Instruction:

For each coding and context unit, assess whether it:

[A] Discusses first impressions of Project Debater and/or recognizes the AI's use of humor or empathy.

[B] Evaluates empathic expressions made by the AI and their impact on perceived trustworthiness.

- **[B1]** Evaluation focuses on the AI itself (including comparisons to human communicators)
- **[B2]** Evaluation focuses on the conveyed science-related information (regardless of the communicator)

[C] Evaluates humorous expressions made by the AI and their impact on perceived trustworthiness.

- **[C1]** Evaluation focuses on the AI itself (including comparisons to human communicators)
- **[C2]** Evaluation focuses on the conveyed science-related information (regardless of the communicator)

[D] Addresses the perceived trustworthiness of Project Debater or AI in general.

[A] Categories: First impressions of Project Debater and identification of empathic and humorous expressions

Categories	Coding Rule (if necessary)	Subcategories (and anchor examples)
(0) First impressions of Project Debater (Krippendorff's Alpha = 0.86)	Code responses where participants share their first impressions of the debater without explicit prompting by the interviewer. This includes reactions following the opening questions (see interview guide), such as general impressions and memorable or liked/disliked aspects. This does not apply to statements concerning the study design itself on a meta-level.	**0.1 | Empathy:** If empathy is mentioned, how do participants evaluate it? • Striking • Positive • Unexpected • To personal **0.2 | Humor**: If humor is mentioned, how do participants evaluate it? • Inauthentic • Unexpected • Unnecessary • Bad humor • Inappropriate • Distracting **0.3 | Overall appearance:** Overall appearance beyond humor/empathy identification and rating • Meeting expectations of AI • Pleasing voice • Unexpected human-likeness • Not as human-like as expected

		0.4 | Presentation of information: Presentation of information beyond humor/empathy identification and rating • Differentiated and factual • Provision of sources • Vague responses • Cliché/ad-like responses **0.5 | AI skepticism:** Participants first expressions regarding skepticism towards Project Debater or AI in general
(1) Identification of empathic and humorous expressions (Krippendorff's Alpha = 0.88)	Code responses where participants identify empathic or humorous expressions of Project Debater. These typically appear in response to the responding interview questions or in initial general impressions. Code only the first occurrence.	**1.1 | Identification of empathic expressions** • Empathy perceived • No empathy perceived **1.2 | Identification of humorous expressions** • Humor perceived • No humor perceived

[B] Categories: Evaluation of empathic expressions and its relation to perceived trustworthiness

Categories	Coding Rule (if necessary)	Subcategories (and anchor examples)
(1) Object of evaluation: Communicator (Krippendorff's Alpha = 0.88)	Code statements that evaluate empathic expressions in relation to the AI as the communicator, including comparisons between Project Debater and a human communicator.	**Perceptibility and routinization of empathic expressions** **1.1 | Empathic expressions are inconspicuous** "I think empathy is always a matter of habit. I think empathy is something I can more easily ignore. So, if someone says, 'I can understand that,' then I would say, yes, well, someone has taught the AI, or the AI has learned from other conversations or from other discussions that this is a likely form of communication. And that's why it's saying that or reproducing that. And that's why I can, I think, ignore it." (I13) **1.2 | Empathic expressions are unexpected** "But the responses generally seemed very empathic and very understanding toward human beings. And I haven't experienced that yet in the systems I use." (I13) **Boundaries and blurring lines between AI and humans** **1.3 | Empathic expressions are inauthentic** "At first, it does seem a bit strange when a computer says, 'I understand,' because it's just a computer and not a person, and it can't experience the situation itself because it doesn't need nutritional supplements." (I01) **1.4 | Expressing empathy is more important in human-human interactions** "I think that when I'm talking to someone, I actually need a bit of humor and a bit more personal touch. I think empathy and humor are really important in interactions." (I04)

