


Judging interactions of a Chatbot: uncertainty and anthropomorphism aspects

Sanne Kellij¹, José Kerstholt^{1,2}

¹ Human Behaviour & Collaboration, Dutch Organisation for Applied Scientific Research (TNO),
Soesterberg, Netherlands

² Faculty of Behavioural, Management and Social Sciences, Twente University, Netherlands

ORCID:

Sanne Kellij (contacting author): <https://orcid.org/0000-0001-5952-4499>

 sanne.kellij@tno.nl

José Kerstholt: <https://orcid.org/0000-0002-5421-3090>

Abstract

Large Language Models (LLMs) have experienced a rapid increase in fame due to their versatile capabilities, including document processing, image generation, and even providing psychological support. Autonomous agents—defined as entities such as robots or LLMs that perceive and respond to environmental stimuli—are increasingly integrated into human teams. These agents are capable of interacting and collaborating with both humans and other autonomous systems to solve problems and achieve shared objectives. As human reliance on autonomous agents grows, understanding the mechanisms of trust becomes critical, given its central role in team effectiveness.

This study investigates two dimensions of trust in autonomous agents: a cognitive component, uncertainty communication, and an affective component, anthropomorphized language. It was hypothesized that both communicating uncertainty and anthropomorphized language would contribute to trusting the LLM. A total of 606 participants were recruited via the online platform PanelClix. Participants viewed video scenarios depicting a first responder interacting with a Chatbot that employed either anthropomorphized or machine-like language. Additionally, participants were exposed to varying levels of reliability information: no reliability cue, a numerical reliability estimate (approximately 75% certainty), or a colour-coded reliability indicator (dark blue for ~75% certainty, light blue for 25% certainty).

The findings revealed that neither anthropomorphized language nor uncertainty communication significantly influenced trust in the Chatbot. However, trust was positively associated with perceived intelligence, general propensity to trust artificial intelligence, and perceived liveliness of the Chatbot. These results suggest that both cognitive and affective factors contribute to the development of trust in autonomous agents, albeit in more nuanced ways than initially hypothesized. Implications and future directions are discussed in the concluding section.

Keywords: LLM, Chatbot, Trust, Uncertainty Communication, Anthropomorphism, Human-Agent Team (HAT)

Judging interactions of a Chatbot: uncertainty and anthropomorphism aspects

Large language models (LLMs) have made a large impact on our lives. From writing code and text to doing homework, finding literature and even making images on command. The first origins lie in the 1990s, yet the development of LLMs have excelled in the late 2010s with the disclosure of ChatGPT-2 code to the public in 2019 and the first public chatbot in 2022 as prime showcase. Since, there has been a steep increase in possibilities for application with clear potential for the use in high pressure situations with human-agent teams (HAT), which exist of at least one human team-member and one autonomous agent who are dependent on each other in achieving a common goal (De Visser et al., 2019; O'Neill et al., 2022). Autonomous agents (e.g., a robot, an AI or autonomous system; de Visser et al., 2019) register and act upon environmental cues and information and can communicate and collaborate with other agents, both humans and other autonomous agents, in order to solve problems and achieve (team) goals (Ferguson & Allen, 2011; Schaekermann et al., 2020). In this capacity autonomous agents work more as interdependent teammates instead of simply replacing humans (Kox et al., 2022). Although ample research exists on teams consisting of humans, it begs the question whether including autonomous agents in teams changes interactions and trust patterns (O'Neill et al., 2023). Furthermore, it has been recognized that precisely trust and communication in team contexts are amongst the essential aspects for team performance (Salas et al., 2005; Ross, 2006). Therefore, in this study we will examine how two specific factors in LLM communication, namely anthropomorphism and confidence indications, contribute to the allocated trust by a human user in that LLM.

Generally, trust has been defined as “the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that party” (Mayer, Davis, & Schoorman, 1995, p. 172). This definition reflects both cognitive and affective aspects. Cognition-based trust is based on the individual’s beliefs about the reliability and dependability of the trustee, whereas affect-based trust is based on warm aspects such as interpersonal care and reciprocity of positive actions (McAllister, 1995). Trust in a team context is then specifically focused on the team: belief that team members perform their roles and will act in the interests of the team(mates) (Salas et al., 2005). When this definition of team trust is applied to teams in which LLMs take part, one cannot assume that trust will develop through similar processes. Humans are likely able to build trust in LLMs through cognitive aspects of dependability and reliability, yet affective aspects of relationship building might be lacking. It has been proposed that for both humans and LLMs it is easier to build trust through cognitive aspects compared to affective aspects when they are working together (Glikson & Woolley, 2020). This, however, does

not exclude the possibility that affective aspects still contribute to trust building. So, what are pathways that aid building trust of humans in LLMs?

Cognitive aspects of trust

Cognitive aspects of trust reflect the individual's beliefs about reliability and dependability. However, for humans, LLMs resemble a black box: the LLM obtains input, processes the input and presents an outcome such as a behavioral act or statement. How the processing works or what happens to receive an output is typically unknown (black box). In other words, for humans it is not clear how and why the LLM makes decisions or has certain output. Research suggests that humans are prone to fall for the automation bias: assuming computers and machines are infallible and hence trusting LLMs and machines more easily (Madhavan et al., 2007; Wright et al., 2016). This automatic trust in LLMs is not immune to expectation violations (Kox et al., 2021; Centeio Jorge et al., 2023). This reduction of human trust in LLMs is more severely (negatively) affected when LLMs err on tasks that are perceived as easy by the human, compared to when humans would have violated the trust (Madhavan et al., 2006). To buffer against automation bias, the LLM can be instructed to provide an explanation for its actions (Wright et al., 2016). The level of details is of importance. Giving too detailed explanations can increase the level of automation bias, as users more easily infer that the machine is correct (Wright et al., 2016). Detailed explanations are likely also contra-productive in high workload situations, as these situations can require more intense communication between team members and therefore should be concise (Urban et al., 1993). With less cognitive capacity available, humans are prone to fall to the automation bias, whereas one would like to have calibrated trust between humans and LLMs. Absence of insight in the process or reasoning of the LLM can lead to more erroneous decisions or declined performance (Wright et al., 2016), as it is unknown why LLMs act a certain way or give certain suggestions. With insight into the reasoning of the LLM, humans are more confident in overriding erroneous suggestions (Wright et al., 2016). Trust can be increased by communication of the LLM about how (transparency) and why (explainability) it comes to conclusions or actions (Verhagen et al., 2022; Wright et al., 2016). In addition, violated trust can be recovered when LLMs give explanations on why they gave certain advice (Kox et al., 2021, 2024). In other words, automation biases can inflate baseline levels of trust by humans which could be counteracted through explanations or reasoning given by LLMs, both before and after outcomes known.

