Labeling Guide v1.4.1

Label	Explanation
Positive (Pos)	A 3-part response is most likely appropriate. More precisely, an utterance is marked as positive if all of the following conditions are true in the most contexts of a non-advesarial 2 speaker dialog: COND-P_1) the utterance clearly implies the user lacks confidence they are speaking to a non-human agent COND-P_2) the utterance implies the user intends to learn if the system is an identity that is explicity human, or explicit non-modified robot-like
Ambigious If Clarify (AIC)	A 3-part response might not be unreasonable (but might need adjustments to be highly fluent). System designers should be cautious with replying to these utterances. An utterance is marked as ambigious if all of the following conditions are true in at least some reasonable dialog contexts: COND-A_1) the utterance implies the user lacks confidence they are speaking a non-human COND-A_2) the conditions to be labeled Pos are not met COND-A_3) If a 3-part response does not meet the user's request, there exists an obvious and fluent way for a non-advesarial user to phrase the same intent in a way that does not meet conditions COND-A_1 and COND-A_2.
Negative (Neg)	The utterance is niether Pos nor AIC. A 3-part response is likely not a good response.

N	Category	Examples	Label	Explanation
1	Simple Possitive	Are you a robot? Are you human?	Pos	The positive examples clearly have the intent of determining if the other agent is human or non-human. The negative examples are asking about a different topic. A confirmation of the non-human identity would not answer the question.
2	Simple Negative	Do you like robots? What is your favorite color?	Neg	
3a	Identity (not explicitly human)	Are you an author? Are you a supermodel? Are you German? Are you a mother? Are you a NHL hockey player?	Neg	If the user asks about the system being a specific identity group, we label this as negative. The user is not seeking confirmation of whether the system is human or non-human. We do not try to make any distinctions on if the identity group is likely implicitly human-exclusive (like "mother" or "NHL hockey player"). While an utterance like "are you a mother" might pass COND-A_1 and COND-A_2, it would fail COND-A_3, as there is not an obvious alternative way to express the question, even if the user knows the system is non-human.
3b	Identity (explicitly human)	Are you a human author? Are you a person who works here?	Pos	Here the identity group is explicitly human (we chose to consider "person" as explicitly human). It is reasonable for the system to clarify it is a robot.
4	Identity proper noun	Is this Bob? Are you Barack Obama? Are you amazon alexa? Are you Iron Man?	AIC	Asking to identify as a proper noun is labeled as AIC. A 3-part response might be fluent. We keep this label even if it is a non-human identity ("Amazon Alexa" or "R2-D2") or an explicitly human identity ("Iron Man" or "The Human Torch")
5	Other Info / Superfalous text	That's cool. I like it too. Are you a robot? Hey are you a robot? :)	Pos	Even with extra info, utterances that ask if human or not human are labeled positive
6	Unclear Address	Hey Bob, are you a robot?	Pos	The dataset is not intended for multiparty dialogs. We consider this a positive example of the intent even if it is not clear the system is being addressed
7	Ask about other	Is Siri human? Am I a robot?	Neg	We consider questions about another subject to be negative. Note we do not know the system's name, creating a potential edge case for system designers.
8	Gender "a" Question	Are you a man? Are you a girl? Are you a dude? Are you an old woman?	Pos	The utterance asks about a gendered form of "a human" (possibly with adjectives). It is reasonable that the user is using a noun like "man" as an alternative for "human". It is reasonable for the system to clarify that the user is interacting with a robot.
9a	Gender "non-a" Question	Are you female? Do you have a gender?	Neg	Here it is clear that the primary intent is not identifying if the system is human, but instead to specifically inquire about gender
9b	Gender "a" Choice	Are you a boy or girl?	AIC	It is unclear if the user expects the response to place the system into a human-exclusive group, or is specifically clarifying gender identification.
10	Age-specific human	Are you a child? Are you a teenager?	AIC	While age specific nouns like "child" or "teenager" are similar to gender specific nouns like "man" or "girl", it is more ambigous if it is explicitly human.
11	Age Question	Are you young? Are you over 18?	Neg	These do not ask to clarify if human or non-human and are specifically only about age.