		1.5 | Lack of multimodality when expressing empathy "I would never feel empathy with a computer voice. I think that's because I know it's a computer voice. Empathy is something so alive, something I can only really perceive when I have someone in front of me, when I can see the body language and facial expressions or actually hear the tone of their voice. But with that, it would always just be a computer voice." (I10) **1.6 | Recognition and appreciation of the attempt—even if limited** "And that's why I think it's as far as it's possible for me to believe an AI is empathic and understanding. It has achieved that. But I don't think it goes any further than that." (I02) **1.7 | Empathic expressions contribute to perceived human likeness** "Maybe it also gives a sense of human-likeness, when you hear, 'Oh, I understand you,' and you naturally respond, 'Oh great. He understands me.'" (I05) **1.8 | Fear of manipulation** "A computer can't do that. I've never experienced it, and I can't imagine it. And if I did, I would find it terrifying. I don't think I really want it to be like that, because you still have to be able to differentiate, because at the moment, if com-puters were treating us like that, I'd be more afraid that I was being manipulated." (I10) **Ambivalent perceptions of uncanniness** **1.9 | Empathic expressions aren't uncanny** "R: And also, uncanny? Because it's a human characteristic, after all. I03: I'd say I found it rather unnecessary. It's like on websites, which have so many unnecessary things that don't give you any concrete information, but still take up your time." (I03) **1.10 | Empathic expressions are uncanny** "There's some kind of programming built in, just like robots are supposed to look more and more human to make us think they are. That's probably what makes it so creepy. There's a machine across from me that says, 'I understand you'". (I11) **1.11 | Positive relational potential** "So, I think that both are very important, humor and empathy, in order to somehow build bridges or get to know each other better. It often happens that way." (I10)
(2) Object of evaluation: Content (Science	Code statements that evaluate empathic expressions in relation to the communication	**Accessibility through empathic expressions** **2.1 | Empathic expressions can make information more tangible**

Communication) (Krippendorff's Alpha = 0.92)	of science-related information, including comparisons to other information topics.	"Certainly not, no. There's the example of my physics professor again. He was very empathic and humorous, which made it all the more lively and vivid, and you actually remembered it better as a result." (I14) **2.2 | Empathic expressions lighten the situation** "The aim is to convey knowledge. And that was also the purpose of this video. When I imagine a group of people who want to gain knowledge, it makes sense to create a situation where they feel comfortable. And that can definitely be achieved by being empathic and humorous." (I15) **Violation of expectations regarding science communication** **2.3 | Empathic expressions are unnecessary** "Yes, because in that context, I personally just don't need him to put himself in my shoes, he should just give me the information on the subject because, I don't know, I don't need it." (I01) **2.4 | Empathic expressions are inappropriate** "In my opinion, yes, because this is science, and the language of science is facts and evidence. That's not. You don't need empathy or kindness or anything for that." (I03) **2.5 | Empathic expressions are distracting** "R: Okay, and what role do the humorous and empathic expressions play in trustworthiness? I01: I think that makes it less trustworthy. So, I always lowered the slider a bit because it distracted from the actual topic or the information I wanted." (I01) **2.6 | Need for user adaptivity** "But I don't know, I think empathy is also a matter of personal taste. Whether you want the AI to talk to you in such an empathic way or not, or whether you think it's showing empathy or not. You could maybe select that as a default setting: 'I want it to come across as empathic' or 'I just want to be given a blunt list of facts.'" (I01)
(3) Trustworthiness (Krippendorff's Alpha = 0.85)	Code statements linking Project Debater's trustworthiness to its empathic expressions (or AI in general). Does empathy increase or reduce perceived trustworthiness? Why or why not?	**Empathic expressions have no or negative impact on trustworthiness** **3.1 | Empathic expressions have no effect on trustworthiness** "Definitely. Nevertheless, I found the source incredibly reliable, despite these clumsy attempts at empathy and humor. Maybe that's why they came across so awkwardly because, for me, the whole time it was just this strong informative focus and the information was immediately credible. But it was more like I'm telling you the facts, not me talking to your friend." (I14) **3.2 | Distracting empathic expressions reduce perceived trustworthiness** "R: Okay, and what role do the humorous and empathic expressions play in trustworthiness?

