Trusting Autonomous Teammates in Human-AI Teams - A Literature Review

Wen Duan
School of Computing
Clemson University
Clemson, South Carolina, USA
wend@clemson.edu

Christopher Flathmann
Human Centered Computing
Clemson University
Clemson, South Carolina, USA
cflathm@clemson.edu

Nathan McNeese
Human-Centered Computing
Clemson University
Clemson, South Carolina, USA
mcneese@clemson.edu

Matthew J Scalia
Human Systems Engineering
Arizona State University
Mesa, Arizona, USA
matthew.scalia@asu.edu

Ruihao Zhang
Human Systems Engineering
Arizona State University
Mesa, Arizona, USA
rzhang82@asu.edu

Jamie Gorman
Arizona State University
Tempe, Arizona, USA
jcgorman@asu.edu

Guo Freeman
School of Computing
Clemson University
Clemson, South Carolina, USA
guof@clemson.edu

Shiwen Zhou
Human Systems Engineering
Arizona State University
Mesa, Arizona, USA
shiwen.zhou@asu.edu

Allyson Ivy Hauptman
TRACE Lab
Clemson University
Clemson, South Carolina, USA
ahauptm@g.clemson.edu

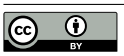
Xiaoyun Yin
Arizona State University
Gilbert, Arizona, USA
shaoinn1@gmail.com

Abstract

As autonomous AI agents become increasingly integrated into human teams, the level of trust humans place in these agents - both as a piece of technology and increasingly viewed as teammates - significantly impacts the success of human-AI teams (HATs). This work presents a literature review of the HAT research that investigates humans' trust in their AI teammates. In this review, we first identify the ways in which trust was conceptualized and operationalized, which underscores the pressing need for clear definitions and consistent measurements. Then, we categorize and quantify the factors found to influence trust in an AI teammate, highlighting that agent-related factors (such as transparency, reliability) have the strongest impacts on trust in HAT research. We also identify under-explored factors related to humans, teams, and environments, and gaps for future HAT research and design.

CCS Concepts

• **Human-centered computing** → **HCI theory, concepts and models**; • **General and reference** → **Surveys and overviews**.



CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713527

Keywords

Artificial Intelligence, Human-Autonomy Teaming, Human-AI Teaming, Human-Agent Teaming, Trust, Trust in Autonomous Teammates

ACM Reference Format:
Wen Duan, Christopher Flathmann, Nathan McNeese, Matthew J Scalia, Ruihao Zhang, Jamie Gorman, Guo Freeman, Shiwen Zhou, Allyson Ivy Hauptman, and Xiaoyun Yin. 2025. Trusting Autonomous Teammates in Human-AI Teams - A Literature Review. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 23 pages. https://doi.org/10.1145/3706598.3713527

1 Introduction

Human-AI teaming (HAT) involves humans and autonomous AI agents working interdependently, where both are distinct, recognized team members with unique roles, aiming to achieve a shared goal [100]. Recently, HCI and related fields have seen an exponential increase in the number of research studies investigating trust in human-AI teaming and collaboration (e.g., [50, 87, 94, 116, 118, 150, 151]), as trust plays a critical role in both effective teamwork [28, 80, 82] and effective use of AI technology [6, 43].

Indeed, HAT research offers an interesting intersection of **teamwork** and **AI technology** domains, both of which regard trust as important, but in distinct ways. In the teamwork domain, trust is often studied from an **interpersonal** perspective, where teammates are trusted as equal partners (e.g., [50, 136, 151]), and where trust is influenced by the presentation and role of AI teammates [96–98]. Conversely, in the AI technology domain, trust is viewed

as a **function** of performance [43, 71, 149], often depending on the AI's ability to fulfill its designed purpose [55]. The differing perspectives on trust reduce clarity and consistency across research efforts, despite the field's rapid growth. This lack of alignment hinders a coherent understanding of trust in HATs, complicates the development of unified theories, and impedes the creation of best practices for designing AI systems that effectively build trust.

To remedy these and provide a coherent understanding of trust in HAT research, this paper provides a systematic review of empirical research that investigated the factors influencing humans' trust in an AI teammate in HAT contexts. We leverage the identification and analysis of these factors to ensure that the research and application of HATs can adequately consider the AI agent, human, team, and environmental factors that impact trust in HATs. Further, this paper serves to identify conceptual and operational inconsistencies across the communities to further drive and standardize future research efforts. Based on these objectives, the following research questions were posed to guide this work:

- **RQ1:** What are the common and different ways trust is conceptualized and operationalized in HAT research?
- **RQ2:** What factors has HAT research shown to impact trust, and how strong is their influence?

Through this review of 57 papers published from 2008 to 2022, we make several contributions to HCI research on HATs. First, we consolidate the existing scientific knowledge of human trust in AI teammates examined in HAT research. In turn, this work acts as a fundamental milestone that synthesizes a currently unorganized domain, which will provide researchers and practitioners with holistic knowledge about trust in HATs. Second, we clarify the existing variety of trust operationalizations and metrics within HAT research, and provide guidelines for future HCI and CSCW work to achieve conceptual-operational alignment in measuring trust in an AI teammate in HATs. Further, this work enhances the awareness and utilization of validated perceptual and behavioral trust metrics in HAT research. We also identify and quantify the factors that are empirically known to impact the trust humans form in their AI teammates in a HAT, which will help HCI researchers and practitioners leverage these factors to foster and manage trust in future HATs.

2 Scope of Human-AI Teaming and Related Constructs

Before proceeding with the review, it is crucial to contextualize and define the scope of Human-AI Teaming (also referred to in this paper as Human-Autonomy or Human-Agent Teaming [1]) within broader research domains. This includes clarifying its relationship to and distinctions from human-automation and human-AI interaction constructs.

In their general overview of human-autonomy teaming, O'Neill et al. [100] canonically define it as the interdependent placement of one or more autonomous technologies alongside one or more humans to complete a shared goal. Despite being a distinct construct, human-autonomy teaming has emerged from automation research as a natural progression driven by advancements in technology. Automation research initially focused on systems "designed to accomplish a specific set of largely deterministic steps (often in a repeated pattern) in order to achieve one of an envisaged and finite set of predefined goals" ([114], p.380). As technology advanced, the scope expanded to autonomy research, emphasizing systems capable of making independent analysis, suggestions or decisions, adapting to dynamic environments, and collaborating with humans [86]. This shift reflects a progression from rigid task execution to more intelligent, flexible, and context-aware capabilities. Take the domain of AI-assisted decision-making for example, an automated loan approval system (i.e., automation) might process applications only when commanded by humans and adheres strictly to predefined criteria, such as income requirements, approving or rejecting applications without any contextual reasoning. In contrast, an autonomous AI would be (or perceived to be) capable of independently analyzing a client's entire financial history, evaluating complex scenarios, dynamically adjusting its criteria [139], and adapting its recommendations as new data becomes available [107]. This **independence** from rigid, human-defined rules or commands, along with the ability to leverage data that may or may not be accessible to or recognized by the humans interacting with the system, is what sets autonomy apart from automation.

Prior reviews in the human-automation space have noted that trust is one of the most important factors to effective use of automation [51, 114], and additional reviews have highlighted how this importance will extend to more autonomous and complex AI systems [8, 43, 88]. Human-AI teaming research should build on existing work on human-automation and human-AI interaction while justifying its unique focus. Specifically, at the core of human-AI teaming is the **interdependence** between human and AI teammates, highlighting how autonomous agents are operationalized through collaborative systems [6]. Unlike broader concepts of human-AI interaction or collaboration, human-AI teams are uniquely defined by their emphasis on complementary roles, where humans and AI work interdependently to achieve shared goals [7, 133]. This interdependence necessitates unique design considerations for AI technologies. Specifically, AI teammates must be designed with interdependent functionality, ensuring that the performance of each teammate is intrinsically linked to the others through role assignment, synchronized actions, etc. [24, 31]. Beyond functional capabilities, AI teammates must also foster social and interpersonal dynamics within the team [33]. What sets an AI teammate apart from other AI applications is its ability to interdependently perform individual tasks while supporting the shared goals of the team both in terms of taskwork and teamwork [34, 39, 151].

3 Review Method

To address the research questions, we conducted the review following the guidelines provided in the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA; [101]). In the following subsections, we outline the procedures, which include the literature search and sampling, paper screening and eligibility

[1] In this review, we use autonomy, agent, and AI interchangeably, as these terms indicate the level of autonomy required for a study to be considered to align with the definition of HATs [100]. However, we acknowledge that not all research differentiate these terms and instead use human-automation, human-machine (terms that connote a lower level of autonomy) to represent what would have been considered autonomy. To ensure a comprehensive review, we included these terms in our search to broaden the scope of the identified literature.

assessment, data extraction and coding, as well as the process for calculating effect sizes to quantify the factors influencing trust.

3.1 Literature Search and Sampling

We first identified an appropriate set of search terms to ensure the search results were as comprehensive as possible. In addition to the two key terms that define the scope of this review - "team" and "trust" - we included a variety of terms researchers have used for human-AI teaming, such as "human-autonomy", "human-agent", "human-machine", "human-automation" [2]. We conducted searches in three databases: ACM Digital Library and IEEE Xplore, both commonly used in HCI research, and Web of Science, which spans a broader range of publication venues across various disciplines.

We conducted two rounds of literature search and screening, because during the first round, which occurred between 5/18/2022 and 6/20/2022, the authors identified that a large number of articles that included trust as a dependent variable also prioritized other dependent variables, such as performance. In turn, multiple articles that empirically explored trust as a secondary objective [3] may not have explicitly discussed trust in their metadata. To ensure a more comprehensive coverage, we performed a second round of search and screening on November 11, 2022 in both full text and metadata areas. To ensure that the second round of search and screening yielded only unique papers, we systematically compared all records from Round 2 against those screened at every stage of Round 1. Each eligibility criterion at each stage of Round 1 was organized into corresponding sub-folders in Zotero, enabling efficient cross-checking. Duplicates identified in the combined meta-folders of Round 1 and Round 2 were removed from the Round 2 dataset to ensure no overlap. A full list of search terms and strategies is provided in Table 1.

3.2 Paper Screening and Eligibility Assessment

We conducted manual screening and selection process involving abstract/title screening and full-text screening. During abstract screening, five of the authors independently assessed all pre-filtered records from the databases for eligibility. They then collaborated to resolve disagreements and refine the eligibility criteria, initially defined by the first author. These iterative discussions resulted in the final version of the inclusion (IC) and exclusion (EC) criteria.