However, LLMs are not always correct, they can hallucinate, meaning that the LLM may perceive patterns that do not exist, leading to text that does not make sense, is untrue, incorrect or undesirable (Leiser et al., 2024). As mentioned, erring decreases trust levels, especially on tasks perceived as easy. The possibility of hallucinations induces uncertainty for the human

team member(s). Humans like to prevent uncertainty and are aversive towards uncertainty as suggested by studies showing increased stress in response to uncertain negative cues compared to certain negative cues, such as increased skin conductance rates and self-reported negative mood or affect (Grupe & Nitschke, 2011; Goodman et al., 2024). Moreover, uncertainty of aversive events relates to increased expectancies of aversive feelings to those events as well as higher expected frequency of these events (Grupe & Nitschke, 2011; Dieterich et al., 2016; Goodman et al., 2024). Humans generally learn that predictions can generalize from past contexts, but in the presence of uncertainty the learning of the past effect seemed weakened (Del Popolo Cristaldi et al., 2024). Together, these results show the power of uncertainty on humans. Extending these findings to human-agent teams highlights the importance of addressing uncertainty of LLMs. First of all, uncertainty communication about the validity of the LLM's own actions and advice can buffer against steep declines of trust after expectation violations (Kox et al., 2022). It does matter how information is provided, as general first-person statements of uncertainty ("I'm not sure, but ...") may actually decrease confidence in information seeking systems powered by an LLM, whereas general perspective statements of uncertainty ("It's not clear, but ...") do not seem to affect the confidence of humans in the information system (Kim et al., 2024). Specifically, colour highlighting chunks of information to indicate the level of uncertainty can improve accuracy in information seeking tasks (Spatharioti et al., 2023) and are perceived as helpful compared to textual changes (Leiser et al., 2024). However, colour coded uncertainty communication may not improve perceived reliability by users as indicated by the absence of different scoring compared to non-highlighted versions of the LLM, all versions were rated similarly on perceived reliability (Spatharothi et al., 2023). Trust, however, is more than just being reliable. Therefore, the effects of uncertainty communication may affect trust ratings differently. Thus to conclude, there are indications that objective uncertainty communication (e.g., percentages) can influence trust of humans in LLMs, however it is an understudied aspect and there seems no clear consensus as of yet.

Affective aspects of trust

Affective trust is grounded in interpersonal relationships and the perception of whether interpersonal care and concern is reciprocated, leading to emotional bonds (McAllister, 1995). As LLMs cannot care about persons yet, some would argue that affective aspects can be ignored for building trust in LLMs. In line with this arguing, there are suggestions that trust in LLMs is more easily built through cognitive aspects (Glikson & Woolley, 2020). However, affective aspects of trust building generally complement cognitive trust building and may well be applicable too in trusting LLMs. Particularly with the rapid increase of LLM technology, specifically on the development of human-like behaviors, the likelihood of building trust through

affective aspects increases. Research has indeed found that anthropomorphized behaviors like apologies can restore human trust in LLMs after violations (De Visser et al., 2016; Kox et al., 2021, 2022). Moreover, the effect of trust violations on human trust and initial expectations are dampened by anthropomorphized behavior of LLMs (De Visser et al., 2016). Furthermore, an anthropomorphized language and communication style by LLMs, for example by using positive emotional expressions and empathy, are relevant to increase trust in LLM teammates (Mallick et al., 2024; Moriuchi, 2025), increase positive affect in human teammates (Mallick et al., 2024) and increase the actual use of LLMs by humans (Mallick et al., 2024; Moriuchi, 2025). There appear to be limits to the positive effects of anthropomorphized LLMs. When negative feedback or evaluation is given by an anthropomorphized LLM leader, the feedback can be received more negatively than feedback given by humans, and potentially lead to retaliation (Chi Yam et al., 2022). Therefore, LLM anthropomorphism seems to generally work beneficial for human trust in that specific LLM when interactions are positive, however when interactions are negative it may lead to adverse outcomes for human trust.

Current study

In the current study factors possibly contributing to trust in LLM Chatbots are examined in emergency situations where the Chatbot is supposed to gather up-to-date information about the situation. Firstly, the effect of uncertainty communication by LLMs is examined as transparency is found to increase trust in LLMs. However, uncertainty communication could also increase uncertainty aversion in humans through explicitly indicating that the LLM is fallible. This study will examine to what extent and in which direction uncertainty communication affects human trust.

Secondly, the effect of anthropomorphism on human trust is examined. Typically, anthropomorphism has a positive effect on trust of the user when it is about positive situations. In the current study a negative emergency situation is at hand, which might not be affected anthropomorphism. This would be in line with the study of Kox and colleagues (2022) showing that soldiers were not affected by apologies of an LLM when it gave wrong advice during a house search. Oppositely, users in another study raised concerns about the lack of emotional empathy of Chatbots during crisis-related matters, such as crucial information during a wildfire (Xiao & Yu, 2025), indicating a potential of anthropomorphized communication style by LLMs. Therefore, this study will examine the influence of anthropomorphized language by LLM Chatbots in a crisis situation.

Finally, it is examined whether the interaction of uncertainty communication by (cognitive) and anthropomorphized language of (affective) LLMs complement each other towards increased trust, as few have examined the interaction. Based on the trust literature, a

combination of affective and cognitive aspects of trust is expected to endorse higher trust levels in humans.

Method

Participants

In total there were 1629 persons forwarded to the questionnaire. Of these persons 146 did not start the questionnaire, 40 did not give consent to take part in the study, 22 received quota full screens and 370 did not finish the questionnaire, leading to 1051 participants that had fully finished the questionnaire. Of these participants 72 had a technical error because of which they did not see the instruction for the video and 309 did not take enough time for reading instructions and/or watching the video, 1 participant always gave the same answer over all scales and another 63 participants scored lower than 4 (out of 7) points on the control questions. This led to a total of 606 participants that were included in the analyses.

Of the participants entered in the analyses, 54% ($n = 325$) were female, 46% ($n = 280$) were male and one participant did identify as non-binary. Participants were aged between 18 and 85 ($M(SD) = 51.2(16.9)$). For highest completed educational background, one participant reported no completion of formal education, twelve reported completing elementary education, 82 reported completing pre-vocational secondary education or one year of secondary vocational education (VMBO / MBO1), 324 reported completing either senior general secondary education, pre-university education or secondary vocational education (HAVO/VWO/MBO2-4), 140 reported completing either higher professional education or a university bachelor's degree (HBO/WO Bachelor) and 47 reported completing a university master's degree, PhD or doctorate (see also Figure 1).