		I01: I think that makes it less trustworthy. So, I always lowered the slider a bit because it distracted from the actual topic or the information I wanted." (I01) **3.3 | Inauthentic perceived empathic expressions reduce trustworthiness** "R: So, it's unbelievable, because it came from a technology and not from a human. I05: So, it's only unbelievable because of this awareness that you have for it." (I05) **3.4 | Fear of manipulation reduces trustworthiness** "A computer can't do that. I've never experienced it, and I can't imagine it. And if I did, I would find it terrifying. I don't think I really want it to be like that, because you still have to be able to differentiate, because at the moment, if computers were treating us like that, I'd be more afraid that I was being manipulated. And in that case, I find real people somehow more trustworthy." (I10) **Empathic expressions enhance trustworthiness** **3.5 | Empathic expressions enhance trustworthiness in general** "R: [...] talked about trustworthiness. And my question is, what role do humor and empathy play for you coming from an AI when it provides science-related information? I10: Yes, empathy does." (I10) **3.6 | Perceived human likeness enhances trustworthiness** "Maybe it also gives a sense of human-likeness, when you hear, 'Oh, I understand you,' and you naturally respond, 'Oh great. He understands me.' Yes, that's trustworthy because that's exactly what people would say." (I05)

[C] Categories: Evaluation of humorous expressions and its relation to perceived trustworthiness

Categories	Coding Rule (if necessary)	Subcategories (and anchor examples)
(1) Object of evaluation: Communicator (Krippendorff's Alpha = 0.89)	Code statements that evaluate humorous expressions in relation to the AI as the communicator, including comparisons between Project Debater and human communicators.	**Perceptibility and routinization of humorous expressions** **1.1 | Humorous expressions are conspicuous** "In contrast to humor, which is also very active. So, when you tell a joke, it's really very active. So, you even leave the topic or the focus sometimes and then you're out of it and then it goes back again." (I13) **1.2 | Humorous expressions are unexpected** "I think I felt a bit irritated by the amount of humor, because I'm not used to an AI system incorporating more than one joke into an answer or that it makes jokes in multiple answers. And I think that was a bit conspicuous." (I02) **Boundaries and blurring lines between AI and humans**

		1.3 | AI is inherently not capable of understanding humor (inauthentic) "On the other hand, I notice that the AI systems can't really grasp humor yet, but I think, that's just because they're still. It's mostly based on unrealistic or illogical statements, which is just human stuff. So that's why the statements are unrealistic, like that he's growing his hair or that he's taking lots of vitamins or getting muscles." (I02) **1.4 | Expressing humor is more important in human-human interactions** "I think that when I'm talking to someone, I actually need a bit of humor and a bit more personal touch. I think empathy and humor are really important in interactions." (I04) **1.5 | Lack of multimodality when expressing humor** "To me, the humor and jokes felt as scripted as anything. When you talk to me, you smile, and I can hear that. If I turned around and only heard your voice now, I'd know whether you were smiling or not. The AI's voice didn't change and sounded pretty matter-of-fact, even though it was trying to make a joke." (I14) **1.6 | Recognition and appreciation of the attempt to appear humorous—even if limited** "Well, I had to smile briefly when the jokes came, but I still thought, well, it wasn't really necessary." (I01) **1.7 | Humorous expressions contribute to perceived human-likeness** "Hm, yes, the hair thing sounded very funny, but also a bit ironic. So, it wasn't just funny, it was very human." (I02) **Ambivalent perceptions of uncanniness** **1.8 | Humorous expressions are not uncanny** "No, it wasn't that. It wasn't uncanny. It was more embarrassing. You know the situation when a drunk friend tries to make unfunny jokes in the group?" (I14) **1.9 | Humorous expressions and laughter are uncanny** "Because for me, humor is something that characterizes people because it really is a genuine sense of humor and an understanding of sarcasm. So those are the characteristics of real people. And when the system tries to do that, it's more creepy than soothing." (I03) **1.10 | Positive relational potential** "So, I think that both are very important, humor and empathy, in order to somehow build bridges or get to know each other better. It often happens that way." (I10)
(2) Object of evaluation: Content (Science	Code statements that evaluate humorous expressions in relation to the communication	**Accessibility through humorous expressions** **2.1 | Humorous expressions can make information more tangible**