- **IC1:** It must be empirical research involving human participants.
- **IC2:** It must involve at least one autonomous agent (or perceived to be autonomous) and at least one human.
- **IC3:** The human participants must work with (or imagine that they work with) autonomous agent(s) interdependently on a task toward a common goal.
- **IC4:** The autonomous agent must demonstrate at least partial autonomy (or perceived autonomy) on Parasuraman et al. [103]'s Level of Automation (LOA) continuum.
- **IC5:** It must explicitly measure humans' trust in or the trustworthiness of the AI agent as a dependent variable, and report results of trust-related measures. [4].
- **EC1:** Physical forms of autonomous agents (e.g., physical robots) should be excluded.
- **EC2:** Tele-operated or remote-controlled agents (e.g., drones, telepresence or surgical robots) should be excluded.
- **EC3:** Off-topic.

Of these criteria, IC2-4 were used to determine that the article can be deemed HAT research following O'Neill and colleagues' [100] definition of Human-Autonomy Teams, where there is at least one human and one autonomous agent, each recognized as occupying a distinct role within the team, working interdependently to achieve a common goal. IC5 ensures that we effectively quantify the factors influencing trust in HATs to perform a robust statistical analysis. This criterion also narrows the focus to experimental research. For EC1, it is important to note that physical embodiment is different from visual representation (i.e., avatar) in the virtual world. Physically embodied agents are tangible, physical entities that occupy space in the real world and can interact directly with people in the physical world. Equipped with sensors, they can perceive their environment, move, and perform physical tasks. Their physical presence and proximity to humans not only make them feel real but also introduce concerns about physical safety, which can influence humans' trust. On the other hand, virtually represented avatars are digital representations of entities that exist within virtual environments, primarily 2D screens [5]. Unlike physical robots, virtual avatars lack physical presence and cannot perform physical tasks in the real world. Therefore, studies involving agents with a visual representation in the virtual world were not excluded based on this criterion. With EC1-2, we intentionally excluded human-robot teaming, as the physical embodiment and interaction have been shown to activate psychological processes that affect trust in unique ways [49], which might introduce confounds.

These criteria were applied to both rounds of paper screening. As shown in Figure 1, during title and abstract screening, we primarily concerned with excluding those that did not involve human participants (IC1) (e.g., computational models, conceptual models), involve physical robots (EC2) or remote-controlled agents (EC2), or were off-topic (EC3). During the full-text screening, we were able to exclude papers after closely examining whether they met the criteria for human-AI teaming (IC2-3), the agent's level of autonomy (IC4), and whether they measured trust (IC5). The two rounds of search and screening resulted in a total inclusion of 57 articles for the final review and analysis of trust influencing factors.

3.3 Data Extraction and Coding

From the final list of articles (N=57), initial coding efforts focused on the factors influencing human participants' trust in the autonomous

[2]The search terms were informed by the authors' extensive experience in HAT research and validated through a quick scan of [100]'s corpus, which identified these terms as commonly referenced across studies. This ensured the search was comprehensive and aligned with established terminology.

[3]Secondary objective refers to studies that measured trust as one of the many dependent variables. But since the study's focus was not trust, it was not mentioned in the title, abstract, or keywords, thereby not identified from the first round of search.

[4]Admittedly, in organizational psychology, trust, trustworthiness, and propensity to trust are interrelated yet distinct constructs [25] We combined these terms to capture as comprehensive a body of literature as possible, as these constructs are not always clearly differentiated and sometimes even used interchangeably in the context of HATs.

[5]Although virtually embodied agents in immersive AR and VR environments are getting mature [137], none of the HAT research we screened involved immersive VR or AR.

Round	Database	Search string	Filters	Results
1	ACM Digital Library	Abstract:(team* AND trust* AND ("human-AI" OR "human-autonomy" OR "human-agent" OR "human-machine" OR "human-automation")) OR Title:(team* AND trust* AND ("human-AI" OR "human-autonomy" OR "human-agent" OR "human-machine" OR "human-automation")) OR Author Keyword:(team* AND trust* AND ("human-AI" OR "human-autonomy" OR "human-agent" OR "human-machine" OR "human-automation"))	Content type: research article + extended abstract + short paper	77
2	ACM Digital Library	AllField:(team*) AND AllField:(trust*) AND AllField:("human-AI" OR "human-autonomy" OR "human-agent" OR "human-machine" OR "human-automation")	Content type: research article + extended abstract + short paper	2012
1	IEEE Xplore	("All Metadata":team*) AND ("All Metadata":trust*) AND ("All Metadata": "human-AI" OR "human-autonomy" OR "human-agent" OR "human-machine" OR "human-automation")	conferences + journals	122
2	IEEE Xplore	("Full Text & Metadata":team*) AND ("Full Text & Metadata":trust*) AND ("Full Text & Metadata": "human-AI" OR "human-autonomy" OR "human-agent" OR "human-machine" OR "human-automation")	conferences + journals	2283
1	Web of Science	TS=(team* AND trust* AND ("human-AI" OR "human-autonomy" OR "human-agent" OR "human-machine" OR "human-automation"))	Document Types: Article + Proceeding paper	314
2	Web of Science	ALL=(team* AND trust* AND ("human-AI" OR "human-autonomy" OR "human-agent" OR "human-machine" OR "human-automation"))	Document Types: Article + Proceeding paper	429

Table 1: Literature Search Strategies and Results. a. Asterisk (*) denotes wildcards that include any number of unknown characters. b. TS in Web of Science searches title, abstract, keyword plus, and author keywords.

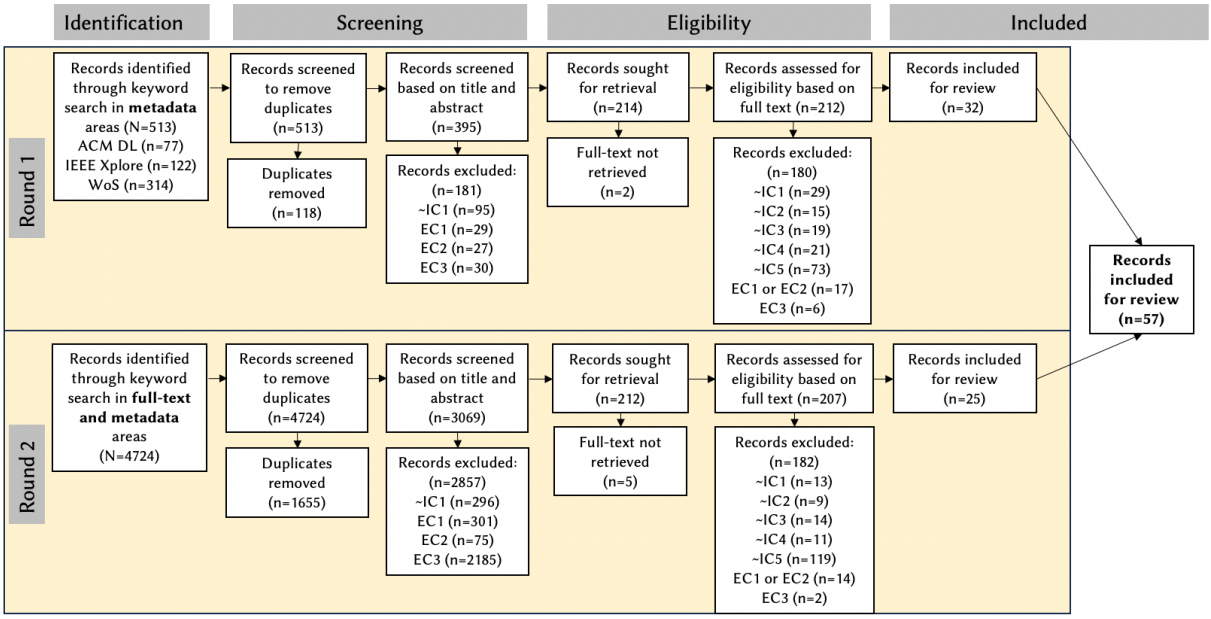


Figure 1: Literature Search Procedures. Note: tilde ("~") means NOT meeting a criterion

teammate, either manifested through subjective evaluation or objective behaviors. We first extracted information on the measurement(s) of trust, categorizing the measures into subjective or objective, documenting the subjective questionnaires, and scales the studies used, adapted, modified, or self-created, categorizing the scales into trust in interpersonal or functional trust. Second, we extracted information regarding the independent variables related to trust and grouped them into agent-related, human-related, team-related, and environment-related factors.

We also extracted information on the definition of trust (if defined), the characteristics of the agent teammate(s), including its control method [120] and its visual representation. For categories of the control method, the agent may be 1) controlled by a computer system autonomously, 2) controlled using the Wizard of Oz technique [29], in which human participants were told that they were working with a computer system while they were actually working with a trained confederate researcher; or 3) using vignettes [3]. We also extracted information on the characteristics of the team, including the size and composition, the nature and type of the team task, the communication between the human and the agent, as well as the environmental characteristics such as the simulation testbed and platform. The year and the country where the research was conducted were documented to provide an overview of the current landscape of HAT research on trust. Six of the authors coded the articles first independently and met up frequently to discuss and resolve disagreements.

3.4 Effect Size Calculation for Meta-analysis of Factors Influencing Trust

A meta-analytical approach [123] was employed to determine the pattern of the findings in the empirical HAT research on factors of trust. To do this, we first extracted all the relevant statistics required to compute the standardized effect sizes (Cohen's d) and the standard errors of the effect sizes. These included the means, standard deviations, and sample sizes for both the control group and the treatment group [14] for pairwise experimental designs. Note that we did not use the effect sizes reported by these studies using partial-eta squared, as partial eta-squared is not standardized and reflects the percentage of variance in the effect rather than the standardized difference between two means. In cases where the required statistics were not reported in the text, table, or graphs, we converted available statistics into effect sizes. For instance, we calculated the standard deviation from the standard error of the mean or confidence intervals of the mean [11]. Despite the effort, we only obtained 25 pairwise effect sizes from 18 studies due to many of the reviewed studies not reporting the statistics completely, especially in cases of insignificant results. The obtained statistics were then entered into SPSS (version 28) to generate the meta-analysis and plots for interpretation of the snapshot of the factors influencing trust in HATs. We used the ranges established by Cohen [23] to interpret the effect sizes (small: d=0.2; medium: d=0.5; large: d=0.8).

4 Overview of the Corpus

4.1 Growing and Global Presence of HAT Research Focusing on Trust and Research Settings

We outline the temporal (Figure 2) and geographical (Figure 3) characteristics of the resulting corpus of 57 articles. The increase in the number of papers over the years suggests a growing interest in research on trust in human-AI teams. Most studies (40/57) were conducted in the USA, followed by Australia (4/57), Germany (4/57), and The Netherlands (3/57).

