Proceedings of the Human Factors and Ergonomics Society Annual Meeting
2025, Vol. 69(1) 1579–1585

DOI: 10.1177/10711813251358255
journals.sagepub.com/home/pro
Sage

Explaining Social Influences in Human-Human-AI Teams: Convergence From Grounded Theory and Structural Topic Modeling

Trevor Patten[1], Stephanie Kim[1], Emanuel Rojas[1], and Mengyao Li[1]

Abstract

Understanding how trust spreads in Human-Human-AI (HHA) teams is critical to designing adaptive AI teammates. While prior research demonstrates that one human's trust can influence another—a phenomenon known *as trust contagion*—the mechanisms in human-AI teams remain unclear. We examined how trust cues affect decision-making in triadic teams consisting of a participant, a confederate, and an AI agent. The confederate's expressed trust in the AI (high, low, neutral) was experimentally manipulated, and team interactions were analyzed using Grounded Theory (GT) and Structural Topic Modeling (STM). GT captured rich, contextual themes, while STM identified semantic patterns at scale. Results showed that high-trust teams reached rapid consensus with minimal discussion, whereas low-trust teams engaged in deliberative trust calibration. We found conceptual convergence between GT and STM, demonstrating the value of integrating human-driven and computational methods to understand trust contagion and proving design implications of AI teammates that adapt to trust dynamics.

Keywords

grounded theory, structural topic modeling, trust contagion, human–AI teaming

Introduction

Trust is essential to the success of Human-AI Teams (HATs), enabling collaboration, decision coordination, and task sharing. Yet trust is not formed in isolation—human teammates influence each other's trust attitudes through a process known as trust contagion. Trust contagion is the process of one agent's trust in a system influencing another human trustor in a process born from emotional contagion theory (Barsade, 2002). When trust is misaligned, even a well-performing AI may be underutilized due to interpersonal distrust; conversely, unjustified trust propagation can result in over-reliance and system failure. Prior research on HATs modeled how trust develops in multi-agent teams (Guo et al., 2024). Despite success in modeling trust, the extent to which social factors, like other trustors, are not well represented (Al-Ani et al., 2014; Feese et al., 2012). Qualitative research on the development of trust has provided potential explanations for how trust develops in hybrid groups. Duan et al. (2025) analyzed how trust and distrust spread in human-AI teams. They successfully identified four mechanisms of trust though particular modes of trust contagion, social information processing and reciprocity, may be unique to humans acting as the source of contagion compared to an AI as the source of the contagion. However, most computational trust models and empirical studies still treat trust as an individual construct, neglecting how team members' attitudes dynamically shape one another's behavior and judgment toward AI systems. To address this gap, we investigated how trust cues from a human confederate influenced team-level trust dynamics and decision strategies in a triadic Human-Human-AI (HHA) team. Our study used a mixed-methods approach, combining Grounded Theory and Structural Topic Modeling to uncover both human-interpretable themes and scalable semantic patterns in trust-related communication.

Grounded Theory and Structural Topic Modeling

To understand the mechanisms of trust contagion, we employed a convergent mixed-method approach that integrates Grounded Theory (GT) and Structural Topic Modeling

[1]Georgia Institute of Technology, Atlanta, USA

Corresponding Author:
Trevor Patten, Georgia Institute of Technology, 2310 Katharine Stinson Drive, Atlanta, GA 27695-7001, USA.
Email: tpatten7@gatech.edu

(STM). GT is a qualitative method that develops theories through a bottom-up approach directly from the data (Corbin & Strauss, 2008). GT follows an iterative process of constant comparison until thematic saturation is reached. It is well-suited for capturing social processes such as trust negotiation and interpersonal influence but is often criticized for its subjectivity, labor-intensive nature, and limitations in scalability. STM complements GT by identifying latent topics in large text corpora using probabilistically mapping words to topics (Roberts et al., 2019). It allows researchers to discover high-level semantic patterns and examine how topic prevalence varies across experimental conditions. When used together, GT and STM offer methodological triangulation: GT provides deep interpretive insight, while STM lends empirical generalizability and cross-condition comparisons (Baumer et al., 2017). To explain trust contagion, we applied a mixed-method approach by integrating GT and STM to code in-game conversations and post-task interviews to uncover behavioral and cognitive themes. This integration enables both theory-building and scalable pattern discovery in understanding how social cues shape trust in Human-Human-AI teams. Our aim for this study was to explore how trust contagion develops across different trust attitudes using a hybrid analysis method of GT and STM.