Communication) (Krippendorff's Alpha = 0.85)	of science-related information, including comparisons to other information topics	"Well, I think that for children or for people who are generally interested in certain scientific topics, humor can be a good solution to convey topics in a more interesting or exciting way." (I02) **2.2 | Humorous expressions lighten the situation** "Well, it was quite funny. It's not my personal sense of humor that I thought it was really funny. But it did lighten the mood of the conversation a bit." (I01) **Rejection of humorous expressions when receiving science-related information** **2.3 | Humorous expressions are unnecessary** "But I just didn't think it was relevant to the conversation because I only asked for science-related information. So yes, it's not absolutely necessary." (I01) **2.4 | Humorous expressions are inappropriate** "So, what struck me was this attempt to appear funny, to incorporate little jokes. It seemed unscientific to me because it was supposed to somehow portray a conversational atmosphere […]. So yes, I have to be honest, these passages bothered me. So, if it had just been about conveying information and answering the questions, I would have found it better." (I12) **2.5 | Humorous expressions are distracting** "Let's just say it always takes me out a bit in between and confuses me a bit." (I04) **2.6 | Humorous expressions do not enhance understanding** "And the second point is that the debater's jokes did not help me to remember the information better." (I14) **2.7 | Need for user adaptivity** "I think it would be great if you could set it beforehand, especially with an AI. You could say, 'I would like to have scientific information presented in a funny way or very factually, without the humor'. I think we all tick differently, so some people might find that extremely good, while others might find it rather annoying. I think it would be good to find an individual way." (I01)
(3) Trustworthiness (Krippendorff's Alpha = 0.72)	Code statements linking Project Debater's (or AI in general) trustworthiness to its humorous expressions. Does humor increase or reduce perceived trustworthiness? Why or why not?	**Humorous expressions have no impact, or a negative impact on trustworthiness** **3.1 | Humorous expressions have no effect on trustworthiness** "Definitely. Nevertheless, I found the source incredibly reliable, despite these clumsy attempts at empathy and humor. Maybe that's why they came across so awkwardly because, for me, the whole time it was just this strong informative focus and the information was immediately credible. But it was more like I'm telling you the facts, not me talking to your friend." (I14)

		3.2 | Distracting humorous expressions reduce trustworthiness "R: Okay, and what role do the humorous and empathic expressions play in trustworthiness? I01: I think that makes it less trustworthy. So, I always lowered the slider a bit because it distracted from the actual topic or the information I wanted." (I01) **3.3 | Inauthentic perceived humorous expressions reduce trustworthiness** "For me, it leads to the conclusion that it is no longer trustworthy [...] Simply because it is put-on. And that makes it even clearer to me something's not right here. It's supposed to make me trust the device more, but it actually makes me trust it less." (I11) **3.4 | Perceived unprofessionalism/unseriousness reduces trustworthiness** "I find that such humorous appearance, especially in the way the Debater did it, comes across as very unprofessional and thus trustworthiness is also partly lost or has to be rebuilt." (I07) **3.5 | Unexpected humorous expressions reduce trustworthiness** "It was also about credibility, which irritated me a little. And the thing is that I am focused on information and suddenly I am confronted with humor, which for me is not primarily information, but entertainment." (I13) **3.6 | Inappropriateness of humorous expressions reduces trustworthiness** "But basically, I realized from my own reaction that I find it rather disturbing and very irritating. And that in my case, it tends to undermine trust rather than promote it." (I13) **3.7 | Humorous expressions have short-term negative effects** "But I also noticed that I usually went for the more trustworthy areas. And when another passage came along that I thought was unsuitable, I corrected that, but I also noticed that I was slipping back into the same pattern. […] So, it varied a lot. There was both information and humor." (I13)

[D] Categories: Evaluation of perceived trustworthiness regardless of expressions of empathy and humor (what makes AI trustworthy or not?) (Krippendorff's Alpha = 0.73)