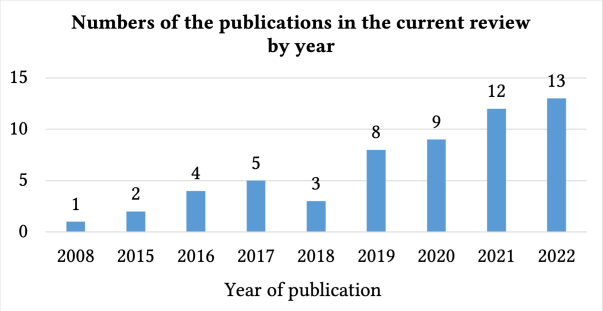


Figure 2: Number of Studies Studying Trust in HATs by Year

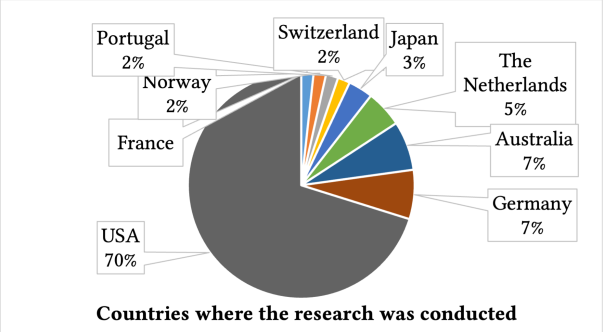


Figure 3: Countries in which the Reviewed Studies Took Place.

4.2 Characteristics of AI Teammates

4.2.1 Agent Visual Representation. As shown in Figure 4, around half of the agents did not have a visual representation. For those that did, we categorized them along two dimensions: humanoid vs. robotic/iconic and static vs. animated. Fourteen had the agent visually represented using a robotic/iconic static image (e.g., an icon on a mission map, see Figure 10a for an example). Two used a robotic/iconic animated avatar (see Figure 10b). In four articles, the agent was represented by a humanoid animated avatar, capable of navigating the virtual environment the same as the human player's avatar (see Figure 10d). One used a humanoid static image to represent the agent (Figure 10c). Additionally, five examined visual representation as an independent variable, and the rest did not specify whether or how the agent was visually represented.

4.2.2 How Was the Interaction between Humans and the AI Teammate Realized. The interaction between humans and the AI teammate can influence humans' trust in the AI. However, not all of the reviewed studies allowed participants to interact with it. We examine whether the HAT research involved actual interaction and how it was realized to contextualize the findings on trust. As shown in Figure 5, in more than half of the research studies (36/57), the AI agent was autonomously controlled by a computer system, engaging in simple, often pre-programmed interactions with humans, such as delivering messages or performing actions, without requiring any human input. Nine studies used the Wizard of Oz (WoZ) paradigm [63, 108], where a confederate researcher played

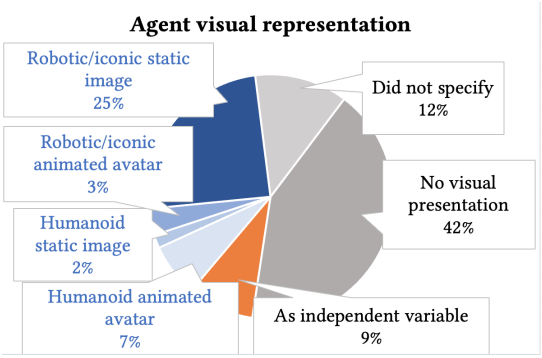


Figure 4: Agent Visual Representation

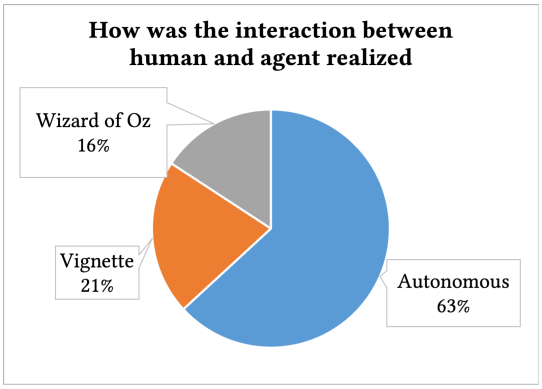


Figure 5: How Was the Interaction between Humans and the Autonomous Teammate Realized

the role of an agent. Participants were made to believe it was a real agent instead of a human controlling it. Twelve out of 57 studies did not allow for the same level of real-time interaction between agent(s) and human participants, using vignette scenarios [3] that manipulated certain variables of interest to study participants' trust in AI teammate(s).

4.3 Characteristics of Team

4.3.1 Team Size and Composition. A majority (44/57) of the articles included in this review explored trust in HAT with teams composed of one human and one agent (see Figure 6). Around twenty percent of the articles (12/57) used a three-member team. Among these twelve studies, nine have teams of two humans and one agent, while two studied teams of one human and two agents; one study manipulated team composition as an independent variable. It's worth noting that more than half (8/12) of the studies that used 3-member HAT were conducted by the same group of researchers, and five of the eight conducted using the same simulation testbed. As found by one of the reviewed articles [116], team composition can have significant impact on humans' trust in their teammate, such that they trusted an agent teammate more when the other teammate was a human, than when the other teammate was also an agent. Additionally, team size and composition can also influence trust perceptions through communication affordances. For instance, multi-member teams deal with more communication challenges such as turn-taking [112] and interruption [126], which can increase chances of miscommunication [148], reduce task efficiency, and increase members' cognitive load [109], potentially posing threats to trust. These communication challenges cannot be accounted for by studying two-member teams.

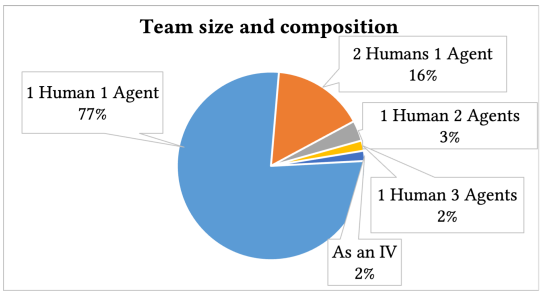


Figure 6: Team Size and Composition

4.3.2 Team Tasks and Contexts. The task environments were categorized into military context, emergency response context, AI-assisted decision-making in various work and life settings, such as image recognition for public safety, healthcare, transportation, etc., and non-military game contexts. More than half (33/57) of the studies were conducted using military task environments (see Figure 7). Consistent with recent meta-review of human-AI teaming [100], military task contexts represent an overwhelming proportion of HAT research.

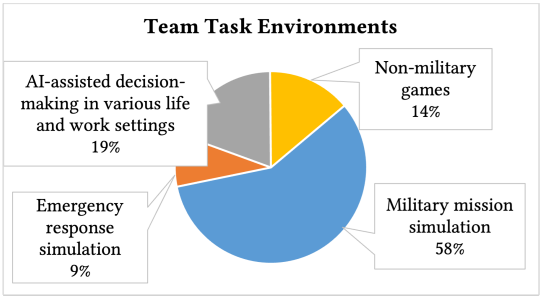


Figure 7: Team Task Environments and Contexts

5 Conceptualizations and Operationalizations of Trust (RQ1)

Before discussing the factors that have been studied in relation to trust in HAT research, it is essential to first understand how trust is conceptualized and operationalized, because how trust is defined and measured (e.g., trusting the AI as a teammate versus a tool) to an extent determines the factors of interest to the researchers.

5.1 Conceptualizations of Trust

Our review reveals that more than half (36/57) of the articles reviewed did not provide or adopt a definition of trust. Among those

that did provide a definition of trust, Lee and See [71]'s definition is most widely adopted, defining trust as

> "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (p.51).

This definition, while formed through an examination of trust in automation and interpersonal trust, is rooted in a perspective where humans serve as operators of machines, making a person's trust predominantly contingent on the apparent qualities of the technology. Several articles have shifted focus away from the "functional" aspect of autonomous agents and instead adopted trust definitions from the field of organizational management such as McAllister [82]:

> "the extent to which a person is confident in, and willing to act on the basis of, the words, actions, and decisions of another" (p.25)

, and Mayer et al. [80]:

> "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 other party" (p.712).

Yet another article [149] provided a definition of trust of their own: "trust is a special case of reliance where one party is relying specifically on the goodwill of the other party". A visual representation of the definitions of trust adopted in the reviewed articles can be seen in Figure 8.

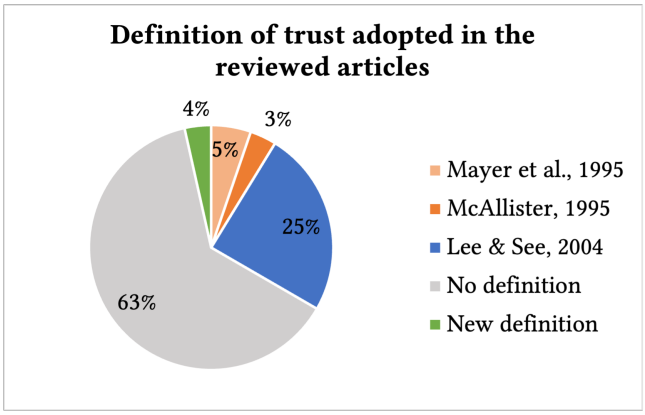


Figure 8: Definition of Trust Adopted in Reviewed Articles

5.2 Operationalizations of Trust - Measuring Trust in and Trustworthiness of Agent Teammates

We identified and categorized all the measurements of trust used in HAT research, first into **subjective** and **objective** measures, with subjective measures further categorized into **interpersonal** trust measures (see Table 2), including the most dominant and seminal interpersonal trust measurement developed by Mayer and Gavin [81]; and **functional** trust measures (see Table 3), including the most dominant measures such as Trust in Automated Systems Scale [59], Trust in Automation Scale [91]. Objective measures of trust (see Table 4) primarily take the form of counting the number or percentage of acceptance of the agent's recommendations or advice (e.g., [36, 48, 67, 68, 122]), and the number of times humans entrusted the agent to do certain tasks [57, 106]. There are also a few studies that used more than one measurement for assessing trust as a dependent variable.

In this section, we first discuss 1) how the autonomous agent is framed and positioned differently in the interpersonal and functional trust measures. We then discuss 2) how the studies that employed both interpersonal trust and functional trust measures yielded discrepant results, indicating the inappropriateness of collating interpersonal trust and functional trust measures while also emphasizing that the use of multiple trust measures for triangulation purposes is desirable. Subsequently, we discuss 3) the relationship between factors of interest in the reviewed studies and their choice of trust measurement. Lastly, we show the 4) temporal aspects of the trust measurements, highlighting the need for dynamic measurement of trust that captures trust fluctuation in real-time.

5.2.1 AI Agent Framing in the Measurements. A noticeable difference in interpersonal and functional trust measurements is the framing and positioning of the trusted entity (i.e., AI agent). The items in interpersonal trust measurements framed the agent as a teammate, either by explicitly calling it a teammate [116, 117], referring to it by a human name [48] or its role on the team [87]; whereas the items in functional trust measurements referred to the agent as the "system" (e.g., [19, 125]) or the name of the "tool" (e.g., [32, 115, 146, 149]).