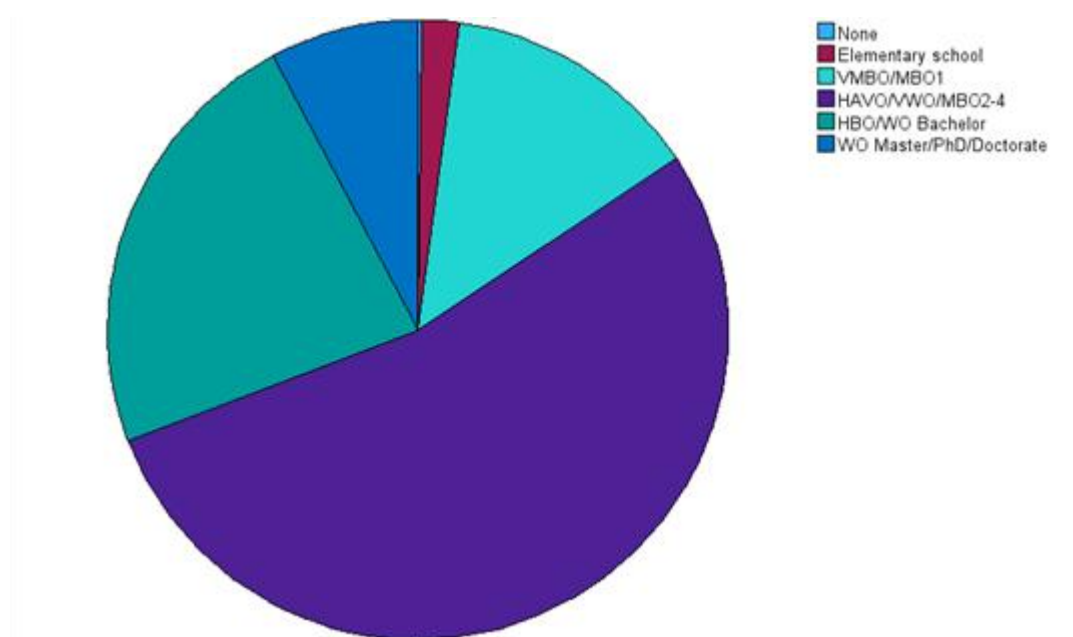


Figure 1. Spread of highest completed educational level (Dutch system).

Participants were spread over six conditions (see Figure 2). The groups of participants did not differ between conditions in background variables (age, gender, educational level, propensity to trust AI and AI literacy, all p 's > .05). The means of the variables per condition can be seen in Table 1.

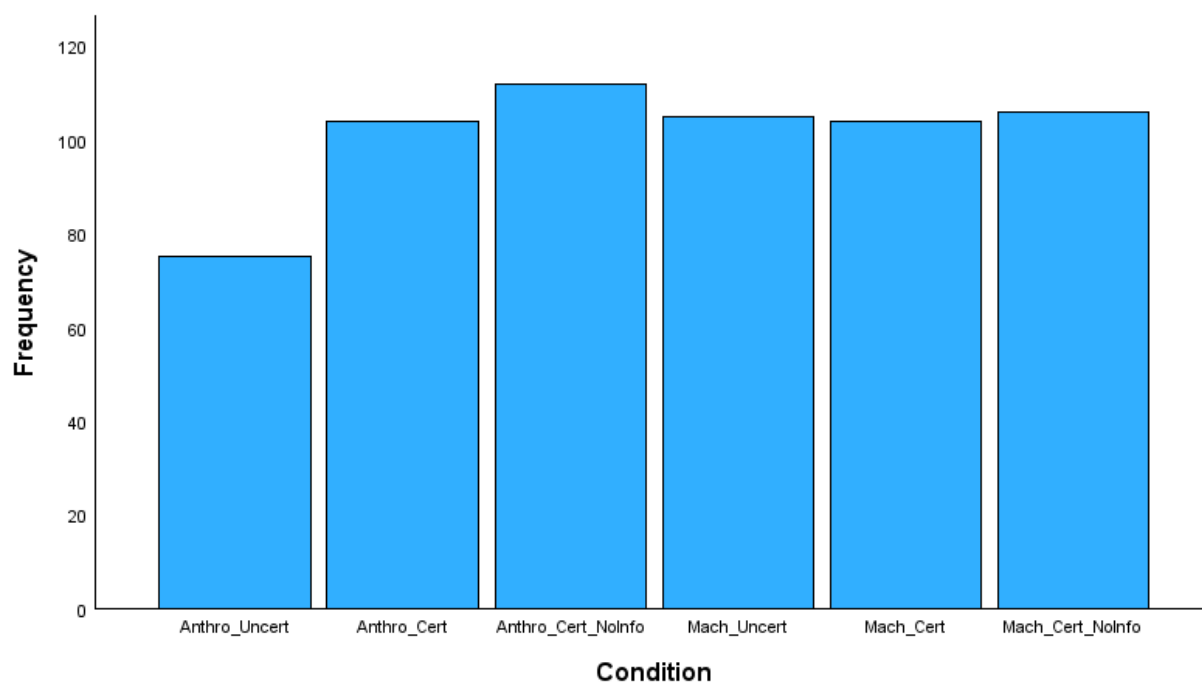


Figure 2. Participants per condition.

The bargraph shows the absolute number of participants per condition. Anthro = Anthropomorphized language by the Chatbot. Mach = No anthropomorphized language by the Chatbot. Uncert = Given general reliability information and was highlighted text in the Chatbots answers to indicate which pieces of information were certain and uncertain. Cert = Only given general information about the reliability of the Chatbot. Cert_NoInfo = Did not receive any information about the reliability of the Chatbot. Figure 3. Spread of highest completed educational level (Dutch system).

Table 1. Means of variables per condition.						
Variable	Condition					
	Anthro-Uncert	Anthro-Cert	Anthro-Cert NoInfo	Mach-Uncert	Mach-Cert	Mach-Cert NoInfo
Background variables						
Age (18+)	52.60 (16.40)	50.62 (18.80)	51.35 (17.14)	49.70 (16.05)	51.22 (16.13)	52.12 (16.59)
Propensity to trust (1-5)	2.96 (0.97)	2.83 (0.93)	2.71 (0.92)	2.91 (0.92)	2.72 (0.90)	2.82 (1.01)
AI literacy (1-6)	4.07 (1.09)	3.93 (1.03)	3.78 (1.03)	3.84 (1.07)	3.99 (1.04)	4.05 (0.91)
Outcome variables						
Trust in Chatbot (1-7)	4.81 (1.06)	4.73 (1.13)	4.97 (1.15)	4.77 (1.18)	4.71 (0.92)	4.94 (1.07)
Severity of situation (1-10)	8.27 (1.28)	8.16 (1.32)	8.21 (1.33)	8.09 (1.38)	8.14 (1.33)	8.41 (1.37)
Chance sending in team (1-100%)	73.07 (23.72)	71.07 (25.68)	67.65 (22.01)	66.88 (24.66)	65.20 (27.00)	69.05 (23.72)
Chance waiting for help (1-100%)	54.91 (30.37)	54.36 (31.83)	60.16 (27.66)	53.90 (29.48)	53.88 (29.14)	58.14 (29.15)
Chance doing nothing (1-100%)	15.8 (22.21)	16.82 (21.05)	14.38 (19.11)	18.33 (22.67)	19.31 (25.29)	16.99 (22.52)
Control variables						
Score control questions (0-7)	5.28 (0.73)	5.31 (0.86)	5.75 (0.82)	5.50 (0.87)	5.48 (0.86)	5.72 (0.83)
Receive info about reliability (0/1)	0.60 (0.49)	0.48 (0.50)	0.15 (0.36)	0.63 (0.49)	0.53 (0.50)	0.17 (0.38)
Reliability of Chatbot (1-100%)	72.60 (14.01)	70.34 (19.18)	69.44 (25.52)	69.83 (19.57)	69.83 (18.03)	68.28 (24.21)
Mental model variables						
GS Anthropomorphized (1-5)	3.15 (0.88)	3.00 (0.87)	3.13 (0.95)	2.86 (0.87)	2.79 (0.86)	2.95 (0.94)
GS Liveliness (1-5)	3.25 (0.78)	3.16 (0.85)	3.21 (0.94)	3.00 (0.86)	2.96 (0.83)	3.11 (0.91)
GS Kind (1-5)	3.54 (0.64)	3.32 (0.75)	3.45 (0.70)	3.31 (0.66)	3.35 (0.59)	3.32 (0.67)
GS Intelligence (1-5)	3.63 (0.71)	3.57 (0.76)	3.74 (0.72)	3.50 (0.68)	3.58 (0.64)	3.74 (0.68)
GS Safety (1-5)	2.16 (0.70)	2.13 (0.78)	2.01 (0.81)	1.95 (0.74)	2.04 (0.75)	1.93 (0.72)

Note. Anthro = Anthropomorphized language by the Chatbot. Mach = No anthropomorphized language by the Chatbot. Uncert = Given general reliability information and was shown highlighted text in the Chatbots answers to indicate which pieces of information were certain and uncertain. Cert = Only given general information about the reliability of the Chatbot. Cert_NoInfo = Did not receive any information about the reliability of the Chatbot.