Method

A team of three—one participant, one confederate, and one AI teammate—played a ten-round trust-based resource allocation game. The study followed a 2 (AI reliability: high vs. low, within-subjects) × 3 (confederate trust: high, low, neutral, between-subjects) design. For reliability, the AI performed with 100% accuracy in the high-reliability condition and 60% in the low-reliability condition. To foster initial trust, the high reliability condition was always presented before the low reliability condition. To manipulate trust contagion, the confederate enacted three trust levels: neutral (fact-based comments), high (positive remarks about the AI), and low (skeptical remarks about the AI).

Participants

We collected a total sample size of $N = 42$ from a student recruitment platform. All participants' ages ranged from 18 to 24 years old. Every condition had the same number of male and female participants. Participants were compensated with one research credit or ten dollars of their choosing.

Procedure

After signing a consent form, participants were acquainted with the confederate and issued a briefing video about the task and their role as commander. The task began shortly after. The task lasted a total of 10 rounds. Within each round, the participant made three decisions: (1) how many points should they give to the AI, (2) how to split contributions between the team and individual rover, and (3) how many points they think the AI would give to the team rover. After the final round, participants engaged in an interview with the experimenter and debriefed on the experiment and the identity of the confederate.

Data

We collected conversational data from the resource allocation game and post-study interviews for further analysis. As the commander, the participant engaged in strategic dialogue with the confederate to navigate resource allocation, deliberating on how to distribute points between an individual rover and a team rover. The human team's goal was to gain the highest total score by optimally allocating points to either an individual rover or a team rover. Participants could also allocate points to the AI with the potential to double and return points but could also fail to do so, introducing uncertainty into the decision-making process. Following the game, participants completed a semi-structured interview where they reflected on their decision-making, interactions with the AI and confederate, and any influence the confederate had on their choices.

Grounded Theory

We adopted the Straussian framework (Corbin & Strauss, 2008) and adapted a codebook development method for multiple coders (Díaz et al., 2023). Each game round served as a data unit eligible for coding. This unit size was agreed upon to balance identifiable turns in conversation while also including enough dialogue to analyze. Post-game interviews were segmented and analyzed with each question-response pair and its follow-ups as a data unit. Given the complexity of conversational data and its multiple possible interpretations, we identified a few non-mutually exclusive codes to capture overlapping themes. Our coding method allowed for the same instance of data to accommodate multiple non-mutual exclusive codes to capture the relative complexity of analyzing a discussion of strategy.

Open coding. The initial open coding process used 10 participants to establish a baseline for serial codebook development. The first coder identified the instances of data for analysis, generated the codebook, and passed the instances of data and the codebook to the next coder. Subsequent coders were able to add or remove codes without seeing where the prior coder applied codes to data (Díaz et al., 2023); thus, the codebook used by the last coder was more sophisticated than the codebook used by the first coder.

Axial coding. The axial coding phase occurred alongside the open coding phase (Corbin & Strauss, 2008). We analyzed an additional eight participants serially between coders to push the codebook to saturation. After each round of coding, coders met to discuss the new codes they added to revise

Table 1. Grounded Theory Axial Codes Descriptions.