These different framings also manifest in the characteristics of the AI agent, its visual presentation, communication capabilities and modalities, and the team characteristics. For instance, we identified that it is through two means that an agent has been framed as a teammate and thus an interpersonal (rather than functional) trust measurement was more likely to be adopted: one is agent visual presentation as a humanoid avatar, the other is human-like communication. The combination of both creates an even stronger teammate framing. We mapped the characteristics of the agent's appearance and communication affordances in the reviewed articles onto a continuum of human-likeness, and color-coded the type of trust measurement they used (Figure 9) to provide a straightforward image.

Specifically, seven out of twenty that measured trust using an **interpersonal** trust measure had the agent both embodied and able to communicate in human natural language; ten had the agent equipped with two-way interactive communication capabilities. In some studies that used a video game platform (e.g., [58, 119]), the agent teammate looked like another human player. Hanna and Richards [48] made the virtual agent a male human avatar named Charlie capable of both verbal and nonverbal communication. Tolmeijer et al. [131] had the agent embodied in a humanoid avatar that communicated recommendations in text using first-person tone of voice. Even though Kox et al. [64] used robotic visual representation in a 3D virtual environment, they had the robot communicate using a male human voice speaking US English. In another study [65], the agent teammate was embodied as a virtual

Interpersonal Trust Measures		
Source	**Example Scale Items**	**Used in Reviewed Articles**
Trust in teammates [81][6]	"I would be willing to let (role or teammate name) have complete control over my task in the team." "If the (role or teammate name) asked me for something, I respond without thinking about whether it might be held against me."	[10, 16, 24, 31, 57, 60, 87, 102, 105, 138, 141]
Trust in teammates [56][7]	"I trust my teammate and would like to continue to participate in other teamwork with my teammate." "My teammate is fair in performing team tasks." "My teammate works responsibly for accomplishing the team task."	[45–47]
Trust in teammates [75]	"I felt confident in the AI teammate I just worked with." "I felt like my AI teammate had harmful motives in the task." "I felt fearful, paranoid, and or skeptical of my AI teammate during the task."	[116, 119]
Trust in teammates [1]	"Most people on this team are basically honest and can be trusted."	[132]
Multi-Dimensional Measure of Trust (MDMT) [134]	16 adjectives on 4 subscales: reliable, predictable, someone you can count on, consistent, capable, skilled, competent, meticulous, sincere, genuine, candid, authentic, respectable, principled, has integrity	[131]
Trust in teammates [27]	"We have complete confidence in each other's ability to perform tasks." "Some members hold back relevant information in this team." "In this team most members tend to keep each other's work under surveillance."	[119]
Authors cited [84] which appears to be a multi-construct model of trust instead of a validated measurement of trust	"My buddy has a lot of knowledge on navigating through this environment." "My buddy puts my interests first." "My buddy is honest."	[64, 65]
Single-item, no source	"how much participants trusted co-players" "Over time, my trust in Charlie's selections increased." "I trusted Teammate A."	[15, 48, 58, 107, 124]

[6] Many cited [80] to be the source of their trust measurement. However, there is no validated scale proposed in [80]. A close examination reveals that the scale comes from [81]

[7] Authors cited [4] but the measurement originated from [56], which was also originated from an earlier version of Mayer & Gavin (2005) [81]'s trust in teammates scale.

Table 2: Interpersonal Trust Measures

drone but communicated again in natural human language using audio messages, but this time with a computerized voice. Others (e.g., [105, 138]) also used a combination of virtual robotic avatar and textual communication using first-person tone. For the rest that did not have agent embodied in any visual form, a majority made the agent teammate capable of communication in human natural language either via audio [16, 102] or text [10, 24, 31, 61, 87, 119, 132]. It's worth noting that unlike fields of research that investigate chatbots and digital voice assistants that use predominantly female voice [13, 17], almost all the HAT research included for this review that employed audio messages used male voice (e.g., [16, 64, 65, 102]). This might suggest that task environments such as military and emergency response contexts that host most HAT research are stereotypically male-dominant. Women and gender non-binary individuals need to be equally represented in future HAT research.

HAT research that adopted **functional** trust measurements primarily evaluated humans' trust in 1) the entire system interface [5, 9, 19, 22, 74, 111, 125, 127, 130, 141], where the autonomous

Functional Trust Measures		
Source	**Example Scale Items**	**Used in Reviewed Articles**
Trust in automated systems [59]	"The system is deceptive." "I am suspicious of the system's intent, action, or outputs."	[19, 22, 57, 90, 111, 121, 127, 130, 141, 143, 144, 146]
Trust in automation [91]	"I believe the (automation name) is a competent performer." "I have confidence in the advice given by the (automation name)."	[9, 12, 44, 73, 74, 79, 146]
Trust in automation [70]	"How much did you trust the (tool name) to (tool function)?"	[5, 30]
Trust in machines [93]	"To what extent does the automated decision aid perform its function properly?"	[146, 149]
Human-computer trust (HCT) [76]	"Xxx performs reliably" "It is easy to follow what xxx does."	[102]
Computer credibility and trustworthiness [40]	trustworthy, good, truthful, well-intentioned, unbiased, honest	[67, 68]
Operator trust in automation [128]	"I know what the automatic system will do in the next 3 minutes."	[125]
Operator trust in automation [92]	competence, predictability, dependability, responsibility, reliability, and faith	[106]
Trust in information systems [129]	"I completely trust the digital agent." "I rely heavily on the digital agent." "I feel comfortable relying on the digital agent."	[35]
Trust in technology [85]	"I trusted the (technology name)" "I could rely on the (technology name)" "I would advise a friend to take advice from the (technology name) if they played the game".	[115]
Metrics for explainable AI [52]	eight items and measured confidence in and predictability, reliability, safety, efficiency, wariness, performance, and likeability of RescueBot	[135]
Trust in systems [140]	"I don't trust the detector at all"	[32, 147]

Table 3: Functional Trust Measures

Behavioral Trust Measures	
How trust was measured behaviorally	**Used in Reviewed Articles**
IRA (Inappropriate Recommendations Accepted) and IRC (Inappropriate Recommendations Correctly adjusted)	[36]
the number of images participants allocated to the automation before each round	[57]
the number of times an operator took manual pictures as an indicator of low trust in UAV teammates.	[106]
Requested and adopted advice	[68]
the number of offers ignored, suggestions requested but declined, suggestions requested and adopted	[67]
the ratio of acceptance by the human of the IVA's request	[48]

Table 4: Behavioral Measures of Trust

teammate had no visual presentation, and communication took the form of human-interface interaction, despite calling it human-agent or human-autonomy teaming; 2) a component of the interface such as an alarm aid [32, 44, 73, 147] that provided objective information (can be considered one-way communication from system interface to the human user) without using first-person tone of voice, and 3) virtual unmanned vehicles or squad members that were not embodied but represented as an icon marking their location on a

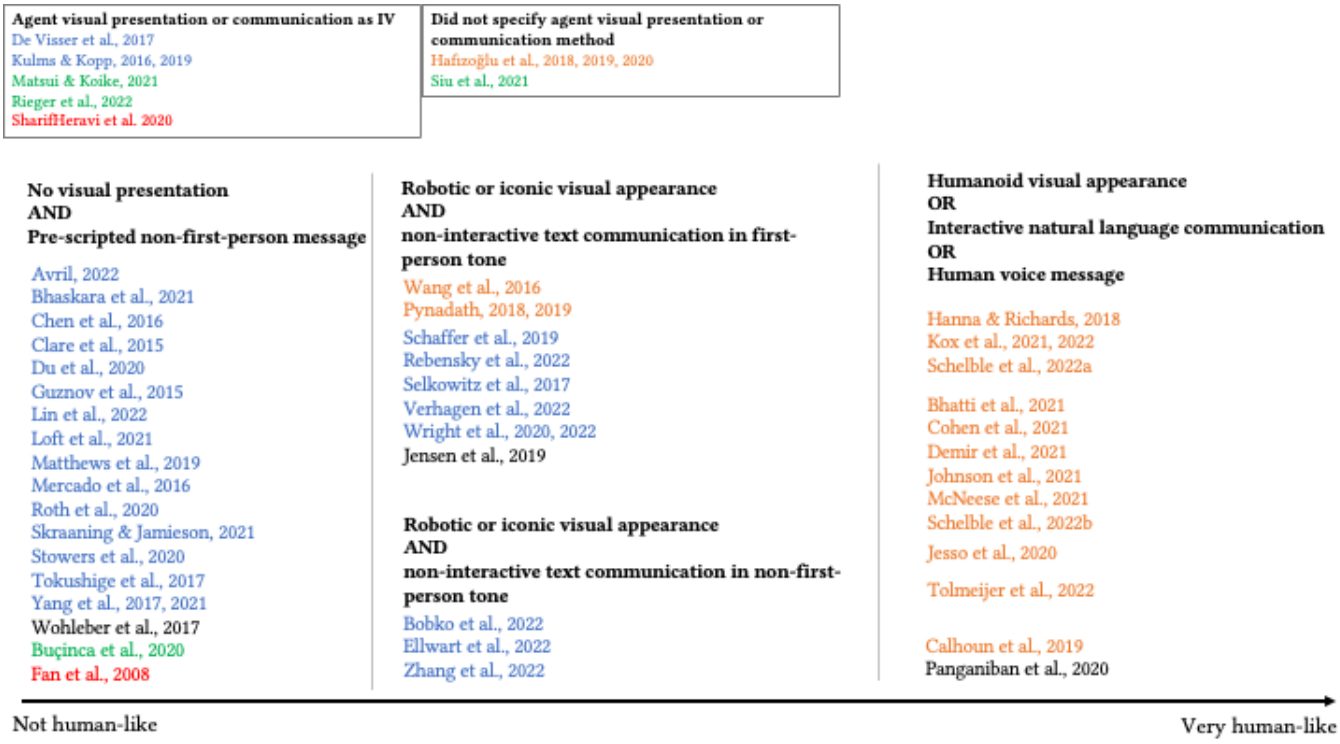


Figure 9: Agent visual presentations and communication affordances of the reviewed articles in relation to human-likeness of the agent and the trust measurement used. Orange are those that used interpersonal trust measurement, Blue are those that used a functional trust measurement, Black are those that used both interpersonal and functional trust measurements, Red are those that only used a behavioral trust measurement, Green are those that used single-item trust measure

map and/or showing their point of view [57, 90, 106, 121, 142, 144], where communication was one-way and visual. A few articles that measured functional trust involved an embodied agent, among which Kulms and Kopp [67, 68] and De Visser and colleagues [30] manipulated the visual representation of agent as an independent variable, presenting the agent avatar using a picture of a human, a cartoon human, or a computer. Others used a simple cartoon sketch of a robot (e.g., [12, 115, 135, 149]). In terms of communication, none of the agents in these articles afforded interactive natural language communication. In [68] and [30] pre-recorded video clips were shown wherein the human and cartoon human voiced utterances like "Let me think", whereas the computer agent "communicated" through blinking light or "symbolic icon". Still others (e.g., [35, 79]) did not provide information about the visual representation or communication capabilities of the agent.