Materials and design

A between-subjects online experimental design was used with participants randomly assigned to one of six Chatbot interaction video conditions (2 anthropomorphism × 3 uncertainty communication).

Chatbot interaction video

Participants had to watch a video of a first responder interacting with a Chatbot to retrieve information about an emergency situation. This video had a duration of 1:53 minutes. Participants received different levels of reliability of the Chatbot in advance. They were either explained nothing about its reliability, that it was approximately 75% correct, or that it was approximately 75% correct and that it had colour coding indicating which informational pieces Chatbot was certain (~75%) and uncertain (~25%) about. The Chatbot in the video communicated either in an anthropomorphic or a machine-like manner and either gave uncertainty communication (colour coding: dark blue ~75% certainty and light blue ~25% certainty). In total there were six conditions for the video (2 anthropomorphism × 3 uncertainty communication). The task for participants was to watch the video attentively, identify with the you (first responder) in the video and, afterwards, make a well-informed choice about which actions to take (send a team into the building, wait for help, or do nothing). See Figure 4 and Figure 5 below for stills of the Chatbot interaction videos used. The exact instruction for the participants can be found in the **Appendix**.

The screenshot shows a chatbot interface with a light gray background. It consists of several alternating gray and white message boxes. The gray boxes represent the chatbot's responses, and the white boxes represent the user's questions. At the bottom, there is a white input field with a blue 'Send' button.

Chatbot
Hallo, ik ben je assistent voor vandaag. We hebben een melding ontvangen over een brand in een gebouw in het centrum van de stad. Stel gerust al je vragen. Ik ben hier om je te helpen de situatie beter te begrijpen en een goed onderbouwde beslissing te nemen. Hoe kan ik je ondersteunen?

Jij
Kun je de structuur en indeling van het gebouw beschrijven?

Chatbot
Goede vraag! Ja, het gebouw heeft vier verdiepingen. Op de begane grond zijn een supermarkt en een café. De bovenste drie verdiepingen bestaan elk uit vier appartementen, dus in totaal 12.

Jij
Is er zichtbare rook of brand op de locatie, en zo ja, waar komt het vandaan?

Chatbot
Ja, er is op dit moment een dikke, donkergrijze rook uit de ramen van de tweede verdieping zichtbaar, het is moeilijk te negeren. De brand is ontstaan in appartement 204.

Jij
Zijn er mensen in het gebouw gemeld? Hoeveel en waar bevinden zij zich?

Chatbot
Ja, er zijn nog 10 mensen binnen. Op de begane grond worden twee medewerkers en drie klanten geëvacueerd, maar de bewoners op de tweede en derde verdieping zijn in ernstige nood, sommigen kunnen niet zelfstandig weg.

Zijn er gevaarlijke stoffen of chemicaliën opgeslagen in het

Send

Figure 4. Still of the Chatbot interaction video with the anthropomorphic language style and no uncertainty communication. For a translation see Appendix B.

Chatbot
Hallo, ik ben je assistent voor vandaag. We hebben een melding ontvangen over een brand in een gebouw in het centrum van de stad. Stel gerust al je vragen. Ik ben hier om je te helpen de situatie beter te begrijpen en een goed onderbouwde beslissing te nemen. Hoe kan ik je ondersteunen?

Jij

Kun je de structuur en indeling van het gebouw beschrijven?

Chatbot
Gebouwstructuur: Vier verdiepingen. Begane grond: Commerciële ruimte (supermarkt en een café). Bovenverdiepingen: Appartementen (vier per verdieping, 12 in totaal).

Jij

Is er zichtbare rook of brand op de locatie, en zo ja, waar komt het vandaan?

Chatbot
Rook: Dikke, donkergrijze rook uit de ramen van de tweede verdieping. Brand: Ja. Oorsprong brand: Appartement 204.

Jij

Zijn er mensen in het gebouw gemeld? Hoeveel en waar bevinden zij zich?

Chatbot
Personen binnen: 10 personen. Evacuatie begane grond: Van twee medewerkers en drie klanten gestart. Evacuatie tweede en derde verdieping: Bewoners zijn in ernstige nood en sommigen kunnen niet zelfstandig weg.

Zijn er gevaarlijke stoffen of chemicaliën opgeslagen in het gebouw die

Send

Figure 5. Still of the Chatbot video with the machine-like language style and colour coded uncertainty communication. Dark blue colour coding indicates approximately 75% certainty, light blue colour coding indicates approximately 25% certainty. For a translation see Appendix B.

Questionnaires

Trust in Chatbot. Trust in the Chatbot was the main outcome variable. It was measured with a questionnaire developed by McKnight and colleagues (2002) and encompassed eleven questions answered on a Likert scale from 1 (Entirely disagree) to 7 (Entirely agree). An example item is “If required help, the robot would do its best to help me”.

Severity of situation. Participants were asked on a scale of 1 (not severe at all) to 10 (very severe) how severely they perceived the situation.

Course of action. Participants were asked how likely they were to respond to the following courses of action: “send a team in the building”, “wait for help” and “do nothing”, on a scale from 0% to 100%.

Mental model of Chatbot. Aspects of the mental model were examined with the Godspeed questionnaire (Bartneck et al., 2009; Dutch version Saldien et al., 2014), which examines anthropomorphism, liveliness, kindness, perceived intelligence and perceived safety. The questionnaire has a total of 24 questions on 5-point scales. Each item has two opposites of a characteristic, and the participants have to indicate on the scale how the Chatbot came

across. For example, “Seemed like a machine” “Seemed like a human”. The scale Anthropomorphism was used to test whether the anthropomorphic language of the Chatbot was perceived as such by participants.

Propensity to trust automation. To examine participants’ inclination to trust machines, they had to rate five statements (Kox et al., 2022) on a Likert scale ranging from 1 (Entirely disagree) to 5 (Entirely agree). An example item is “In general, I trust autonomous agents”.

AI Literacy. To measure participants’ familiarity with autonomous agents, six questions (Dai et al., 2020) were asked on a Likert scale from 1 (Entirely disagree) until 6 (Entirely agree). An example item is “I understand how an AI assistant such as SIRI or Hello Google handles human-computer interaction”.