Code	Description
Confederate considerate	Participants directly requested decision-support from the confederate
Confederate influence	Participants either accepted or rejected the decision-support of the confederate
a. Receptive	Participants accepted the confederate's decision support
b. Resistant	Participants rejected the confederate's decision support
Minimal debate	Coders identified minimal conflict or deliberation between human teammates
Attitude similarity	Participants were either aligned or unaligned with the trust attitude confederate
a. Similar humans	Participants described a high degree of agreement with the confederate
b. Dissimilar humans	Participants described a high degree of disagreement with the confederate
Seeking AI abandonment	Participants identified trusting the AI as redundant to powering the team rover
Error salience	Participants drew additional attention to an error the AI made
Reliability assessment	Participants described the AI in terms of reliability, consistency, or predictability
AI goal inference	Participants described the AI is in pursuit of a goal or fulfilling its own motive

the final codebook, which contains ten major axial codes. We completed the axial coding procedure with all coders recoding the initial 10 participants from the open coding phase with the final codebook. Table 1 is an overview of our codebook. To preserve semantic consistency, certain codes were designated as mutually exclusive. In the same turn of conversation, a participant could not be both similar in attitude but also dissimilar. The codes for Confederate Influence and Attitude Similarity reflect this with mutually exclusive subcodes. An instance of data could only be coded with Minimal Debate if it lacked Confederate Considerate and Confederate Influence as those codes require the exertion of influence, whereas harmony would imply a shared mental model.

The simple intercoder reliability of a majority response with two out of three coders agreeing on the exact set of codes was 61%. From this point, a single coder coded the remaining 29 participant's data with the final set of axial codes. This coder analyzed a total of 640 instances of data.

Structural Topic Modeling

We prepared for STM analysis by pulling our text data into the *Quanteda* and *STM* packages in R. First, we segmented text data by conversation turns, cleaned and tokenized the text by removing punctuation, numbers, English stop words, and custom stop words (e.g., "rover," "hmm"). All words were lemmatized to their roots to avoid any duplicates (e.g., try, tried, trying) with *spacyr* (Honnibal et al., 2020). Next, we selected the ideal number of topics with the *searchK* function from the *stm* package with the between-subjects variable as our covariate. We iterated over *K* from 3 to 15 in increments of 2 and identified 11 as the ideal number of topics using metrics of held-out likelihood, low residuals, and maintaining semantic coherence. After determining the number of topics, we passed the data into the *stm* function searching for prevalence with the between-subjects variable as our covariate. We identified three topics as semantically incoherent resulting in their ejection. Topics that had a strong conceptual connection to each other were merged. Four topics were deemed semantically similar enough to be considered as two topics resulting in a total of six topics. Topics were labeled by generating probable quotes associated with each topic's FREX terms with the *findThoughts* function at a .1 threshold (Weston et al., 2023).

Results

Grounded Theory Results

Our GT analysis produced eight axial codes. We identified two major themes to unify the codes: AI-Related and Human Team-Related. The AI-Related axial codes all described how teams responded to the AI's behaviors. Human Team-Related codes described how humans reasoned with each other to reach a decision for interacting with the AI. Our process of coding revealed that the prevalence of particular axial codes varied with the between-subjects condition. We identified where particular axial codes were highly prevalent by observing the frequency of counts as descriptives. Ultimately, our GT analysis produced a model for how trust contagion propagates as a function of a teammate's level of trust.

The model in Figure 1 describes how an AI teammate generates information that the human team factors into their subsequent decision to trust the AI. The loop begins with the human team deciding on a strategy for trusting the AI. The outcome of trusting the AI is understood by humans in terms of reliability, the AI's objectives, and the salience of the error. Team members must use their understanding of the AI to inform a subsequent trust decision. How they reach their decision is a function of the individuals seeking decision support and the team's similarity on attitudes. Ultimately, a decision on how to interact with the AI is made, either with minimal debate or through critical discussion on an optimal approach.

