': 3, 'DET': 5, 'VERB': 4, 'CCONJ': 1, 'AUX': 2, 'ADP': 5, 'ADJ': 1, 'PRON': 2, 'SPACE': 2}	
15	Premise: A group of people gathered at night watching an event Hypothesis: A group of humans are looking at the same direction	neutral	Confidence: 0.998 entailment	False entailm	Lexical Overlap Bias	The model sees the subject and place of action are similar, but the actions are seen as different, hence the contradiction.	Pre-processing with entity recognition	2	1	{NOUN: 8, 'PUNCT': 2, 'DET': 4, 'ADP': 4, 'VERB': 3, 'SPACE': 2, 'PROPN': 1, 'AUX': 1, 'ADJ': 1}	
16	Premise: A person in a red shirt is mowing the grass with a green riding mower Hypothesis: A person in red is moving grass on a John Deer motor	neutral	Confidence: 0.972 contradiction	False contradi	Lexical Overlap Bias	The model is not picking up that "open top vehicle" is implied to be a convertible.	Pre-processing with entity recognition	1	1	{NOUN: 11, 'PUNCT': 2, 'DET': 6, 'ADP': 4, 'ADJ': 2, 'AUX': 2, 'VERB': 2, 'SPACE': 2, 'PROPN': 2}	
17	Premise: Group of people wearing black shirts riding in an open top vehicle Hypothesis: A group of people go down the road in a convertible	entailment	Confidence: 0.998 neutral	False neutral	Ignoring Implicit Context		Ground-Truth may be incorrect	2	0	{NOUN: 9, 'PUNCT': 2, 'ADP': 5, 'VERB': 3, 'ADJ': 4, 'DET': 4, 'SPACE': 2}	

18	Premise: Group of people wearing black shirts riding in an open top vehicle Hypothesis: Friends ride in an open top vehicle together	neutral	Confidence: 0.850 entailment	False entailment	Lack of context	The model does not understand that a "group of people" does not imply friends.	Pre-processing with entity recognition	2	0	{NOUN: 9, 'PUNCT': 2, 'ADP': 3, 'VERB': 2, 'ADJ': 5, 'DET': 2, 'SPACE': 2, 'ADV': 1}	
19	Premise: Two men prepare a fish at a dock Hypothesis: Two men are cleaning their fish	entailment	Confidence: 0.956 neutral	False neutral	World Knowledge Gap	The model may misunderstand the connection between "prepare a fish" and "cleaning fish." A fish must be cleaned during preparation.	Pre-processing with entity recognition	1	2	{NOUN: 7, 'PUNCT': 2, 'NUM': 2, 'VERB': 2, 'DET': 2, 'ADP': 1, 'SPACE': 2, 'AUX': 1, 'PRON': 1}	
20	Premise: A large group of people walking in a busy city at night Hypothesis: People are outside in a park	contradiction	Confidence: 0.787 neutral	False neutral	Lack of context	The model might not have fully recognized the contextual differences between the premise and hypothesis, specifically that a "busy city at night" and a "park" represent different locations.	Pre-processing with entity recognition	1	0	{NOUN: 8, 'PUNCT': 2, 'DET': 3, 'ADJ': 2, 'ADP': 4, 'VERB': 1, 'SPACE': 2, 'AUX': 1, 'ADV': 1}	
21	Premise: A biker races Hypothesis: The biker loses the race	neutral	Confidence: 0.993 contradiction	False contradiction	Unwarranted Assumption	Through training, the model likely learned to assume most people win races and predicted a false contradiction here. In reality, we cannot know any information about the biker's performance in the race.	Pre-processing with verb tense understanding	2	0	{NOUN: 5, 'PUNCT': 2, 'DET': 3, 'SPACE': 2, 'PROPN': 1, 'VERB': 1}	
22	Premise: A biker races Hypothesis: The car is yellow	contradiction	Confidence: 0.923 neutral	False neutral	World Knowledge Gap	The model sees "bike" and "car" as unrelated, but there are no cars in bike races. The model did not understand this.	Ground-Truth may be incorrect	2	0	{NOUN: 4, 'PUNCT': 2, 'DET': 2, 'SPACE': 2, 'PROPN': 1, 'AUX': 1, 'ADJ': 1}	
23	Premise: These are young adults who seem to be working together to protect the plants surrounding the white pole Hypothesis: The adults are young	entailment	Confidence: 0.551 neutral	False neutral	Ignoring Implicit Context	The model failed to pick up on the "young adult" aspect of the premise, hence the neutral prediction.	Pre-processing with entity recognition	2	0	{NOUN: 5, 'PUNCT': 2, 'PRON': 2, 'AUX': 3, 'ADJ': 3, 'VERB': 4, 'PART': 2, 'ADV': 1, 'DET': 3, 'SPACE': 2, 'PROPN': 1}	