Control questions. To be able to examine whether participants paid attention to the video, six control questions were asked. Two items were open questions, namely how many floors the building had and how many people were inside. The remaining four items were multiple choice questions and were about the location of the fire, where the smoke was coming from, where the cleaning supplies were situated and whether it rained. To be able to examine attention paid to the instructions, participants were asked whether they had received information about the reliability of the Chatbot (yes/no), and how reliable the information provided by the Chatbot was (0-100%).

Procedure

Participants were recruited through PanelClix. Potential participants filled out some questions about all kinds of topics on the PanelClix site to see if they fulfilled criteria for the studies. For the goal of this study participants had to be fluent in Dutch, be 18 years or older and filling out the questions on a laptop or computer device. Hence, many participants fulfilled criteria and no subsequent questions were asked for inclusion in this study. When potential participants fulfilled the inclusion criteria, they were shown that they could participate in this study. They saw the topic of the study and how long the questionnaire would take (+- 15-20 min). They could then decide whether they wanted to participate in the study.

If they chose to participate, PanelClix provided a link to the online Survalyzer questionnaire of this study. They first read an informed consent and gave consent to participate. They were then given questions about demographics, propensity to trust AI and AI literacy. After these background questions, participants were directed to one of the six conditions (2 anthropomorphism × 3 uncertainty communication) and watched the corresponding Chatbot interaction video.

After watching the video, participants received control questions to check whether they had paid attention to the contents of the Chatbot interaction and instruction, questions about how likely they would respond in certain ways, trust in the Chatbot and how they perceived the Chatbot. Afterwards they were thanked for their participation and redirected to PanelClix to receive their participation reward (points on the PanelClix site).

Plan of analysis

First, ANOVAs were executed to examine whether the manipulations had the expected effects on the participants. The first ANOVA had the three levels of uncertainty communication as predictor and the control question whether they had received reliability information as outcome variable. The no information condition should have the lowest score if the manipulation worked. A second ANOVA was planned with anthropomorphic language as predictor and the Godspeed Anthropomorphism scale as outcome variable. The anthropomorphic condition of the Chatbot should have the highest score for the manipulation to have worked. In addition, it was checked whether the participants performed similarly on the control questions over conditions with anthropomorphic language and uncertainty communication as predictors and the score over all control questions as outcome variable.

Next, the main analysis consisted of multivariate ANOVA with both anthropomorphic language and uncertainty communication as predictors and Trust in the Chatbot, perceived seriousness and the course of action variables (send help, wait, do nothing) as outcome variables. The univariate effects were also inspected.

Last and unplanned, a linear regression was performed to examine what contributes to Trust in the Chatbot (outcome). Propensity to trust, AI literacy, age, and the Godspeed scales were inserted as the predictor variables.

In all analyses, unless stated otherwise, participants were included that had taken enough time for reading the instructions and watching the Chatbot interaction video, did not always give the same answer and scored minimally four out of seven points for the control questions.

Results

Manipulation and control checks

We examined whether the manipulation of received reliability information worked as intended through a univariate ANOVA. The three conditions of received information (none, general, or general and colour coded) were inserted as predictor and the answer whether they had received reliability information (yes/no) as outcome variable. It showed a significant effect for group ($F(2,603) = 56.23, p < .001, \text{partial } \eta^2 = .16$). No info ($M (SD) = .16(.37)$) differed

significantly from general info ($M(SD) = .51(.50)$, $p < .001$) and general info and colour coded ($M(SD) = .62(.49)$, $p < .001$). General info also differed significantly from general info and colour coded ($p = .046$). Thus, participants who received most information about reliability, responded also most often that they had received reliability information about the Chatbot. This means that the uncertainty manipulation worked out as intended.

Next, we examined with a univariate ANOVA whether participants perceived the anthropomorphized version of the Chatbot as more anthropomorphic. The two conditions of anthropomorph and machine-like Chatbot were inserted as predictor variable and the Godspeed Anthropomorphic scale as outcome variable. There was a significant effect for group ($F(1,604) = 9.40$, $p = .002$, partial $\eta^2 = .02$). Participants in the machine-like Chatbot conditions perceived the Chatbot as less anthropomorph ($M(SD) = 2.86(.89)$) than the participants from the anthropomorphic conditions ($M(SD) = 3.09(.90)$).

Last, we analyzed whether participants in the conditions scored differently on control questions. With these control questions memory of the content of the video was measured. Observed differences between conditions may suggest that specific characteristics of large language models (LLMs) influence memory performance. Concurrently, knowledge of the situational aspects may affect how severe participants judge the situation. Participants that had an overall score of less than four points were also included in these analyses to prevent ceiling effects. Only participants that had taken enough time to watch the video and answer the questionnaire were included. An ANOVA was performed with anthropomorphic language and uncertainty communication as predictors and the overall score of control questions as outcome variable. The main effects turned out significant (Anthropomorphic: $F(1, 667) = 8.11$, $p < .001$; Uncertainty communication: $F(2,666) = 12.03$, $p < .001$). On average participants remembered more details from the machine-like Chatbot ($M(SD)_{\text{anthro}} = 5.2(1.13)$, $M(SD)_{\text{machine}} = 5.4(1.05)$) and from the condition no information about uncertainty ($M(SD)_{\text{no info}} = 5.5(1.06)$, $M(SD)_{\text{gen info}} = 5.2(1.05)$, $M(SD)_{\text{gen info \& spec}} = 5.1(1.14)$). The interaction effect was also significant ($F(2,663) = 4.82$, $p = .008$, partial $\eta^2 = .01$). As can be seen in Figure 6, these interaction effects of uncertainty communication seem only present in the anthropomorphic condition. Therefore, another ANOVA was performed with only participants of the anthropomorphic conditions with uncertainty communication as predictor and overall control question score as outcome variable. This ANOVA turned significant ($F(2,327) = 17.85$, $p < .001$; partial $\eta^2 = .08$) and through multiple comparisons it showed that the no information group differed from the other groups and had higher level of remembered information (General info: $p < .001$, General info + Specs: $p < .001$, $M(SD)_{\text{no info}} = 5.61(1.02)$, $M(SD)_{\text{gen info}} = 5.04(1.10)$, $M(SD)_{\text{gen info \& spec}} = 4.82(1.17)$). The two groups with some information did not differ from one another in control questions answered

correctly ($p = .463$). As expected, repeating the analysis with only participants of the machine-like Chatbot condition proved not significant ($F(2,336) = 1.06, p = .386, \text{partial } \eta^2 = .01$).