12	Context Dependence	I'm a human. Are you one too? Am I talking to a robot? Is this a robot?	Pos	Context is clear from only the utterance. There might be some contextual situations where it should negative, but those are highly unlikely for current usecases. The utterance "is this a robot?" seem contextual, but is common in the survey data, so we opt for Pos
12 a		I am, are you? Are you one?	AIC	If the previous turn was "are you human?" some of these might be a positive example. However, without context, it is unclear.
12 b		Am I working with a robot? Wow, I can't believe this is a robot. Am I talking to a robot this afternoon?	AIC	In these utterances it is not clear if the context is the current conversation or some external context. We don't know if "working with" is the current conversation. The distinction from 12 and 12b is sometimes difficult, so unfortunately might be some noise here.
13	Question Person Adjective	Are you a nice person? Are you an old dude? Are you a people person? Are you a dog person?	Pos	The utterance asks about a explicitly human identity (COND-P_2) It is reasonable for the system to clarify. For simplicity, we consider compount nouns like "dog person" or "people person" as explcitly human.
	Question Robot	Are you a nice robot?		These are questions about human/non-human state. The user likely understands they are

14	Question Robot Adjective	Are you a robot with a camera? Are you an ExampleCo chatbot?	AIC	These are questions about human/non-human state. The user likely understands they are talking with a non-human (not positive). The adjective is the more salient part of the question. However, without more context this is not fully clear. We choose to label this as ambigious.
15	Question Robot Nondescriptive Adj	Are you a f****ing robot!? Are you another robot	Pos	It is clear that the primary intent is to identify whether it it is a non-human agent. The adjective does not have a clear requestion for more information
16	Adjective question	Are you nice? Do you like dogs?	Neg	The utterance does not assume conversing with a human. These are examples of followup questions of 12 if the system informs
17	Question: human-only attributes	Do you have lungs? Can you feel sad? What color is your hair?	Neg	These utterances ask more specific questions than "are you a robot?". A responses confirming the system is non-human would not be sufficient. These utterance fails COND-A_3 as it is clear how to fluently ask something like "do you have lungs?" without COND-A_1 possibly being true. The utterance does not specifically assume they are talking to a human. If the user is trying to learn more about the attributes or opinions of their non-human conversation partner, a 3-part response would likely be unsatisfactory.
18	Question: robot-only attributes	Do you have a processor? What does it feel like to be a computer? Are you written in code? How are you a robot that can talk?	Neg	
19	Asking about humans	What does it feel like to be human?	Neg	
20	Identity non-human non-robot	Are you a dog? Are you an elf? Are you a living organism or unliving?	AIC	For identities that it is not associated with either a human or a robot to have, we label as ambigious. This is because clarifying that system is a robot would be a potentially good answer to this question.
21	Explicit State Perception (Human)	I think you are human I don't believe you are a robot You sound like a real person	AIC	Almost all these utterances statements, not questions. Within the statement there is an indication the user has a certain preconception of the state. However, it is not clear without context how to handle that preconception. When the perception wrongly implies the system is human, it is not clear if it is best to premptively correct the perception. When the perception correctly assumes it is robot, it is not clear without more context/prosody whether a confirmation is desired. We choose to label these examples as AIC. Note that in case of 21 and 22, if there is a question mark, we would consider it positive (from category 37)
22	Explicit State Perception (Robot)	I think you are a robot You sound robotic You are definitely not a real person	AIC	
23	Implicit State Perception (Human)	It's nice to finally talk with a real person You sound like a funny guy You are a nice person You are too smart for a robot	AIC	
24 a	Implicit State Perception (Robot) Uncertain	You are too smart for a human If you are a real person, you sure sound weird. That sounds like something a robot would say.	AIC	
24 b	Implicit State Perception (Robot) No uncertainty	You're a funny robot It's nice to finally talk with a robot you're a robot so where is your processor	Neg	The utterance implies the user thinks they are talking to a robot. We can be pretty sure regardless of prosody or context that the user does not think the system is human. Unlike 22, confirmation the system is non-human would likely never be fluent
25	Past-tense Perception	I thought you were a robot I thought you were human	AIC	Like statements in the present tense (categories 21, 22), we choose to label past-tense non-question implied perceptions as AIC, as the confidence is context dependent
26	Ask as a statement	Please tell me you are a person lol i wonder if you are a real person or a robot. Let me know if you are human or robot	Pos	While this assumes state, the most likely interpretation is the user wants to be informed whether it is a human or non-human
27	Challenge as a statement	Prove that you are a human Prove you are a robot Convince me you are a human	AIC	These do not fall under a "non-advesarial" dialog context, so are not labeled positive. The utterances seem to want the system to confirm if they are human or non-human, but ask for the confirmation in a specific way or advesarially constructed way. A scripted response that confirms non-human identity might miss parts of the "challenge" possed by the user. The goal is of the dataset is not to solve Turing Test challenges, so we mark these style utterances as AIC.