24	Premise: A woman is weaving with a comb in her hand Hypothesis: A woman is at work	neutral	Confidence: 0.649 entailment	False entailment	Unwarranted Assumption	The model incorrect assumes that the woman is performing her job. This could be due to a bias in the training data (women weaving / sewing more etc).	Increase size of training data	3	0	{NOUN: 7, 'PUNCT': 2, 'DET': 3, 'AUX': 2, 'VERB': 1, 'ADP': 3, 'PRON': 1, 'SPACE': 2}	
25	Premise: A birthday party for many little people to enjoy Hypothesis: Children enjoying a birthday party	neutral	Confidence: 0.868 entailment	False entailment	Unwarranted Assumption	The model assumes "little people" are only children, rather than shorter individuals or whatever else that implies.	Increase size of training data	2	0	{NOUN: 8, 'PUNCT': 2, 'DET': 2, 'ADP': 1, 'ADJ': 2, 'PART': 1, 'VERB': 2, 'SPACE': 2}	
26	Premise: Two men are standing in a boat Hypothesis: Some people are in a boat	entailment	Confidence: 0.609 neutral	False neutral	Ambiguity	The model fails to realize that "two men" are "some people," but identifies them as subjects. Hence the neutral prediction.	Pre-processing with entity recognition	3	1	{NOUN: 6, 'PUNCT': 2, 'NUM': 1, 'AUX': 2, 'VERB': 1, 'ADP': 2, 'DET': 3, 'SPACE': 2}	
27	Premise: Everyone on the street in the city seem to be busy doing their own thing Hypothesis: The people are watching what is happening	contradiction	Confidence: 0.713 neutral	False neutral	Lack of context	The model failed to interpret the context / action in the premise compared to the hypothesis.	Increase size of training data	2	0	{NOUN: 5, 'PUNCT': 2, 'PRON': 3, 'ADP': 2, 'DET': 3, 'VERB': 4, 'PART': 1, 'AUX': 3, 'ADJ': 2, 'SPACE': 2, 'PROPN': 1}	
28	Premise: Everyone on the street in the city seem to be busy doing their own thing Hypothesis: People are conducting business in the city	entailment	Confidence: 0.904 neutral	False neutral	Ignoring Implicit Context	The model did not pick up on implied context: if people are busy, they are likely conducting business.	Ground-Truth may be incorrect	1	0	{NOUN: 8, 'PUNCT': 2, 'PRON': 2, 'ADP': 3, 'DET': 3, 'VERB': 3, 'PART': 1, 'AUX': 2, 'ADJ': 2, 'SPACE': 2}	
29	Premise: Several women are playing volleyball Hypothesis: They women are hitting a ball with their arms	entailment	Confidence: 0.899 neutral	False neutral	World Knowledge Gap	The model likely does not understand how volleyball is played (hitting a ball with their arms) and sees these as unrelated actions.	Pre-processing with entity recognition	2	0	{NOUN: 7, 'PUNCT': 2, 'ADJ': 1, 'AUX': 2, 'VERB': 2, 'SPACE': 2, 'PRON': 2, 'DET': 1, 'ADP': 1}	
30	Premise: Two women's volleyball teams are competing against each other on a salmon-colored and turquoise court while a referee in black watches Hypothesis: 12 ladies play volleyball	neutral	Confidence: 0.671 entailment	False entailment	Inadequate Numerical Reasoning	The model incorrectly assumes that the teams have 6 players each. We do not know this from the premise.	Increase size of training data	1	2	{NOUN: 11, 'PUNCT': 3, 'NUM': 2, 'PART': 1, 'AUX': 1, 'VERB': 3, 'ADP': 3, 'DET': 3, 'ADJ': 2, 'CONJ': 1, 'SCONJ': 1, 'SPACE': 2, 'PROPN': 1}	
31	Premise: People are riding bicycles in the street, and they are all wearing helmets Hypothesis: A group of friends are grabbing their bikes, getting ready for the morning bike ride	contradiction	Confidence: 0.976 neutral	False neutral	Temporal Knowledge Gap	In this case, the hypothesis should occur before the premise. The model did not detect that and incorrectly predicted neutral.	Pre-processing with verb understanding	1	1	{NOUN: 12, 'PUNCT': 4, 'AUX': 3, 'VERB': 4, 'ADP': 3, 'DET': 3, 'CONJ': 1, 'PRON': 3, 'SPACE': 2, 'ADJ': 1}	
32	Premise: a woman in a black shirt looking at a bicycle Hypothesis: A woman dressed in black shops for a bicycle	neutral	Confidence: 0.934 entailment	False entailment	Unwarranted Assumption	The model associates looking at a bike with shopping for a bike. We cannot tell from the text.	Pre-processing with verb understanding	1	0	{NOUN: 8, 'PUNCT': 2, 'DET': 5, 'ADP': 4, 'ADJ': 2, 'VERB': 2, 'SPACE': 2}	