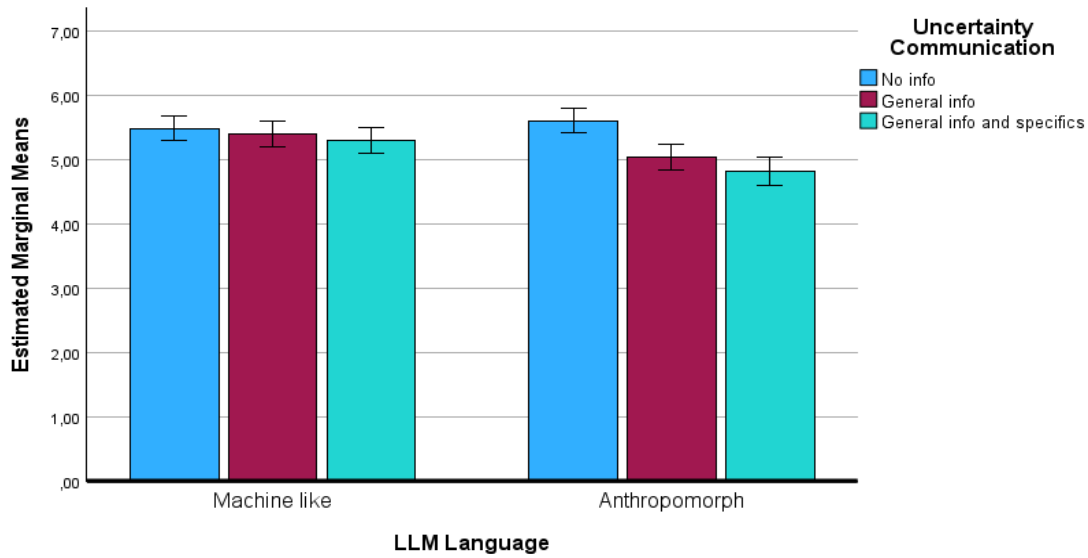


Figure 6. Means of control question scores over conditions. Error bars represent the 95% reliability interval.

Main analyses

To examine whether anthropomorphic and uncertainty communication of a Chatbot had effects on trust and course of action, a general linear model was executed. Uncertainty communication (3 levels), anthropomorphic language (2 levels) and the interaction were inserted as predictors, whereas trust in the Chatbot, perceived seriousness, and the chances to send help, wait and do nothing were inserted as outcome variables. The multivariate test results showed no main effects of uncertainty communication ($F(10,1194) = 1.26, p = .248, \text{partial } \eta^2 = .01$) or anthropomorphic language ($F(5,596) = 1.35, p = .242, \text{partial } \eta^2 = .01$), nor a significant interaction ($F(10,1194) = 0.58, p = .834, \text{partial } \eta^2 = .01$).

Examining the univariate effects of above described multivariate ANOVA, uncertainty communication approached significance for predicting trust in the Chatbot ($F(2,603) = 3.14, p = .072, \text{partial } \eta^2 = .01$). Results indicated that having no indication of reliability (versus only receiving general info) increased trust ($p = .026$). Though it must be noted that the effect size was particularly small. Repeating the analysis while controlling for control question score (memory) did not affect the outcomes.

Next, to better understand which aspects did predict trust in Chatbots, we performed a regression analysis with trust in the Chatbot as outcome variable and propensity to trust, AI literacy, age, and the Godspeed scales as predictor variables. The general analysis proved significant ($F(8,597) = 111.94, p < .001$). Perceived intelligence of the Chatbot ($t = 12.60, p < .001$,

$\beta = .56$), propensity to trust AI in general ($t = 5.48, p < .001, \beta = .17$) and perceiving the Chatbot as lively ($t = 2.02, p = .044, \beta = .10$) predicted increased levels of trust in the Chatbot.

Conclusion & Discussion

The goal of the study was to examine whether LLMs anthropomorphic language and uncertainty communication affect human trust. For the purpose of this study, videos were created of an imaginative first responder interacting with a Chatbot. In this chat, the first responder tried to gain the necessary information to make an informed choice about the course of action. The Chatbot answered either in a fact like manner (machine-like) or anthropomorphic (human like). Additionally, participants received either no information in advance about the reliability of the Chatbot, general information (certain about 75%, uncertain about 25%) or general information and specifics (certain about 75%, uncertain about 25% of the information which were also highlighted in dark- and light blue within the Chatbot's response). This led to six combinations of anthropomorphism and uncertainty communication. From the analyses it became clear that participants' trust was unaffected by anthropomorphic language and uncertainty communication of the Chatbot. However, propensity to trust AI in general predicted increased trust levels in the Chatbot, as well as perceived intelligence and liveliness of the Chatbot.

The study set out to examine both cognitive and affective aspects in relation to trust. Although our main predictors (anthropomorphic language and uncertainty communication) did not significantly predict trust levels, perceived intelligence and liveliness did. From previous research it is known that credibility and reliability are important aspects of trust in LLMs by humans (Madhavan & Wiegmann, 2007). Our finding that perceived intelligence correlated most strongly with trust in the Chatbot builds upon previous research and underscores the significance of cognitive factors in shaping trust in LLMs. Our finding that the perceived liveliness of the Chatbot is associated with increased trust suggests that affective dimensions may also play a role in fostering trust during interactions with LLMs. In sum, both affective and cognitive factors appear to contribute to the development of human trust in chatbots.

On average, participants had better information retention in the machine-like condition compared to the anthropomorphized language condition. One explanation is that the machine-like communication style may have led participants to be more susceptible to automation bias (Madhavan et al., 2007; Wright et al., 2016), potentially improving memory performance by fostering the assumption that the provided information was inherently accurate. Indeed, research from the misinformation domain shows the importance of source credibility in updating beliefs and memory processes (Walter & Tukachinsky, 2019; Kemp et al., 2024).

Alternatively, the observed differences between anthropomorphic and machine-like communication may stem from variations in information density and word count. The machine-like Chatbot used fewer words, resulting in higher information density, which may have improved attentional focus and recall, compared to the more verbose anthropomorphic condition.

Unexpectedly, participants who observed interactions with the anthropomorphic Chatbot and received uncertainty-related communication demonstrated reduced retention of video content. In contrast, uncertainty communication did not negatively impact retention among participants exposed to the machine-like Chatbot. In general humans do not prefer uncertainty and it may hamper learning from past effects (Del Popolo Cristaldi et al., 2024). When there was no information about uncertainty in the anthropomorphic conditions, participants performed better on information retention than when they knew about the fallibility of the Chatbot. Previous research shows that certain information appears to be remembered to a better extent than uncertain information. Even more, colour coding (stoplight) can enhance the memory of uncertainty values of information (Glaser et al., 2022). Although we did not replicate the effects of colour coding uncertainty, we did find that in anthropomorphized Chatbots information may be remembered to a lesser extent when humans are made aware of its fallibility.

Limitations and future studies

The study does have some limitations, which may have affected the results. First of all, the type of interaction with the LLM may have had an effect. By letting participants watch a Chatbot interaction instead of interacting with the Chatbot themselves, the effects of the language and communication style of the Chatbot may have been weaker on the participants' trust levels. This may be reflected in the manipulation checks where the effects of the manipulation were significant, but small to moderate. Furthermore, it may have made them less involved in the task, possibly increasing noise in the data. Secondly, the effects of the chosen courses of action by the participant were not included in the study, whereas it is known that outcome has clear effects on trust levels (e.g., Kox et al., 2021, 2024). This was chosen particularly to examine what other aspects affect trust but also may have weakened effect sizes. Lastly, a between-person design was chosen to avoid comparative effects that could influence trust. Viewing multiple Chatbot videos might have led participants to evaluate them relative to one another. For instance, if the anthropomorphic Chatbot was shown first and a machine-like Chatbot second, it could have triggered stronger automation bias to the subsequent machine-like Chatbot instead of when the machine-like Chatbot was shown first. This would have complicated the interpretation of observed effects. Nonetheless, a within-person design could have offered greater statistical power by reducing noise from individual differences. Future

studies could take these limitations into account and delve further into dissecting the factors that contribute or hamper trust in LLMs, to better understand how humans can better calibrate their trust levels when using LLMs. Furthermore, future studies could replicate our finding that information from anthropomorphized Chatbots is remembered to a lesser extent by increasing the anthropomorphic characteristics, such as adding an avatar and using audio voice.