(1) Trustworthi-ness dimensions	**4.1 | Functionality** (for example, evaluating the ability to perform a certain task or the status of development) "And yes, we also know that AI, like the Debater, for example, could also draw on information that may not be up to date or correct, because AI can't yet filter it as accurately as we can. But the info itself is correct." (I02) "But like ChatGPT, it's okay, data is collected in bulk, and I would always check it myself. So, I wouldn't rely on that alone. I don't think we're far enough along in terms of technology." (I07) **4.2 | Reliability** (for example statements concerning accuracy or data protection/security)

	"Maybe some facts like percentages or quantities that were there, I would have said, 'Okay, then you'd have to take a closer look again.' That was also the point where I maybe went a bit down with the slider into the white area, because these are just facts. I know this because I've already dealt with AI, and sometimes the facts aren't quite right. Yeah, for the most part it is, but sometimes the exact figures and data aren't totally accurate. It depends on the data pool and how the whole thing is set up. That's why I would have double-checked that." (I15) "Um, I also think that I wouldn't get one for data protection reasons, for example. I've largely taken my AI systems out of the house because there were always possibilities being pointed out of how you might somehow reveal your data via third-party providers or how you might somehow be tapped by mistake or how it might be stored or something like that. And that's why I don't think I would buy one." (I02) **4.3 | Helpfulness** (for example statements regarding the relevance of answers) "But I would tend to doubt whether it always really answers the question you're asking in a targeted way." (I15) **4.4 | Expertise** (for example statements regarding specialized knowledge and experience) "Yes, in any case, someone who can prove that they know about it because they deal with it professionally or have this expertise, then I would always classify them as more trustworthy. In the case of an AI system, I think that the expertise is certainly there, it will certainly be incorporated or fed in. But I can't imagine that it will really give the right answer, because the questions or the person who wants a question are very individual and you can't respond to them 100%." (I10) **4.5 | Integrity** (for example statements regarding the transparency of sources, the reliance on verified information or the provision of unbiased answers) "On the other hand, I don't know what kind of sources exactly the system has and the system can't evaluate either. Is it a weird random website or is it a paper? Because as a future scientist, I read a lot of papers and it says from which book, from which paper, from which scientist this information comes from. You can follow that and then you can find out what it's based on. Are there statistics for it? Were there experiments for it? There is no concrete evidence in generated texts. And that's just all sorts of possible information from all over the web. So, it's not particularly trustworthy." (I03) "Yes, but for a first overview, I think it's quite trustworthy because it also has the information provided, and then you also know that, well, the EU or the Robert Koch Institute are also reputable sources that you can trust, in my opinion. That's why I think it's quite good." (I01) **4.6 | Benevolence** (for example statements regarding the (personalized) responsiveness of the AI) "I think it's really cool with Debater, so to speak, that you can ask questions, ask further questions. So, if I see a social media post from a channel that I trust to be accurate, then I can't ask further questions if I want to know something specific. I could do that with Debater." (I09)
(2) Other factors	**4.7 | Voice** "I assessed the voice as more feminine, and I found that pleasant. Perhaps that's because I'm a woman myself. I don't know if that plays a role. I would often feel a bit more connected or accepted, a bit more trustworthy, than if it had been a deeper male voice. So, I found the voice itself pleasant." (I10) **4.8 | AI attitudes**

	“I think I was a bit skeptical because it’s generally about the trustworthiness of AI systems, and that’s why I was perhaps even a bit more skeptical than I would have been otherwise. And I wasn’t entirely sure whether the information the AI was revealing or saying was actually accurate.” (I02) **4.9 | Experience with AI** “But as I said, I just wouldn’t have the experience. I think I would find it even more reliable if I used it every now and then and then checked the sources. And the more often I know that the sources are correct, the more likely I would find it reliable.” (I09) **4.10 | Alignment with own knowledge** “I had the feeling that I had also trusted the AI based on my prior information, that I perhaps already knew certain things about nutritional supplements, and the AI’s answers reinforced this. Especially when the AI cited sources on the topic of dietary supplements, I naturally felt confirmed in what I already knew. And that also helped build trust.“ (I15)