Our review highlights that the researchers' framing of an agent (either intended or unintended) often manifests in how it's visually represented and its verbal communication capability and modality, which correlates with the selection of interpersonal versus functional measurement of trust. Studies that have the agent visually represented in a humanoid avatar and/or enable the agent to communicate verbally and interactively using natural language tend to use interpersonal trust measures, whereas those that don't involve agent visual representation or interactive communication tend to adopt function trust measures.

5.2.2 Results Discrepancy Using Multiple Measurements. Another interesting finding from the review is that, among the three studies that used a combination of interpersonal trust and functional trust measures, two yielded discrepant results. For instance, Wohleber et al. [141] found that using an interpersonal trust measure, no significant effect of agent transparency was found, but only a significant main effect of communication framing (critical versus complimentary communication) such that participants were more trusting when the agent was critical. However, using the Trust in Automated Systems measure [59], transparency was found to have a significant effect such that participants trusted the high transparency agent more; but communication framing was not significant. Panganiban et al. [102] reported no significant effect of team type (interdependence among team members) or type of agent transparency (neutral versus benevolent) using an interpersonal trust measure [80]; but found a significant effect of team type using a trust in technology measure [76], such that individuals dependent on the agent reported higher trust than those independent of the agent. While Jensen et al. [57] claimed to have used both functional and interpersonal measures of trust, they unfortunately only reported results for the functional measure. Based on the two studies, it seems that interpersonal trust and functional trust measures tend

to yield different results, which warrants attention to purposefully choosing the measurement that aligns with the conceptualization and positioning of the trusted agent entity.

5.2.3 Independent Variable of Interest and Its Impact on Trust Measurement Selection. The meta-review suggested that studies that used functional trust measures tended to investigate agent-related performance-based factors (see Section 6 for our categorization of the factors) such as the agent's reliability and transparency. For instance, nine out of 11 studies that examined reliability as a factor, and 14 out of 17 studies that examined transparency as a factor adopted a functional measure of trust (see Tables 5, 6). Studies that used interpersonal trust measures tended to look at the agent-related behavior-based factors, such as communication style, apology, and blame that emphasized the interpersonal aspects. Additionally, most of the studies that examined a team-related factor (e.g., role, interdependence, team composition) adopted an interpersonal measure of trust. The rest of the factors have a relatively equal representation of studies adopting interpersonal and functional trust measures.

5.2.4 Temporal Aspects of Trust Measurement. As trust is known to fluctuate over time and over multiple interactions with an agent through direct and indirect interactions [54], it is best to measure trust prior to interactions with an agent, during, and after each exposure to properly quantify the amount of trust held within the agent and the team. Marsh and Dibben [77] identified three layers of trust that can be viewed as representing trust fluctuation over time. These layers are dispositional, situational, and learned trust, which were applied to trust in human and agent interaction by Hoff and Bashir [51]. Dispositional trust refers to a human's overall long-term based tendency to trust an agent or HAT independent of context or system arising from both biological and environmental influences [51]. As an individual's dispositional trust in agents and HATs in general can alter or form their trust in future agents or HATs, it is recommended to measure this layer of trust before any interaction with an agent(s) or HAT takes place [113]. Situational trust is influenced by the environment and context-dependent variations in an individual's mental state that occur during interaction [51], and best measured behaviorally during the interactions of a HAT. Learned trust is formed by the evaluations of all the past experiences or current interaction an individual has with a specific agent or HAT [51], and best measured using questionnaires after a HAT finishes a sequence of interactions that may influence a human's evaluation of trust during said sequence of interaction. Overall, this dynamic and temporal aspect of trust seems to be overlooked throughout the human- teaming literature where we discovered that out of the 57 studies included in this review, five studies accounted for dispositional trust through trust propensity scales (e.g., [102, 116, 119, 138, 146]), six measured a form of situational trust during an interaction (e.g., [36, 48, 67, 68, 106, 122]), 52 studies recorded learned trust once after all the interactions among that HAT took place, and six studies recorded a teammate's learned trust after each interaction (i.e., a mission or trial) took place (e.g., [24, 57, 61, 64, 65, 119]). To properly measure trust, researchers must account for the layers of trust that shape how trust fluctuates and develops over time. Future studies should account for dispositional, situational, and learned trust by measuring trust propensity prior to the start of a HAT's interaction, measuring trust behaviorally during interactions, and by recording an individual's trust evaluation of the agent(s) and HAT prior, during, and after each interaction.

The review of trust measurements in HAT research highlights several opportunities for methodological improvement. First, there exists the need for conceptual-operational alignment. Very few studies included in our review explicitly defined or adopted a definition of trust, which obscures a clear perspective and positioning of the agent as a teammate versus a tool, which in turn affects the operationalization of trust and its ability to accurately capture humans' attitude towards the agent. Second, the discrepant results from using multiple measures highlight the need for triangulation techniques (e.g., triangulating self-reports, derived latent trust factors by factor analyzing measures, behavioral measures, and even qualitative reflections) that increase the content validity of the findings. Third, the dominance of post-hoc subjective measures of trust also highlights the need for trust to be viewed dynamically where an individual's dispositional, situational, and learned trust in an agent teammate must be measured and accounted for [51, 77]. The dynamic nature of trust, specifically situational and learned trust, requires real-time measurement of trust to allow for timely adjustment from the agent and/or trust repair.

6 Factors Influencing Trust Outcomes in HATs (RQ2)

6.1 Overview and Categorization of Factors

Grounded in the review, we identified the independent variables hypothesized to have an impact on trust in HAT research and categorized them into agent-related, human-related, team-related, and environment-related factors (see Tables 5 6).

In the current review of the existing HAT literature, the agent-related factors influencing trust in HATs have been the focus. We classified these factors into three categories based on different aspects of the agent: agent attribute, agent performance, and agent behavior.

Attribute-based factors of trust primarily manipulated anthropomorphic characteristics of the agent, including the humanness of voice [16], agent embodiment [67, 68, 78, 104, 105], framing of the trusted target (agent vs. human vs. system) [58, 107, 131]. Performance-based factors of trust are those that are intrinsically related to the agent's designed function, its competence in performing the function, and features that facilitate humans' understanding and evaluation of the agent performance of the function. These include the level of autonomy (LOA), transparency, reliability, explanation, etc. Behavior-based factors of trust are those that are not related to the agent's designed function and emphasize the humanness of agent behavior. These include the agent's ethical behavior [119], trust repair strategies such as apology (e.g., [64]) or blame [57], and communication style, such as being critical or complimentary [141], directive vs. nondirective [144]. In the data extraction phase, the authors noticed inconsistent use of terms with respect to the factors. For instance, Panganiban and colleagues [102] uses the term transparency to denote what is communication style (neutral vs. benevolence), whereas Verhagen and colleagues [135] uses communication style to denote what is a combination of

Factor category	Factor influencing trust	Factor examined in reviewed articles	Trust measurement category
Agent-related			
Attributed-based	Anthropomorphism	[122]	behavioral
		[58, 78, 107]	single-item
		[30]	functional
		[67, 68]	behavioral+functional
		[16, 131]	interpersonal
	Agent appearance (robot-like vs. dog-like)	[104]	behavioral
Behavior-based	Apology	[64, 65, 119]	interpersonal
		[105]	interpersonal + behavioral
	Communication style	[102, 141]	interpersonal + functional
		[144]	functional
	Agent ethical behavior	[119]	interpersonal
	Blame	[57]	behavioral + interpersonal + functional
	Collaboration strategy	[132]	interpersonal
Performance-based	Transparency	[9, 12, 19, 73, 74, 79, 90, 111, 121, 125, 127, 135, 143]	functional
		[16, 132]	interpersonal
		[141]	interpersonal + functional
		[122]	behavioral
	Reliability	[5, 12, 30, 32, 44, 115, 130, 143, 146, 147]	functional
		[36]	behavioral
		[138]	interpersonal
		[57]	behavioral + interpersonal + functional
		[67, 68]	behavioral+functional
		[107]	single-item
	Level of autonomy (LOA)	[106]	behavioral + functional
		[115]	functional
		[131]	interpersonal
	Uncertainty communication	[64, 65]	interpersonal
		[147]	functional
		[105]	interpersonal + behavioral
	Explanation	[115] [149]	functional
		[138]	interpersonal
		[15]	single-item

Table 5: Agent-Related Factors Examined in the Reviewed Articles to Influence Trust in HATs

transparency and explanation type. We corrected the terms after reading thoroughly into the detailed descriptions by three of the authors of the manipulations to reach an agreement as to what the independent variable examined.

6.2 Meta-Analysis of Factors

Due to incomplete statistical reporting, we were only able to generate a meta-analysis for 8 agent-related factors, including 18 effect sizes from 14 articles, as shown in Figure 11 (See end of manuscript).

6.2.1 Agent-related Factors. Agent-related factors have an overall large effect on trust (Cohen d = 1.12). Among them, reliability and transparency have an aggregation of multiple articles examining their effect on trust; each of the rest of the factors has only one article represented.

Reliability, operationalized as high versus low suggested by accuracy percentages, has an overall large effect on trust (d=1.13). This is further guaranteed by looking at the effect sizes of the individual studies where all the lines are to the right of and do not cross over

Factor influencing trust	Factor examined in reviewed articles	Trust measurement category
Human-related		
Gaming experience	[22]	functional
Training	[60]	interpersonal
Commitment	[48]	interpersonal
SDSC (sociodigital self-comparisons) in favor of the nonhuman agent	[35]	functional
Drug condition (oxytocin or placebo)	[30]	functional
Team-related		
Team member interdependence	[102]	interpersonal + functional
	[135]	functional
Team composition	[116]	interpersonal
Team performance	[87]	interpersonal
Role on the team	[10, 24, 31]	interpersonal
Staffing solution (single- vs. multi-unit)	[125]	functional
Human-AI expertise complementarity	[149]	functional
Environment-related		
Time	[24, 60, 64, 65, 119]	interpersonal
	[122]	behavioral
	[149]	functional
Trust priming	[22]	functional
	[45–47]	interpersonal
Danger level	[73]	functional
Risk level	[74]	functional
Scenario difficulty	[125]	functional

Table 6: Factors Related to Human, Team and Environment That Were Examined in the Reviewed Articles to Influence Trust in HATs

the null-effect line, suggesting an invariably significant effect of reliability. All the studies with sufficient statistical data for effect size computation examined agents without any visual representation. Additionally, Fan et al. [36] reported smaller effect sizes for the impact of reliability on trust compared to other studies. However, due to the small sample size, it is unclear whether the observed variance was driven by the use of behavioral versus subjective functional measures of trust, differences in team composition, or a combination of both.