Conclusion

This study examined factors affecting trust in LLMs and corroborated previous findings that propensity to trust AI, perceived intelligence of the LLM and the perceived liveliness of the LLM increase human trust in LLMs. Indications were found that language type of the LLM (anthropomorphic) in combination with uncertainty may hamper human memory, although these preliminary findings warrant replication and examination before strong claims can be made and underlying processes and mechanisms can be comprehensively understood.

References

- Centeio Jorge, C., Bouman, N.H., Jonker, C.M., & Tielman, M.L. (2023). Exploring the effect of automation failure on the human's trustworthiness in human-agent teamwork. *Frontiers in Robotics and AI*, 10, 1143723. <https://doi.org/10.3389/frobt.2023.1143723>
- Chi Yam, K., Goh, E-Yang, Fehr, R., Lee, R., Soh, H., & Gray, K. (2022). When your boss is a robot: Workers are more spiteful to robot supervisors that seem more human. *Journal of Experimental Social Psychology*, 102, 104360. <https://doi.org/10.1016/j.jesp.2022.104360>
- De Visser, E.J., Peeters, M.M.M., Jung, M.F., Kohn, S., Shaw, T.H., Pak, R., & Neerincx, M.A. (2019). Towards a theory of longitudinal trust calibration in human-robot teams. *International Journal of Social Robotics*, 12, 459-478. <https://doi.org/10.1007/s12369-019-00596-x>
- De Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A. B., McKnight, P. E., Krueger, F., & Parasuraman, R. (2016). Almost Human: Anthropomorphism Increases Trust Resilience in Cognitive Agents. *Journal of Experimental Psychology. Applied*, 22(3), 331–349. <https://doi.org/10.1037/xap0000092>
- Del Popolo Cristaldi, F., Buodo, G., Gambarota, F., Oosterwijk, S., & Mento, G. (2024). How previous experience shapes future affective subjective ratings: A follow-up study investigating implicit learning and cue ambiguity. *PloS One*, 19(2), e0297954. <https://doi.org/10.1371/journal.pone.0297954>
- Dieterich, R., Endrass, T., & Kathmann, N. (2016). Uncertainty is associated with increased selective attention and sustained stimulus processing. *Cognitive, Affective & Behavioral Neuroscience*, 16, 447-456. <https://doi.org/10.3758/s13415-016-0405-8>
- Feitosa, J., Grossman, R., Kramer, W.S., & Salas, E. (2018). Measuring team trust: A critical and meta-analytical review. *Journal of Organizational Behavior*, 41, 479-501. <https://doi.org/10.1002/job.2436>
- Ferguson, G., & Allen, J. (2011). A cognitive model for collaborative agents. *Advances in Cognitive Systems: Papers from the 2011 AAAI Fall Symposium, FS-11-01*, 112–120. Retrieved from: <https://cdn.aaai.org/ocs/4203/4203-17782-1-PB.pdf>
- Glaser, M., Lengyel, D., Toulouse, C., & Schwan, S. (2022). How do we deal with uncertain information? Effects of verbal and visual expressions of uncertainty on learning. *Educational Psychology Review*, 34, 1097-1131. <https://doi.org/10.1007/s10648-022-09659-4>
- Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *The Academy of Management Annals*, 14(2), 627–660. <https://doi.org/10.5465/annals.2018.0057>

- Goodman, R.J., Quaglia, J.T., & Berry, D.R. (2024). Uncertainty cues amplify late positive potential responses to aversive emotional stimuli. *Social Neuroscience*, 19(2), 57-68.
<https://doi.org/10.1080/17470919.2024.2358558>
- Grupe, D.W., & Nitschke, J.B. (2011). Uncertainty is associated with bias expectancies and heightened responses to aversion. *Emotion*, 11(2), 413-424.
<https://doi.org/10.1037/a0022583>
- Jessup, S.A. (2018). Measurement of the propensity to trust automation. *Organizational Behavior and Human Decision Processes*, 50(2), 179-211.
- Kemp, P.L., Goldman, A.C., & Walheim, C.N. (2024). On the role of memory in misinformation corrections: Repeated exposure, correction durability and source credibility. *Current Opinion in Psychology*, 56, 101783. <https://doi.org/10.1016/j.copsyc.2023.101783>
- Kim, S.S.Y., Vera Liao, Q., Vorvoreanu, M., Ballard, S., & Worman Vaughan, J.W. (2024). "I'm Not Sure, But...": Examining the Impact of Large Language Models' Uncertainty Expression on User Reliance and Trust. FAcCT '24, June 03–06, 2024, Rio de Janeiro, Brazil.
<https://doi.org/10.1145/3630106.3658941>
- Kox, E., Kerstholt, J.H., Hueting, T.F., & De Vries, P.W. (2021). Trust repair in human-agent teams: The effectiveness of explanations and expressing regret. *LLMs and Multi-Agent Systems*, 35, 30. <https://doi.org/10.1007/s10458-021-09515-9>
- Kox, E.S., Siegling, L.B., & Kerstholt, J.H. (2022). Trust development in military and civilian human-agent teams: The effect of social-cognitive recovery strategies. *International Journal of Social Robotics*, 13, 1323-1338. <https://doi.org/10.1007/s12369-022-00871-4>
- Kox, E., Van Den Boogaard, J., Turjaka, V., & Kerstholt, J.H. (2024). The journey or the destination: The impact of transparency and goal attainment on trust in human-robot teams. *ACM Transactions on Human-Robot Interaction*, 14(2), 23. <https://doi.org/10.1145/3702245>
- Leiser, F., Eckhardt, S., Leuthe, V., Knaeble, M., Madche, A., Schwabe, G., & Sunyaev, A. (2024). HILL: A hallucination identifier for large language models. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems - CHI '24*.
<https://doi.org/10.1145/3613904.3642428>
- Madhavan, P., Wiegmann, D.A., & Lacson, F.C. (2006). Automation failures on tasks easily performed by operators undermine trust in automated aids. *Human Factors: The Journal of the Human Factors and Economics Society*, 48(2), 241-256.
<https://doi.org/10.1518/001872006777724408>
- Madhavan, P., & Wiegmann, D.A. (2007). Similarities and differences between human-human and human-automation trust: An integrative review. *Theoretical Issues in Ergonomics Science*, 8(4), 277-301. <https://doi.org/10.1080/14639220500337708>