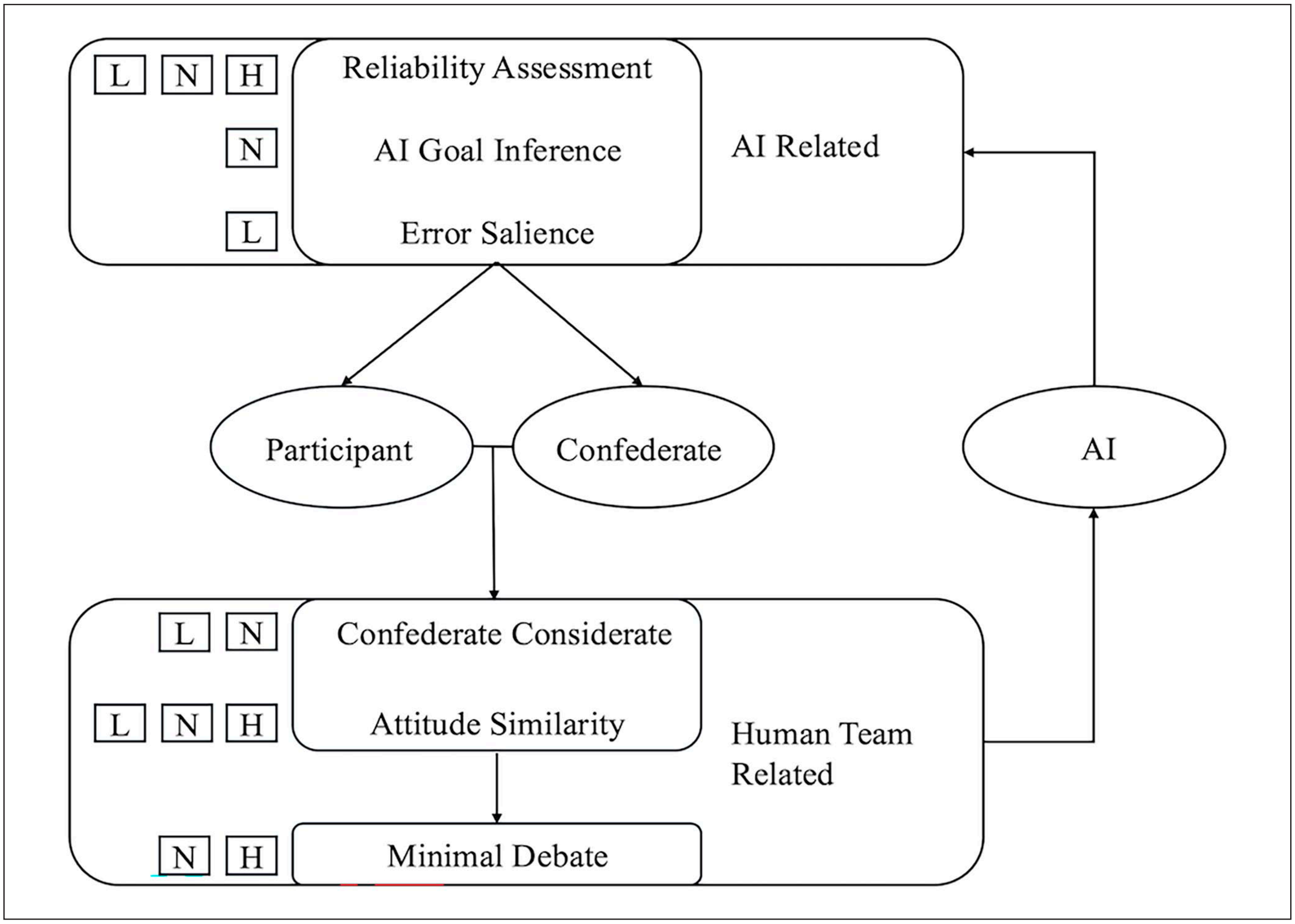

Figure 1. Conceptual framework of trust contagion based on grounded theory results.
Note. H = high; N = Neutral; L = low denote the most common trusting condition for axial codes.