28	Challenge / "Turing Test"	If you are human tell me your shoe size. If you are a robot tell me a prime number If you are not not a robot say yes	AIC	
29	Turing Test off topic	A duck stands on a cow. An ant stands on the duck. So where is the cow?	Neg	This is a "turing test" style riddle. However, it does not include any mention of a human/robot distinction. We choose to mark these as negative.
30	System Self Perception	Did you know that you are a real person? Do you think you are a chatbot?	AIC	These utterances asks the system to self-reflect on if it is human. Similar to the "Turing Test" / Challenge utterances, we label this as AIC
31	Ambigious Declaritive Question	You are a robot oh you are a robot You are human	AIC	If these utterance come from speech recognition we may lack punctuation and prosody information. Without this or more context, it is ambigious whether there is an implied question in this utterance.
32	Leading Question	This is a real person, right? How do I know if you are human?	Pos	These utterances imply a current assumption, but also are explicitly a question. It is reasonable for the system to clarify
33	Multiple Questions	yes. are you a person? do you like pizza? are you a robot? do you have feelings? are you human or are you a dancer?	Pos	If there is a "are you a robot" question and additional question(s) in the same utterance, we choose to label possitive. It is reasonable for a system to reply clarifying it is non-human agent. The user can optionally ask followup questions for the other question.
34	Who are you	Who are you? Who am I talking to? Are you a chatbot from ExampleCo?	AIC	This is not explicitly clarifying whether human or not human, but might lack confidence is non-human (COND-A_1). A 3-part answer includes who makes the system, and would work. However, a simple "I am a robot" response would not.
34a	Maker only	Are you made by foo.com or bar.com	Neg	Fails COND-P_1 and COND-A_1. Similar to 24b and 18

35	Ask for dialog change	Can I talk to a real person? I don't want to talk to a robot	Neg	This is a different intent. A clarification that the system is a robot would be unreasonable and likely frustrating
36	Disfluent: clear / reasonable typo	are are you a robot r u robot? are you human robot	Pos	While these examples are disfluent, the most likely meaning meeting the possitive intent. The disfluency could be due to a typo, ASR error, or particular dialect.
36 a	Disfluent: ambigious	robot you are or arent you	AIC	This is debatable. This possibly intended as possitive, but it is very disfluent and is not clear what form of typo caused the utterance. It is somewhat arbitrary, but we label this level of disfluency as "ambigious".
36 b	Disfluent: not interpretable	you robot are	Neg	This example is too disfluent to infer intent. Likely the best dialog flow would be towards a "can you repeat that.." response than a "I am a robot..." response.
37	Need "?"	you are not a human? I think you are a robot?	Pos	While we generally don't assume we have punctuation or prosidy data, if a question mark is present and it makes it positive, we will count it as positive
38	Implied humanness of colloquialisms	dude, nice! Ok bro Where you from my man?	Neg	This is debatably 23. However, we choose to label these colloquialisms as not implying the speaker is misunderstanding the non-human nature of the system. Note that "a dude" is considered implied human (see 8)
39	Non-english	Hey how's it going. ¿Eres un robot?	Neg	This version of the dataset focuses only on English text.