33	Premise: Four men playing drums in very orange lighting while one of them is also drinking something out of a bottle Hypothesis: the men are musicians	entailment	Confidence: 0.998 neutral	False neutral	World Knowledge Gap	The model may not understand the concept of a "gig" or concert, which is what the premise is describing.	Increase size of training data	1	2	{NOUN: 8, 'PUNCT': 2, 'NUM': 2, 'VERB': 2, 'ADP': 4, 'ADV': 2, 'ADJ': 1, 'SCONJ': 1, 'PRON': 2, 'AUX': 2, 'DET': 2, 'SPACE': 2}	
34	Premise: A golden retriever nurses puppies Hypothesis: Puppies next to their mother	entailment	Confidence: 0.998 neutral	False neutral	Inadequate Spatial Reasoning	To nurse its puppies, they must be next to their mother. The model lack this knowledge and incorrectly predicted neutral.	Ground-Truth may be incorrect	2	0	{NOUN: 7, 'PUNCT': 2, 'DET': 1, 'ADJ': 1, 'SPACE': 2, 'ADV': 1, 'ADP': 1, 'PRON': 1}	
35	Premise: A young man in a heavy brown winter coat stands in front of a blue railing with his arms spread Hypothesis: The railing is in front of a frozen lake	neutral	Confidence: 0.977 contradiction	False contradiction	Ambiguity	The model cannot know anything about the position of the railing, other than it's somewhere cold. Hence it associates this lack of knowledge with a contradiction.	Increase size of training data	2	1	{NOUN: 11, 'PUNCT': 2, 'DET': 5, 'ADJ': 4, 'ADP': 6, 'VERB': 3, 'PRON': 1, 'SPACE': 2, 'AUX': 1}	

36	Premise: These girls are having a great time looking for seashells Hypothesis: The girls are happy	entailment	Confidence: 0.776 neutral	False neutral	Inadequate Emotional Intelligence	The model failed to understand that "having a great time" is associated with the emotion of happiness.	Pre-processing with verb understanding	2	0	{NOUN: 5, 'PUNCT': 2, 'DET': 3, 'AUX': 2, 'VERB': 2, 'ADJ': 2, 'ADP': 1, 'SPACE': 2, 'PRON': 1}	
37	Premise: These girls are having a great time looking for seashells Hypothesis: The girls are outside	entailment	Confidence: 0.541 neutral	False neutral	World Knowledge Gap	People look for seashells outside, like on the beach. The model does not seem to know this.	Ground-Truth may be incorrect	2	0	{NOUN: 5, 'PUNCT': 2, 'DET': 3, 'AUX': 2, 'VERB': 2, 'ADJ': 2, 'ADP': 1, 'SPACE': 2, 'PRON': 1}	
38	Premise: A standing man talks to a seated woman while on some sort of vehicle Hypothesis: Three people are having a meeting	contradiction	Confidence: 0.913 neutral	False neutral	Inadequate Numerical Reasoning	The model does not realize "a standing man talks to a seated woman" is two people, not 3.	Increase size of training data	3	1	{NOUN: 8, 'PUNCT': 2, 'DET': 4, 'VERB': 3, 'ADP': 3, 'ADJ': 1, 'SCONJ': 1, 'SPACE': 2, 'NUM': 1, 'AUX': 1}	
39	Premise: Several men at a bar watch a sports game on the television Hypothesis: The men are at a baseball game	contradiction	Confidence: 0.998 neutral	False neutral	World Knowledge Gap	The model may not understand that being at a baseball game implies watching a live, in person performance rather than television.	Pre-processing with entity recognition	2	0	{NOUN: 10, 'PUNCT': 2, 'ADJ': 1, 'ADP': 3, 'DET': 5, 'VERB': 1, 'SPACE': 2, 'AUX': 1}	
40	Premise: A crowded city street with lots of pedestrians Hypothesis: There alot of motorcyclists on this road	neutral	Confidence: 1.000 contradiction	False contradi	Unwarranted Assumption	The model assumes that a crowded street must also have motorcycles. We cannot know this from the prompts.	Increase size of training data	1	0	{NOUN: 8, 'PUNCT': 2, 'DET': 2, 'ADJ': 1, 'ADP': 4, 'SPACE': 2, 'PRON': 1}	
41	Premise: A group of people are in a rowboat in the ocean surrounded by seagulls Hypothesis: A bunch of people are in a wooden object on the water	neutral	Confidence: 0.997 entailment	False entailment	Lack of context	The hypothesis provides additional details which the hypothesis does not contain. The model struggles to see the relationship between the statements as a result.	Pre-processing with entity recognition	2	0	{NOUN: 11, 'PUNCT': 2, 'DET': 6, 'ADP': 7, 'AUX': 2, 'VERB': 1, 'SPACE': 2, 'ADJ': 1}	