AI-Related Factors included codes such as *Reliability Assessment, AI Goal Inference*, and *Error Salience*. These codes reflected how participants evaluated the AI's behavior to inform trust and strategy decisions. Reliability was a constant concern across all three trust conditions; however, *AI Goal Inference* and *Error Salience* were most prevalent in one condition each. *Error Salience* was most prevalent with a low-trust confederate. The *AI Goal Inference* code was most prevalent with a neutral confederate with phrases that acknowledged the AI in pursuit of a goal like the following:

Participant: I think the AI is trying to reach 200.

Human Team-Related codes captured interpersonal coordination and social influence. Codes such as *Confederate Considerate, Attitude Similarity*, and *Minimal Debate* revealed how participants either consulted or aligned with their teammate. We noticed participants frequently asked low and neutral-trust confederates for more decision-making support compared to high-trust confederates. Neutral confederates were unable to provide opinions on their level of trust though participants were still interested to know what their partner was thinking. In contrast, high-trust condition teams were verbally efficient with very few words exchanged for the team to settle on a decision. Often, the high-trust confederate teams maintained a strategy for multiple rounds; thus, limited communication was required. For example:

Participant: All right, it all right, so 10 [points] again?

High-Trust Confederate: Yeah, sure.

Neutral and high-trust confederates often experienced less deliberation when making a decision than the intensive discussions seen with low-trust confederates, though the reasons for this appear to be related to participants acknowledging neutral teammates as not worth consulting and high trust teammates as willing to maintain a successful strategy.

STM Topics

We produced six topics from our STM analysis that described *team actions* and *emergent team states*. We identified team three actions: *Team Goal Prioritization*, teams strategizing about powering the team rover; *Strategy Adaptation*, teams altered their strategies based on prior AI behaviors, adjusting their allocation patterns; and *Cooperative Decision-Making*, teams employing methods of cooperation. The three emergent team states were as follows: *Team Formation*, participants identifying with team; *Experiencing Uncertainty*, teams experience doubt towards the AI; and *Human Agreement*, where the human team exchanged minimal information to confirm team actions. We used FREX terms and probable quotes to interpret our STM topics and label them.

Team Goal Prioritization. This topic described the heavy emphasis placed on reaching the point threshold required to activate the team rover. The top FREX terms were: "*sense*," "*activate*," "*active*," "*reach*," "*already*," "*invest*," "*triple*." These terms identify the dominance of the team rover as an objective. Participants often identified that reaching the team rover point threshold as the most optimal goal. For example:

> Probably not the best we could have done because we activated the rover, but I'm sure we could have gotten higher points through a different strategy.

Method for Strategy Creation. Participants often had a plan of action to achieve the greatest possible final score. Participants calibrated the number of points to assign to the AI based on prior performance. Top FREX words included "*teammate*," "*human*," "*commander*," "*decision*," "*opinion*," "*mostly*," "*take*." This was expected given the importance of reliability for trust calibration, as shown in the quote below:

> My decisions were mainly influenced by the outcomes from the AI teammate. Specifically, I understood my human teammate's skepticism, but at the same time, we wanted to reach a goal which was pretty easy to reach.

Accommodating for Cooperation. Participants had their own methods for establishing a cooperative dynamic with the confederate. Top FREX words included "half," "put," "allocate," "mean," "great," "number," and "least." A strategy of compromise was common though occasionally participants would totally align themselves with the confederate's attitude without an argument.

Team Formation. The FREX words included "us," "try," "course," "say," "wow," "see," "behave" describing how participants aligned themselves with the confederate's perceived trust level. These terms indicated that over the course of the task, participants converged on a cooperative dynamic with the confederate.

Experiencing Uncertainty. Teams experienced constant uncertainty. Top FREX words included "time," "double," "last," "guess," "whole," "hopefully," "chance." Participants often tested the AI's reliability through guesswork. Other participants took greater risks as their chances of powering the team rover grew slimmer throughout the task. Words like "guess," "hopefully," and "chance" express the level of uncertainty participants had during the task:

> And then I guess, yeah, it has 22, maybe like 18, I think. Something around that. Hopefully it picks up on it.