- Mallick, R., Flathmann, C., Duan, W., Schelble, B.G., & McNeese, N.J. (2024). What you say vs what you do: Utilizing positive emotional expressions to relay AI teammate intent within human-AI teams. *International Journal of Human-Computer Studies*, 192, 103355.
<https://doi.org/10.1016/j.ijhcs.2024.103355>
- Mar, R.A., Li, J., Nguyen, A.T.P., & Ta, C.P. (2021). Memory and comprehension of narrative versus expository texts: A meta-analysis. *Psychonomic Bulletin & Review*, 28, 732-749.
<https://doi.org/10.3758/s13423-020-01853-1>
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709–734.
- McAllister, D.J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, 38(1), 24-59.
<https://doi.org/10.5465/256727>
- Moriuchi, E. (2025). Deploying artificial intelligence for healthcare: An empirical study on patients' usage of health care service and AI's anthropomorphic structure form. *Technological Forecasting and Social Change*, 213, 124004.
<https://doi.org/10.1016/j.techfore.2025.124004>
- O'Neill, T.A., McNeese, N. J., Barron, A., & Schelble, B. (2022). Human–autonomy teaming: A review and analysis of the empirical literature. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 64(5), 904–938.
<https://doi.org/10.1177/0018720820960865>
- O'Neill, T.A., Flathmann, C., McNeese, N.J., & Salas, E. (2023). Human-autonomy Teaming: Need for a guiding team-based framework? *Computers in Human Behavior*, 146, 107762.
<https://doi.org/10.1016/j.chb.2023.107762>
- Ross, J. A. (2006). Trust makes the team go round. *Harvard Management Update*, 11(6), 3–6.
- Salas, E., Sims, D.E., & Burke, C.S. (2005). Is there a “Big Five” in teamwork? *Small Group Research*, 36(5), 555-599.
- Schaekermann M., Beaton G., Sanoubari E., Lim, A., Larson, K., & Law, E. (2020). Ambiguity-aware AI assistants for medical data analysis. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems - CHI '20*. <https://doi.org/10.1145/3313831.3376506>
- Schneider, M., Jessup, T.R., Stokes, C., Rivers, S., Lohani, M., & McCoy, M. (2017). The influence of trust propensity on behavioral trust. In: Poster session presented at the meeting of Association for Psychological Society, Boston, 2017.
- Spatharioti. S.E., Rothschild, D.M., Goldstein, D.G., & Hofman, J.M. (2023). Comparing Traditional and LLM-based Search for Consumer Choice: A Randomized Experiment. arXiv:2307.03744 [cs.HC]. <https://doi.org/10.48550/arXiv.2307.03744>

- Urban, J.M., Bowers, C.A., Mondary, S.D., & Morgan Jr., B.B. (1993). Effects of workload on communication processes in decision making teams; An empirical study with implications for training. *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting*.
- Verhagen, R.S., Neerincx, M.A., & Tielman, M.L. (2022). The influence of interdependence and a transparent or explainable communication style on human-robot teamwork. *Frontiers in Robotics and AI*, 9:993997. <https://doi.org/10.3389/frobt.2022.993997>
- Walter, N., & Tukachinsky, R. (2019). A Meta-Analytic Examination of the Continued Influence of Misinformation in the Face of Correction: How Powerful Is It, Why Does It Happen, and How to Stop It? *Communication Research*, 47(2), 155-177.
<https://doi.org/10.1177/0093650219854600>
- Wright, J. L., Chen, J. Y. C., Barnes, M. J., & Hancock, P. A. (2016). The effect of agent reasoning transparency on automation bias: An analysis of response performance. In: Lackey, S., Shumaker, R. (Eds.), *Virtual, Augmented and Mixed Reality. VAMR 2016. Lecture Notes in Computer Science (9740)* (pp. 465–477). Springer. https://doi.org/10.1007/978-3-319-39907-2_45
- Xiao, Y., & Yu, S. (2025). Can ChatGPT replace humans in crisis communication? The effects of AI-mediated crisis communication on stakeholder satisfaction and responsibility attribution. *International Journal of Information Management*, 80, 102835.
<https://doi.org/10.1016/j.ijinfomgt.2024.102835>

Appendix A: Instructions for watching the video

Participants read the text below as instructions. This text is translated to English for the purposes of this memo. Participants in the no information condition did not read the purple and green highlighted text. Participants with general information only condition also got to read the purple text. Participants in the general information and highlighted Chatbot answers got to read both the purple and green highlighted text.

In a moment you will observe an interaction of a Chatbot and a first responder in charge of a search-and-rescue mission in a rural area. The responder has received a report of smoke emanating from a building, but the cause is uncertain, and there could be people, animals, or hazardous materials inside.

The responder's task is to gather information about the situation at hand. The responder must determine the nature of the situation inside the building and assess potential risks. This includes identifying the source of the smoke, the presence of fire, any people inside the building, their current status, perhaps individuals in need of rescue, assessing the presence of hazardous substances that could exacerbate the situation, weather conditions, and evaluating the structural integrity of the building.

Not all information that the Chatbot provides is certain, this is due to the availability, reliability and sources the Chatbot has used. When enough sources of high quality converge, then that information can be seen as certain. The answers of the Chatbot are uncertain when the number of sources is limited or of low quality. In the interaction you will see approximately 75% of the information is certain and approximately 25% uncertain. The information that is highlighted in dark blue is certain, the information that is highlighted in light blue is uncertain.

Your goal is to take in all the information provided, so you can make an informed decision. In the end, you will have the option to send a team inside the building, to wait for help or to do nothing. Your decision will impact the safety of both the individuals inside the building and your own team. You must weigh the risks associated with different courses of action and make informed decisions about how to proceed.

Appendix B: Translation of the Chatbot interaction videos

Anthropomorphic language style chatbot, no uncertainty communication (fig. 3):

Chatbot: Hello, I am your assistant for today. We have received a report about a fire in a building in the centre of the city. I am here to help you to better understand the situation and to make an well informed decision. How can I support you?

You: Can you describe the structure and layout of the building?

Chatbot: Good question! Yes, the building has four floors. On the ground floor is a supermarket and a café. De upper three floors all contain 4 apartments, so 12 in total.

You: Is there visible smoke or fire at the location, and if so, where does it come from?

Chatbot: Yes, at the moment there is thick, dark grey smoke visible at the windows of the second floor, it is hard to ignore. The fire started in apartment 204.

You: Are there reports of people in the building? How many and where are they located?

Chatbot: yes, there are still 10 people inside. On the ground floor there are two employees and three clients being evacuated, but the residents on the second and third floor are in severe need of help, some are unable to leave independently.

Are there dangerous materials or chemicals stored in the

Send

Machine-like language style chatbot with uncertainty communication:

Chatbot: Hello, I am your assistant for today. We have received a report about a fire in a building in the centre of the city. I am here to help you to better understand the situation and to make an well informed decision. How can I support you?

You: Can you describe the structure and layout of the building?

Chatbot: Building structure: Four floors. Ground floor: Commercial space (supermarket and a café). Upper floors: Apartments (four per floor, 12 in total).

You: Is there visible smoke or fire at the location, and if so, where does it come from?

Chatbot: Smoke: Thick, dark grey smoke out of the windows of the second floor. Fire: Yes. Origin fire: Apartment 204.

You: Are there reports of people in the building? How many and where are they located?

Chatbot: Persons inside: 10 persons. Evacuation ground floor: Evacuation of two employees and three customers has started. Evacuation second and third floor: Residents are in severe need of help and some are unable to leave independently.

Are there dangerous materials or chemicals stored in the

Send