For transparency, the effect is only small to medium (d=.31), with more than half of the studies not yielding significant effects. Only two studies ([90, 121]) demonstrated a significant effect of transparency on trust. It's worth noting that almost all studies examining the role of transparency operationalized it using Chen et al. [20]'s "Situation Awareness-based Transparency (SAT) framework," where Level 1 provides information about the agent's current state, goals, intentions, and plan of action; Level 2 provides information about the agent's reasoning process behind the action; and Level 3 provides information regarding projected consequences and uncertainty, including the likelihood of failure. For pairwise comparisons, the treatment condition always includes more levels of information than the control condition (i.e., Levels 1+2 vs. Level 1 or Levels 1+2+3 vs. Levels 1+2). The fact that the two studies that yielded significant effects compared the treatment condition (i.e., Levels 1+2+3) to the control condition (i.e., Level 1) that are two levels apart rather than conditions that are one level apart (as did the rest of the studies) might explain the significance. All studies investigating the effect of transparency on trust employed a military task within a one-human, one-agent team composition. The variance in effect sizes does not appear to be explained by either the agent's visual representation or the type of trust measurement used.

Other agent-related factors such as agent's ethical behavior (ethical vs. unethical; d=.83), apology (presence vs. absence; d=4.09), blame (internal vs. pseudo-external; d=.98), uncertainty communication (presence vs. absence; d=2.41) all have significant and large positive effects on trust. Collaboration strategy (cooperative vs. individual; d=.68) and explanation (presence vs. absence; d=.59) have medium effects on trust.

6.2.2 Human-related Factors. For human-related factors, we were only able to obtain 1 effect size for 1 factor: participants' gaming experience, which is shown to have a medium effect (d=.56) on trust.

6.2.3 Environment-related Factors. We identified one factor related to environment: trust priming. Specifically, studies primed participants by exposing them to a trustworthy agent versus untrustworthy one prior to the main interaction. This factors is shown to have small to medium effect on trust (d=.39). Notably, Clare et al. [22] reported larger effect size than the other studies. However, due to the limited number of studies, no conclusions can be drawn about

whether the observed variance was driven by the use of functional versus interpersonal measures of trust, differences in agent visual representation or the task context, or a combination of these factors.

6.2.4 Team-related Factors. For team-related factors, only one article included the statistics required for computing Cohen's d. Team composition has a large effect size on trust (d=.98), such that humans trusted the agent teammate more in teams with two humans and one agent teammate, than in those with two agents and one human teammate [116].

There were several challenges in conducting the meta-analysis. First, the information in some of our reviewed studies was incomplete and imprecise, making it hard and in some instances impossible to obtain the statistics required. Second, studies that did not find statistical significance tended not to report the statistics required to conduct a meta-analysis. Additionally, non-significant results are often not published, falling prey to the "file drawer problem"[110]. These can result in biased findings.

7 Discussion

This systematic review and meta-analysis has demonstrated that existing understanding of trust in HATs are rapidly evolving. This understanding is often advanced from two key perspectives, namely, how AI teammates can be trusted as a functional system and as an interpersonal teammate. However, not all research has adopted these or any consistent framing to study trust in HATs. This thus highlights the need for conceptual clarity and consistency as well as conceptual-operational alignment. In this section, we provide some guidelines for HCI and CSCW communities to establish a common foundation for clarifying and aligning conceptual definitions with operational measures to ensure methodological rigor and comparability across studies. In doing so, we aim to facilitate a more coherent and cumulative research trajectory within the HCI and CSCW communities to help standardize research practices, promote knowledge synthesis, and ultimately contribute to the development of effective strategies for cultivating trust in human-AI collaboration. Additionally, through the systematic literature review and meta-analysis, we also identified a number of gaps in the existing trust research in HATs and provided directions for future research efforts to address.

7.1 Need for Conceptual Clarity and Consistency

A number of the articles we reviewed adopted one of two popular definitions of trust, namely, trusting the automation [71] and interpersonal trust [80]. However, a great number of articles also chose not to adopt any standard definition of trust. These inconsistencies demonstrate that as the perspective of considering agents as teammates gains more popularity in HAT research [2, 94, 116, 150], challenges will arise in accurately defining what it means to trust an AI teammate. Our systematic review suggests that the emerging field of HAT research on trust lies at the intersection of positioning trust as both functional and interpersonal, drawing heavily on a trust in automation framework [71] and a human team trust framework [80] respectively. However, an agent teammate is not solely functional nor solely interpersonal, as the agent must meet the functional and social requirements. Definitions of trust may be better served not by using one of two polarizing perspectives but rather using a teaming continuum that describes agent teammates based on their functional and interpersonal capabilities. An agent teammate may be more appropriately positioned somewhere on this continuum depending on its designed purpose, its relation to the human, and the teaming and environmental context requiring differing functional and interpersonal competency. In other words, traditional definitions of both interpersonal trust and functional trust may not be readily applicable to all HAT research. Rather, it is beneficial for empirical studies to define trust along this continuum in a way that best serves their purpose in their specific context and that suits their framing of the agent in the team. Scoping these definitions on a common continuum will also ensure a degree of consistency and relatability across research efforts.

Leveraging this continuum will require empirical work to first identify the dimensions of human trust in the AI teammate specific to the HATs being researched, which calls for a bottom-up approach that leverages grounded theory [42], qualitative, or certain quantitative methodologies. For example, Hauptman et al. [50] work represents an excellent example that uncovers the qualities contributing to human trust in an AI colleague through interviews with professionals who actually work with AI on a daily basis. Within the context of this work, these qualities would serve as critical points of interest in both defining and measuring trust in future empirical research that shares a similar context. Alternatively, researchers could elicit lay people's perceptions and expectations of their (imagined) AI teammate or partner not only through a grounded theory approach but also by putting them into experimental or pseudo-experimental scenarios, as did Musick et al. [95] and Zhang et al. [151]. While this methodology would likely be less contextually specific due to not using an inductive methodology, replicating these approaches would be especially helpful in crafting quantitative measurements for trust in a specific context. Lastly, researchers could uncover the trust-related factors represented in HAT using quantitative approaches like exploratory factor analyses. Scalia [113] is an example of this approach where HAT and all-human teams were studied in a military style experimental context. Team member responses to trust surveys were factor analyzed to determine the underlying factors in HATs and contrasted those to the factors found in all-human teams.

In sum, the variance in defining trust in most of the reviewed HAT research highlights the need for conceptual clarity, without which empirical progress could be hindered. However, the dynamic and varying nature of HATs means that trust cannot receive a singular and universal definition, as doing so will likely prevent trust from being accurately characterized. As such, this discussion presents a potential scoping mechanism that allows researchers to identify trust on a functional-interpersonal continuum inductively formulated by this review. The use of this continuum will provide a level of construct consistency across research efforts while also ensuring that trust within individual efforts is accurately defined. It is important to note that while the functional-interpersonal continuum of trust may loosely resemble how the neighboring field of HRI [18, 21] conceptualizes robot attributes along the competence and warmth dimensions, which originates from the Stereotype Content Model [38] that theorizes how people form impressions and stereotypes of others, it differs from and extends beyond this

dichotomy in significant ways. First, the competence-warmth dichotomy views trustworthiness as a construct closely tied to (and a sub-construct of) warmth and largely distinctive from competence [37, 38]. In contrast, our functional-interpersonal continuum conceptualizes trust as a multidimensional construct that intersects both functional and interpersonal dimensions, with the relative importance of each dimension varying depending on the context. Second, perceptions of warmth and competence are established measurements [38], using a fixed set of adjectives to describe an individual or a robot, regardless of how the agent is framed or positioned through research design or instrument wording. In contrast, our functional-interpersonal continuum is not a measurement tool but a conceptual framework intended to guide future research in deliberately conceptualizing and operationalizing trusted agents within HAT contexts. .

Guidelines for future research:

- Provide a clear definition of trust that aligns with the conceptualization and framing of the agent in the research.
- Consider both functional and interpersonal capabilities when conceptualizing and framing the agent, while understanding they may not carry equal importance for each HAT context.
- Leverage grounded theory and qualitative approaches for identifying the nuances and dimensions of trust unique to HAT contexts.

7.2 Ensuring Conceptual-Operational Alignment in the Definition and Measurement

While not all reviewed articles provided a clear conceptual definition of trust in HATs (36/57), those that did, 21 to be exact, demonstrate acceptable conceptual-operational alignment with respect to the measurement for trust. For instance, studies adopting a more functional definition [71] also used trust-in-automation measures; and those adopting interpersonal or organizational trust definitions [80, 82] employed corresponding interpersonal trust measures, with a few exceptions ([60, 149]). While conceptual definitions may exist on a continuum, key constructs and their relationship within the conceptual definition should align with those in an operational definition, and, in turn, should be captured in measurement. In survey-based measurement, the mere terminology referencing the AI system has been shown to affect lay people's perceptions and evaluations of the system [69]. Therefore, researchers should take extra care in ensuring that how they refer to the AI agent in their measurement (e.g., "my buddy," "teammate," "Charlie" versus "the system," "the digital aid," etc.) aligns with their overall framing and positioning of the continuum detailed above.

The call for conceptual-operational alignment also underscores the need for measurement consistency, including the use of validated multi-dimensional measurements of trust instead of single-item measures, as trust is far from a unidimensional, narrowly scoped concept [41], and the location of a HAT on the trust continuum will need to be determined by multiple factors of consideration. As indicated in our review, studies adopting functional trust measures tended to emphasize the AI teammate's ability and thus primarily examined performance-based factors, whereas those adopting interpersonal trust measures tended to emphasize the benevolence aspect of trust and were more interested in behavior-based factors. This alignment in turn aids in the ability to form rational and theory-driven hypotheses while also enabling research participants to better relate measurement constructs to the novel system they interact with. In pursuing this alignment, researchers should not shy away from adapting a measure to better align with their definition of trust; however, researchers should state clearly how the adopted measurements are adapted. Further, while researchers may choose to measure only one dimension of trust if the agent's framing is unequivocally clear (e.g., clearly a tool), they should avoid conflating interpersonal and functional trust measurements by averaging ratings across dimensions into a single score, as doing so risks producing misleading or inconsistent results. Instead, interpersonal and functional trust measures (as well as behavioral measures) can be used in parallel for triangulation to increase validity.

In addition to measurements, our review suggests that other operational aspects could shape human participants' expectations of the AI agent, and thus should also align with the researcher's purposeful positioning of the AI agent on the functional-interpersonal continuum. These aspects range from the agent's visual representation, communication capability and modality, to the interdependence between the AI and humans. We do not imply that interfaces cannot be trusted as teammates. However, framing the AI agent as a system interface is likely to elicit different aspects of trust than framing it as a teammate. When the AI agent is framed as a system interface, certain dimensions (e.g., ability) are likely to be weighted more in trust evaluation than when the AI agent is framed as an embodied teammate who communicates like a human being. Importantly, the trusted agent being weaker in certain dimensions is likely not easily offset by its being stronger in others (i.e., humans might not trust a system that screws up and explains its failure, but still trust a teammate that does the same). In turn, if one operationalizes their AI teammate in a contrasting way to how they operationalize trust, then potentially significant but nonsensical results could manifest. Thus, research should ensure an alignment with what the AI teammates are designed to do and what humans should trust them to do, which will ensure internal validity, statistical reliability, and replicability.