Method of Cooperation. Similar to trends identified from the GT analysis, we noticed the relative rarity that a participant would act without seeking assurance from the confederate. These interactions presented as a sign of both the confederate and the participant acting as one unit in agreement. FREX words signal agreeableness with words like "*agree*," "*good*," "*continue*," "*cool*," "*error*," "*agree*," "*good*," "*idea*," and "*halfway*."

Results Comparison and Discussion

Our analyses revealed conceptual overlaps in the GT and STM results (see Figure 2). The GT analysis contributed towards an overall process for trust contagion in team settings, but the prevalence of these ideas was better supported by the computational approach of STM. These analyses conceptually converge, which allowed us to clarify the significance of our observations. We identified three core themes common across both analyses: (1) Trust Calibration through Reliability Inference, (2) Mechanisms of Social Influence in Trust Contagion, and (3) Human Team Efficiency via Attitude Alignment.

For the first theme of trust calibration, GT codes such as *Reliability Assessment* and *AI Goal Inference*, and STM topics like *Method for Strategy Creation* and *Experiencing Uncertainty*, revealed that participants continuously updated their trust based on AI behavior. This calibration process was most pronounced in the low-trust condition, where participants engaged in deliberate discussions and recalibrated their strategy following AI errors. STM reinforced this by showing heightened uncertainty semantics in low-trust teams. Reliability appeared to be a major focus within low-trust teams through more frequent strategic discussions than high-trust teams. This dynamic aligns with prior literature on dynamic trust calibration processes in response to performance and social signals over time.

Our second theme uncovered the differential processing of trust contagion. The GT code *Confederate Considerate* revealed that participants often sought or deferred to the confederate's judgment, particularly in the neutral and low-trust conditions. STM topics like *Method of Cooperation* further emphasized this behavioral convergence. Both analyses captured how the confederate's expressed trust shaped the