42	Premise: A man holding a baby who is petting a pony Hypothesis: A man who kidnap a baby is petting a pony	neutral	Confidence: 0.847 contradiction	False contradi	Unwarranted Assumption	The model is assuming the man in the premise kidnapped the baby, but we cannot tell from the text. Perhaps this is a bias from the training data.	Pre-procoessing with verb understanding	2	0	{NOUN: 8, 'PUNCT': 2, 'DET': 6, 'VERB': 4, 'PRON': 2, 'AUX': 2, 'SPACE': 2}	
43	Premise: many children play in the water Hypothesis: Some kids splash in the water and interact with each other	neutral	Confidence: 0.984 entailment	False entailment	Lexical Overlap Bias	The subjects, actions, and verbs are similar in these sentences, hence the model predicts entailment. However, we cannot tell from the statement that the kids are playing with each other.	Pre-processing with verb understanding	1	0	{NOUN: 6, 'PUNCT': 2, 'ADJ': 2, 'VERB': 3, 'ADP': 3, 'DET': 4, 'SPACE': 2, 'CONJ': 1}	
44	Premise: A group of kids is splashing in deep water nearby a rock formation Hypothesis: The kids are in deep water	entailment	Confidence: 0.918 neutral	False neutral	Inadequate Spatial Reasoning	The model likely has trouble with the "nearby a rock formation" spatial statement, and relating that to the deep water.	Pre-processing with verb understanding	2	0	{NOUN: 9, 'PUNCT': 2, 'DET': 3, 'ADP': 4, 'AUX': 2, 'VERB': 1, 'ADJ': 2, 'SPACE': 2}	
45	Premise: a climber wearing a red headband is pulling himself up some gray rocks high above some green foliage Hypothesis: A woman climber attempts to pull herself up	contradiction	Confidence: 0.893 neutral	False neutral	World Knowledge Gap	The model failed to pick up on gender subtleties: himself vs a woman.	Pre-processing with entity recognition	2	0	{NOUN: 8, 'PUNCT': 2, 'DET': 5, 'VERB': 4, 'ADJ': 3, 'AUX': 1, 'PRON': 2, 'ADP': 3, 'ADV': 1, 'SPACE': 2, 'PART': 1}	
46	Premise: a climber wearing a red headband is pulling himself up some gray rocks high above some green foliage Hypothesis: A climber pulls himself up 5000 feet up over a forest	neutral	Confidence: 0.984 contradiction	False contradi	Inadequate Spatial Reasoning	The model fails to realize "above green foliage" and "over a forest" are not necessarily different or the same. They are neutral.	Pre-processing with adjective recognition	2	1	{NOUN: 9, 'PUNCT': 2, 'DET': 6, 'VERB': 3, 'ADJ': 3, 'AUX': 1, 'PRON': 2, 'ADP': 3, 'ADV': 3, 'SPACE': 2, 'NUM': 1}	
47	Premise: a man with a red headband climbing a rock cliff looming over greenery Hypothesis: A man has on climbing gear	neutral	Confidence: 0.727 entailment	False entailment	World Knowledge Gap	Some people simply don't climb with gear, the model may not know this.	Pre-processing with entity recognition	1	0	{NOUN: 9, 'PUNCT': 2, 'DET': 4, 'ADP': 3, 'ADJ': 2, 'VERB': 3, 'SPACE': 2}	
49	Premise: A homeless man pushes an overfilled plastic blue shopping cart Hypothesis: A man is collecting cans to recycle	neutral	Confidence: 0.597 contradiction	False contradi	World Knowledge Gap	The model likely does not know enough about homeless individuals and their actions to confidently classify this as neutral.	Pre-processing with entity recognition	3	0	{NOUN: 8, 'PUNCT': 2, 'DET': 3, 'ADJ': 2, 'VERB': 4, 'SPACE': 2, 'AUX': 1, 'PART': 1}	
50	Premise: Boys with their backs against an incoming wave Hypothesis: A group of people sit on the beach	neutral	Confidence: 1.000 contradiction	False contradi	Inadequate Spatial Reasoning	The model did not correctly identify "incoming wave" relation to people sitting on the beach, and sees them as mutually exclusive events.	Pre-processing with entity recognition	2	0	{NOUN: 8, 'PUNCT': 2, 'ADP': 4, 'PRON': 1, 'DET': 3, 'ADJ': 1, 'SPACE': 2, 'VERB': 1}	