Trust is already a multi-dimensional, complex construct in both human team literature and trust in automation literature. The positioning of the trusted agent in human-AI teaming contexts increases this complexity. In such contexts, conceptual and operational alignment is crucial for reducing the perplexity surrounding trust as a concept, for identifying dimensions of trust that are more or less important when the agent is viewed as a tool or a teammate, for distinguishing trust in the agent from team-level trust and other related concepts and advancing the empirical literature.

Guidelines for future research:

- Make sure the reference to the agent in the measurement items align with the conceptual framing.
- Use validated multi-dimensional measurements of trust instead of single-item measures.
- Avoid merging functional and interpersonal trust measurements into a single averaged score.

7.3 Future Directions

Our review suggests that the experimental set-up for a majority of HAT research on trust predominantly uses two-member teams consisting of only one human and one AI agent (77%), with AI and human team members performing tasks in a military or emergency task environment (67%). These provide several gaps and directions for future HCI and CSCW research to explore.

First, results from two-member teams may not extend to teams with more than two members. In multi-member teams, trust in an AI agent (or lack thereof) may be contagious and spread from one member to another through word-of-mouth or other social processes [28], which cannot be captured by studying two-member teams. This transitive property of trust is further explained in Huang and colleagues' [54] distributed dynamic team trust framework, wherein human-AI team trust is thought to change through direct and indirect interactions. The trust a human team member may have in an agent teammate can be influenced by a third human or agent teammate or another stakeholder related to the team such as a commanding officer, a subordinate, or a HR representative in charge of training. In the future, humans and AI agents may be required to work together in more complex teaming environments involving multiple HATs, where trust can be even more important to all stakeholders. Thus, the dynamic nature of trust and its associated real-time measurement, as well as the detection of trust or distrust spread within and across HATs are imperative and should be developed and validated in complex teaming contexts.

It is understandable that most human-AI teaming research on trust had the agent perform tasks predominantly in military and emergency contexts. AI agents are integrated into human teams to help execute tasks that are too dangerous for humans [66]. Humans' trust in them rests upon their precision, accuracy, and reliability in what they are trained for. However, with recent advancements in generative AI capable of creating novel contents and beyond, future HATs might require human and AI team members to collaborate on other types of tasks, such as creativity tasks, decision-making tasks, resolving conflicts of interest or conflicts of opinions [83], and the like. Trust perceptions for those scenarios will likely be impacted by different factors, including factors not yet identified.

With respect to communication between human and agent team members, it is disappointing that a great majority of human-AI teaming research on trust did not have their agent equipped with natural language processing and generation capabilities, let alone coupling NLP and NLG with task execution. Studies that did allow for natural language communication invariably employed the Wizard of Oz technique. While natural language capabilities are not a panacea, they can significantly contribute to the collaborative and interactive dimensions [72, 145] of human-AI teaming, provided they are implemented in a way that aligns with the team's goals and context. With rapid advancements in large language and multimodal models (e.g., ChatGPT), AI agents in HATs should soon be able to communicate with humans in natural language and execute tasks according to the conversation. As such, new dimensions of trust may be introduced. For instance, powerful models are trained with enormous datasets that may involve misinformation, which may result in the model producing biased or even harmful outputs and orders for the agent to perform. In such cases, calibrating human trust in the AI team member to an appropriate level is more advantageous than fostering blind trust [89]. Building on insights from our review, this calibration should incorporate both functional and interpersonal dimensions to inform the development of multifaceted trust calibration techniques, as highlighted in recent work [34].

Our review also suggests that agent-related factors, especially agent performance-based factors such as reliability and transparency of the agent teammate, have been the focus in HAT research on trust. And the finding that higher reliability leads to greater trust does not need further replication and validation in the same context and experiment setup. Future research should look more into behavior-based factors such as how the agent teammate should acknowledge their imperfection and mistake to gain humans' trust (e.g., trust repair), how their manner of communication (e.g., politeness, humor, tone of voice) may impact various dimensions of trust, and how repeated interactions with an agent over time may lead to fluctuating perceptions of trust.

It might be expected that HAT research would place greater emphasis on the teaming aspect, leading to more exploration of team-related factors. However, factors such as team composition, size, and role assignment remain underexplored, highlighting opportunities for future research. First, research that uses "teaming" language should explicitly explain how the collaboration between humans and AI in their study is *interdependent* in nature, addressing both taskwork (functional interdependence) and teamwork (interpersonal interdependence). This distinction is essential to differentiate human-AI teaming from human-AI interaction. Additionally, multidimensional team factors, including skill differentiation, authority differentiation, and temporal stability [53], warrant further investigation. For example, temporal stability offers a particularly compelling area of study. In HAT contexts, teams often consist of humans and agents collaborating temporarily for a specific task. This ad hoc nature of teaming raises important questions, as trust within teams may require time to develop. Consequently, the dynamics of trust in short-term teams may differ significantly from those in longer-term teams (e.g., those collaborating over months or years). Investigating these temporal variations could provide valuable insights into trust formation and maintenance in HATs. Further, recent research has shifted its focus to how teams function through intermittent and interdependent interactions, rather than solely on their composition, a concept referred to as "teamness" [26]. This concept provides a framework for examining multidimensional constructs and offers a pathway to classify teams as HATs and agents as teammates, based on their functional and interpersonal dynamics rather than their constituent elements. Leveraging these frameworks, a theoretically grounded approach could be developed to objectively determine whether an agent qualifies as a team member or is merely functioning as a tool.

7.4 Limitations

While this review provides a much needed synthesis of research on trust in HATs, it has several limitations that future research should address. First, the focus is solely on trust, though HATs also rely on other critical constructs such as communication, coordination,

interdependence, and cognition, all of which merit similar attention. Second, the review is constrained by the timeframe of the literature analyzed. While ongoing research continues to expand the understanding of trust in HATs, rapid advancements in AI technology are simultaneously shaping the field and influencing HAT research. This review offers a systematic analysis of trust in HATs, but ongoing dialogue and updates will be needed to reflect future progress in HATs and the evolution of AI technologies. Third, the reviewed research has predominantly been conducted in Global North countries, leaving a critical gap in understanding perspectives from the Global South, where socio-economic factors, technological infrastructure, access to AI, as well as regulatory frameworks and ethical considerations may differ significantly [62, 99], potentially influencing how trust is built and maintained within HATs. Cross-cultural approaches that incorporate perspectives from a wider range of global contexts would help to create more inclusive and effective frameworks for understanding trust in HATs. Fourth, this review intentionally excludes robotic systems, focusing instead on digital AI technologies, which have grown substantial enough to warrant their own dedicated analysis. However, as robotic systems increasingly integrate AI, they cannot be overlooked. A separate review is needed to explicitly examine trust in AI-enabled human-robot teams, addressing unique considerations such as physical safety. Finally, when performing a meta-analysis on specific design characteristic impacts in HATs, this review found a number of empirical studies with incomplete reporting. While this limitation is not the result of this review process, it does bring up an important challenge in the HAT domain. In particular, future work needs to improve the fidelity of its reporting to ensure future HAT reviews can more accurately assess the domain.

8 Conclusion

As artificial intelligence rapidly matures and integrates into teams to form human-AI teams (HATs) in various domains, trust becomes a critical concern as humans and AI agents work more closely and interdependently with one another. A holistic understanding of the current state of the science of trust in human-AI teaming is much needed before an explosion of HATs is implemented in the real world. This work offers such an understanding by documenting the characteristics of the AI agent, team, and environment that delineate the boundaries of the extant knowledge of trust in HATs, identifying and categorizing the measurements of trust, as well as the factors examined to have an impact on trust. By synthesizing the unorganized domain, we have consolidated the existing scientific knowledge of human trust in AI teammates examined in HAT research, to provide researchers and practitioners with a coherent understanding of this area. We also provide guidelines and directions for HCI and CSCW researchers to standardize research methodologies in conducting trust research in HATs, and to broaden the scope of such research.

References

[1] Kathleen Allen, Richard Bergin, and Kenneth Pickar. 2004. Exploring trust, group satisfaction, and performance in geographically dispersed and co-located university technology commercialization teams. In *VentureWell. Proceedings of Open, the Annual Conference*. National Collegiate Inventors & Innovators Alliance, 201.

[2] Zahra Ashktorab, Casey Dugan, James Johnson, Qian Pan, Wei Zhang, Sadhana Kumaravel, and Murray Campbell. 2021. Effects of communication directionality and AI agent differences in human-AI interaction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[3] Christiane Atzmüller and Peter M Steiner. 2010. Experimental vignette studies in survey research. *Methodology* (2010).

[4] Benoit A Aubert and Barbara L Kelsey. 2003. Further understanding of trust and performance in virtual teams. *Small group research* 34, 5 (2003), 575–618.

[5] Eugénie Avril. 2022. Providing different levels of accuracy about the reliability of automation to a human operator: impact on human performance. *Ergonomics* 66, 2 (2022), 217–226.

[6] Gagan Bansal, Alison Marie Smith-Renner, Zana Buçinca, Tongshuang Wu, Kenneth Holstein, Jessica Hullman, and Simone Stumpf. 2022. Workshop on Trust and Reliance in AI-Human Teams (TRAIT). In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. ACM, New Orleans LA USA, 1–6. doi:10.1145/3491101.3503704

[7] Gagan Bansal, Tongshuang Wu, Joyce Zhou, Raymond Fok, Besmira Nushi, Ece Kamar, Marco Tulio Ribeiro, and Daniel Weld. 2021. Does the whole exceed its parts? the effect of ai explanations on complementary team performance. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–16.

[8] Michaela Benk, Sophie Kerstan, Florian von Wangenheim, and Andrea Ferrario. 2024. Twenty-four years of empirical research on trust in AI: a bibliometric review of trends, overlooked issues, and future directions. *AI & SOCIETY* (2024), 1–24.

[9] Adella Bhaskara, Lain Duong, James Brooks, Ryan Li, Ronan McInerney, Michael Skinner, Helen Pongracic, and Shayne Loft. 2021. Effect of automation transparency in the management of multiple unmanned vehicles. *Applied Ergonomics* 90 (2021), 103243.

[10] Shawaiz Bhatti, Mustafa Demir, Nancy J Cooke, and Craig J Johnson. 2021. Assessing communication and trust in an ai teammate in a dynamic task environment. In *2021 ieee 2nd international conference on human-machine systems (ichms)*. IEEE, 1–6.

[11] J Martin Bland and DouglasG Altman. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet* 327, 8476 (1986), 307–310.