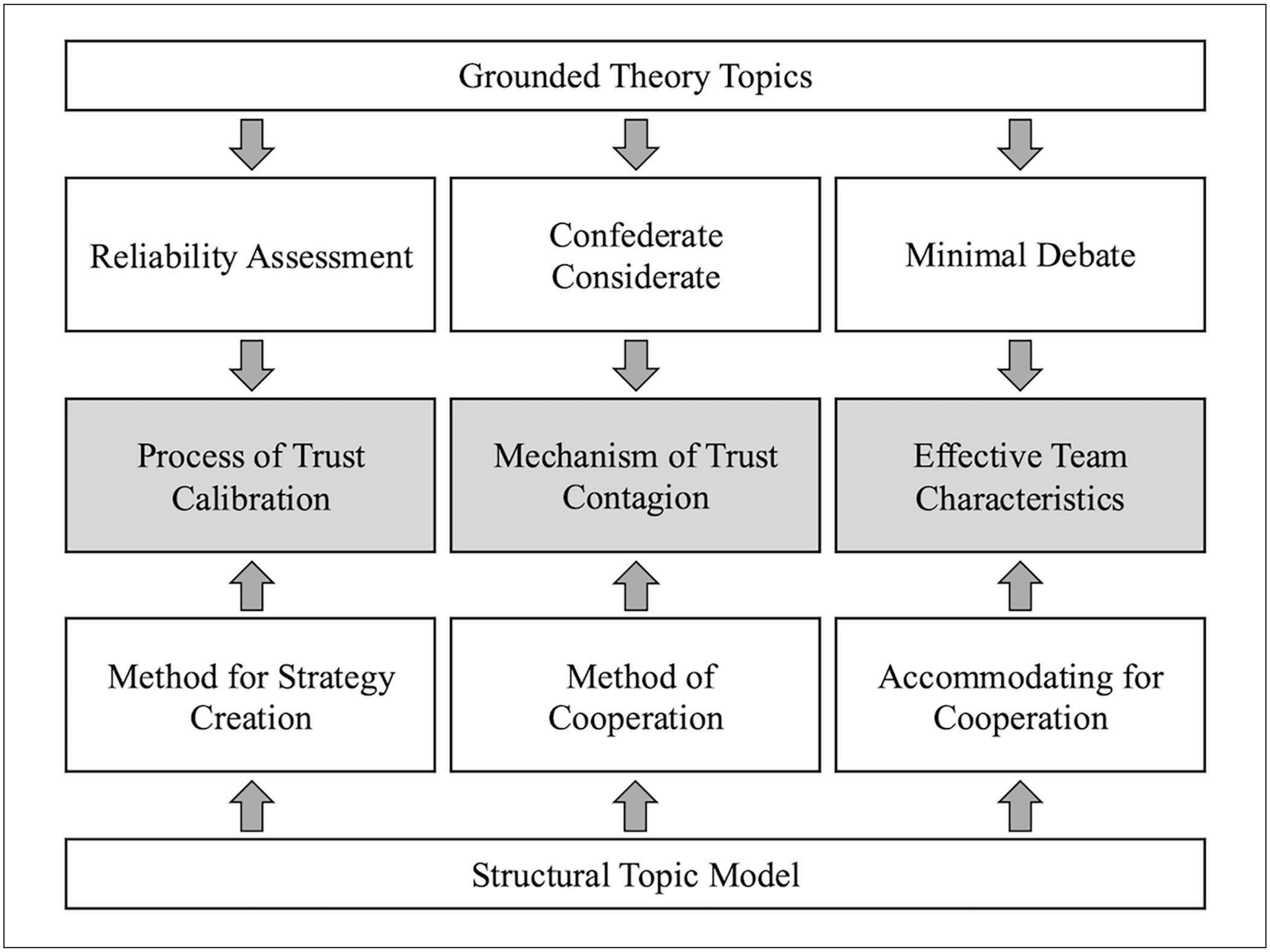

Figure 2. Converged topics (middle) from GT results (top) and STM results (bottom).

participant's behavior, even when the participant held formal authority. Notably, both high-trust and low-trust confederates influenced participants, but the resultant behaviors differed: High-trust confederates fostered rapid, implicit trust alignment, whereas low-trust confederates triggered deliberative, explicit discussions about AI reliability. This pattern matched the mechanism of trust contagion from Duan et al. (2025) called "reciprocity," whereby trust expressions signal a normative stance that others respond to. Reciprocity engenders parties to adopt each other's relative level of trust. Expressions of high trust communicate that there is nothing to be concerned with and the strategy can be maintained, and expressions of low trust communicate to teammates that something is amiss with the current level of trust so effort must be directed towards reevaluating the current level of trust (Duan et al., 2025).

A third theme involved the efficiency and harmony of human team dynamics. GT codes such as *Minimal Debate* and *Attitude Similarity* and STM topics like *Accommodating for Cooperation* highlighted that high-trust teams operated with reduced verbal exchanges—often confirming decisions with brief affirmations. This reflects a shared mental model (Endsley, 1995), where less communication is needed once alignment is established. Conversely, low-trust teams demonstrated more explicit negotiation, suggesting that trust misalignment increases cognitive and communicative load. These patterns underscore that trust contagion influences not only the content of team decisions but also the style and intensity of collaboration.

There are two main limitations in this study. First, while we experimentally manipulated trust contagion using a scripted confederate, some off-script interactions were necessary to maintain organic dialogue. Although the confederate followed condition-specific scripts closely, variability in delivery may have introduced uncontrolled social cues. Future research could mitigate this limitation by either leveraging participants' dispositional trust or inducing trust differentially at baseline, thereby avoiding the need for a confederate altogether. Second, GT analyses are challenging in part due to the complexity of data they can capture. Our coding scheme evolved to a more general scope over time to help coders reach agreement.

Conclusion

Our study examined trust contagion in Human-Human-AI (HHA) teams by integrating Grounded Theory (GT) and Structural Topic Modeling (STM) to uncover the mechanisms behind the trust spread. GT excels in uncovering contextual mechanisms, while STM offers scalable, condition-sensitive trends. Through this mixed method approach, we found that low-trusting and neutral teammates prompted greater deliberation and strategy revision, while high-trust teams exhibited implicit agreement and minimal discussion, aligning with shared mental model theory. This suggests that low-trust communication fosters explicit trust calibration, whereas high-trust communication reinforces existing trust through implicit reinforcement. Our proposed framework describes trust contagion as a dynamic interplay between AI behavior, human trust alignment, and team communication patterns. Our converged findings from STM and GT demonstrate that qualitative and computational methods can converge on conceptually similar ideas, similar to Baumer et al. (2017). We believe that this mixed method could be effective when analyzing large volumes of data for complicated behaviors like trust contagion. By understanding how trust spreads in mixed human-AI teams, we can develop AI teammates that dynamically adjust their behaviors to maintain appropriate trust levels and optimize team performance.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Emanuel Rojas https://orcid.org/0009-0003-9979-0013
Mengyao Li https://orcid.org/0000-0002-0819-4693

References

Al-Ani, B., Marczak, S., Redmiles, D., & Prikladnicki, R. (2014). Facilitating contagion trust through tools in Global Systems Engineering teams. *Information and Software Technology*, *56*(3), 309–320. https://doi.org/10.1016/j.infsof.2013.11.001

Barsade, S. G. (2002). The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly*, *47*(4), 644–675. https://doi.org/10.2307/3094912

Baumer, E. P. S., Mimno, D., Guha, S., Quan, E., & Gay, G. K. (2017). Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? *Journal of the Association for Information Science and Technology*, *68*(6), 1397–1410. https://doi.org/10.1002/asi.23786

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research (3rd ed.): Techniques and procedures for developing grounded theory*. SAGE Publications, Inc.

Díaz, J., Pérez, J., Gallardo, C., & González-Prieto, Á. (2023). Applying inter-rater reliability and agreement in collaborative grounded theory studies in software engineering. *Journal of Systems and Software*, *195*, Article 111520. https://doi.org/10.1016/j.jss.2022.111520

Duan, W., Zhou, S., Scalia, M. J., Freeman, G., Gorman, J., Tolston, M., McNeese, N. J., & Funke, G. (2025). Understanding the processes of trust and distrust contagion in Human–AI Teams: A qualitative approach. *Computers in Human Behavior*, *165*, Article 108560. https://doi.org/10.1016/j.chb.2025.108560

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*(1), 32–64. https://doi.org/10.1518/001872095779049543

Feese, S., Arnrich, B., Tröster, G., Meyer, B., & Jonas, K. (2012). *Quantifying behavioral mimicry by automatic detection of nonverbal cues from body motion* [Conference session]. Proceedings of ASE/IEEE International Conference on Social Computing, Kaifeng, China, pp. 520–525. https://doi.org/10.1109/SocialCom-PASSAT.2012.48

Guo, Y., Yang, X. J., & Shi, C. (2024). TIP: A trust inference and propagation model in multi-human multi-robot teams. *Autonomous Robots*, *48*(7), 20. https://doi.org/10.1007/s10514-024-10175-3

Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). SpaCy: Industrial-strength natural language processing in Python [Python]. https://github.com/explosion/spaCy

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An R package for structural topic models. *Journal of Statistical Software*, *91*(2), 1–40. https://doi.org/10.18637/jss.v091.i02

Weston, S. J., Shryock, I., Light, R., & Fisher, P. A. (2023). Selecting the number and labels of topics in topic modeling: A tutorial. *Advances in Methods and Practices in Psychological Science*, *6*(2), 25152459231160105. https://doi.org/10.1177/25152459231160105