[12] Philip Bobko, Leanne Hirshfield, Lucca Eloy, Cara Spencer, Emily Doherty, Jack Driscoll, and Hannah Obolsky. 2023. Human-agent teaming and trust calibration: a theoretical framework, configurable testbed, empirical illustration, and implications for the development of adaptive systems. *Theoretical Issues in Ergonomics Science* 24, 3 (2023), 310–334.

[13] Sylvie Borau, Tobias Otterbring, Sandra Laporte, and Samuel Fosso Wamba. 2021. The most human bot: Female gendering increases humanness perceptions of bots and acceptance of AI. *Psychology & Marketing* 38, 7 (2021), 1052–1068.

[14] Michael Borenstein, Harris Cooper, L Hedges, and J Valentine. 2009. Effect sizes for continuous data. *The handbook of research synthesis and meta-analysis* 2 (2009), 221–235.

[15] Zana Buçinca, Phoebe Lin, Krzysztof Z Gajos, and Elena L Glassman. 2020. Proxy tasks and subjective measures can be misleading in evaluating explainable AI systems. In *Proceedings of the 25th international conference on intelligent user interfaces*. 454–464.

[16] Christopher S Calhoun, Philip Bobko, Jennie J Gallimore, and Joseph B Lyons. 2019. Linking precursors of interpersonal trust to human-automation trust: An expanded typology and exploratory experiment. *Journal of Trust Research* 9, 1 (2019), 28–46.

[17] Julia Cambre and Chinmay Kulkarni. 2019. One voice fits all? Social implications and research challenges of designing voices for smart devices. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–19.

[18] Colleen M Carpinella, Alisa B Wyman, Michael A Perez, and Steven J Stroessner. 2017. The robotic social attributes scale (RoSAS) development and validation. In *Proceedings of the 2017 ACM/IEEE International Conference on human-robot interaction*. 254–262.

[19] Jessie YC Chen, Michael J Barnes, Anthony R Selkowitz, Kimberly Stowers, Shan G Lakhmani, and Nicholas Kasdaglis. 2016. Human-autonomy teaming and agent transparency. In *Companion Publication of the 21st International Conference on Intelligent User Interfaces*. 28–31.

[20] Jessie Y Chen, Katelyn Procci, Michael Boyce, Julia Wright, Andre Garcia, and Michael Barnes. 2014. Situation awareness-based agent transparency. *US Army Research Laboratory* April (2014), 1–29.

[21] Lara Christoforakos, Alessio Gallucci, Tinatini Surmava-Große, Daniel Ullrich, and Sarah Diefenbach. 2021. Can robots earn our trust the same way humans do? A systematic exploration of competence, warmth, and anthropomorphism as determinants of trust development in HRI. *Frontiers in Robotics and AI* 8 (2021), 640444.

[22] Andrew S Clare, Mary L Cummings, and Nelson P Repenning. 2015. Influencing trust for human–automation collaborative scheduling of multiple unmanned vehicles. *Human factors* 57, 7 (2015), 1208–1218.

[23] Jacob Cohen. 2013. *Statistical power analysis for the behavioral sciences*. Routledge.

[24] Myke C Cohen, Mustafa Demir, Erin K Chiou, and Nancy J Cooke. 2021. The dynamics of trust and verbal anthropomorphism in human-autonomy teaming. In *2021 IEEE 2nd international conference on human-machine systems (ICHMS)*. IEEE, 1–6.

[25] Jason A Colquitt, Brent A Scott, and Jeffery A LePine. 2007. Trust, trustworthiness, and trust propensity: a meta-analytic test of their unique relationships with risk taking and job performance. *Journal of applied psychology* 92, 4 (2007), 909.

[26] Nancy J Cooke, Myke C Cohen, Walter C Fazio, Laura H Inderberg, Craig J Johnson, Glenn J Lematta, Matthew Peel, and Aaron Teo. 2023. From teams to teamness: Future directions in the science of team cognition. *Human Factors* (2023), 00187208231162449.

[27] Ana Cristina Costa and Neil Anderson. 2011. Measuring trust in teams: Development and validation of a multifaceted measure of formative and reflective indicators of team trust. *European Journal of Work and Organizational Psychology* 20, 1 (2011), 119–154.

[28] Ana Cristina Costa, C Ashley Fulmer, and Neil R Anderson. 2018. Trust in work teams: An integrative review, multilevel model, and future directions. *Journal of Organizational Behavior* 39, 2 (2018), 169–184.

[29] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. 1993. Wizard of Oz studies—why and how. *Knowledge-based systems* 6, 4 (1993), 258–266.

[30] Ewart J De Visser, Samuel S Monfort, Kimberly Goodyear, Li Lu, Martin O'Hara, Mary R Lee, Raja Parasuraman, and Frank Krueger. 2017. A little anthropomorphism goes a long way: Effects of oxytocin on trust, compliance, and team performance with automated agents. *Human factors* 59, 1 (2017), 116–133.

[31] Mustafa Demir, Nathan J McNeese, Jaime C Gorman, Nancy J Cooke, Christopher W Myers, and David A Grimm. 2021. Exploration of teammate trust and interaction dynamics in human-autonomy teaming. *IEEE Transactions on Human-Machine Systems* 51, 6 (2021), 696–705.

[32] Na Du, Kevin Y Huang, and X Jessie Yang. 2020. Not all information is equal: effects of disclosing different types of likelihood information on trust, compliance and reliance, and task performance in human-automation teaming. *Human factors* 62, 6 (2020), 987–1001.

[33] Wen Duan, Nan Weng, Matthew J Scalia, Ruihao Zhang, Jessica Tuttle, Xiaoyun Yin, Shiwen Zhou, Guo Freeman, Jamie Gorman, Gregory Funke, et al. 2024. Getting Along With Autonomous Teammates: Understanding the Socio-Emotional and Teaming Aspects of Trust in Human-Autonomy Teams. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. SAGE Publications Sage CA: Los Angeles, CA, 10711813241272123.

[34] Wen Duan, Shiwen Zhou, Matthew J Scalia, Xiaoyun Yin, Nan Weng, Ruihao Zhang, Guo Freeman, Nathan McNeese, Jamie Gorman, and Michael Tolston. 2024. Understanding the Evolvement of Trust Over Time within Human-AI Teams. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW2 (2024), 1–31.

[35] Thomas Ellwart, Nathalie Schauffel, Conny H Antoni, and Ingo J Timm. 2022. I vs. robot: Sociodigital self-comparisons in hybrid teams from a theoretical, empirical, and practical perspective. *Gruppe. Interaktion. Organisation. Zeitschrift Für Angewandte Organisationspsychologie (GIO)* 53, 3 (2022), 273–284.

[36] Xiaocong Fan, Sooyoung Oh, Michael McNeese, John Yen, Haydee Cuevas, Laura Strater, and Mica R Endsley. 2008. The influence of agent reliability on trust in human-agent collaboration. In *Proceedings of the 15th European conference on Cognitive ergonomics: the ergonomics of cool interaction*. 1–8.

[37] Susan T Fiske, Amy JC Cuddy, and Peter Glick. 2007. Universal dimensions of social cognition: Warmth and competence. *Trends in cognitive sciences* 11, 2 (2007), 77–83.

[38] Susan T Fiske, Amy JC Cuddy, Peter Glick, and Jun Xu. 2002. A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology* 82, 6 (2002), 878–902.

[39] Christopher Flathmann, Wen Duan, Nathan J Mcneese, Allyson Hauptman, and Rui Zhang. 2024. Empirically Understanding the Potential Impacts and Process of Social Influence in Human-AI Teams. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–32.

[40] Brian J Fogg and Hsiang Tseng. 1999. The elements of computer credibility. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 80–87.

[41] Christoph Fuchs and Adamantios Diamantopoulos. 2009. Using single-item measures for construct measurement in management research: Conceptual issues and application guidelines. *Die Betriebswirtschaft* 69, 2 (2009), 195.

[42] Barney G Glaser, Anselm L Strauss, and Elizabeth Strutzel. 1968. The discovery of grounded theory; strategies for qualitative research. *Nursing research* 17, 4 (1968), 364.

[43] Ella Glikson and Anita Williams Woolley. 2020. Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals* 14, 2 (2020), 627–660.

[44] Svyatoslav Guznov, Alexander Nelson, Joseph Lyons, and David Dycus. 2015. The effects of automation reliability and multi-tasking on trust and reliance in a simulated unmanned system control task. In *HCI International 2015-Posters' Extended Abstracts: International Conference, HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015. Proceedings, Part II 17*. Springer, 616–621.

[45] Feyza Merve Hafizoglu and Sandip Sen. 2018. Reputation based trust in human-agent teamwork without explicit coordination. In *Proceedings of the 6th international conference on human-agent interaction*. 238–245.

[46] Feyza Merve Hafizoğlu and Sandip Sen. 2019. Understanding the influences of past experience on trust in human-agent teamwork. *ACM Transactions on Internet Technology (TOIT)* 19, 4 (2019), 1–22.

[47] Feyza Merve Hafizoğlu and Sandip Sen. 2020. Comparing human trust attitudes towards human and agent teammates. In *Proceedings of the 8th International Conference on Human-Agent Interaction*. 50–59.

[48] Nader Hanna and Deborah Richards. 2018. The impact of multimodal communication on a shared mental model, trust, and commitment in human–intelligent virtual agent teams. *Multimodal Technologies and Interaction* 2, 3 (2018), 48.

[49] Kerstin S Haring, Kelly M Satterfield, Chad C Tossell, Ewart J De Visser, Joseph R Lyons, Vincent F Mancuso, Victor S Finomore, and Gregory J Funke. 2021. Robot authority in human-robot teaming: Effects of human-likeness and physical embodiment on compliance. *Frontiers in Psychology* 12 (2021), 625713.

[50] Allyson I Hauptman, Wen Duan, and Nathan J Mcneese. 2022. The Components of Trust for Collaborating With AI Colleagues. In *Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing*. 72–75.

[51] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors* 57, 3 (2015), 407–434.

[52] Robert R Hoffman, Shane T Mueller, Gary Klein, and Jordan Litman. 2018. Metrics for explainable AI: Challenges and prospects. *arXiv preprint arXiv:1812.04608* (2018).

[53] John R Hollenbeck, Bianca Beersma, and Maartje E Schouten. 2012. Beyond team types and taxonomies: A dimensional scaling conceptualization for team description. *Academy of Management Review* 37, 1 (2012), 82–106.

[54] Lixiao Huang, Nancy J Cooke, Robert S Gutzwiller, Spring Berman, Erin K Chiou, Mustafa Demir, and Wenlong Zhang. 2021. Distributed dynamic team trust in human, artificial intelligence, and robot teaming. In *Trust in human-robot interaction*. Elsevier, 301–319.

[55] Alon Jacovi, Ana Marasović, Tim Miller, and Yoav Goldberg. 2021. Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in ai. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 624–635.

[56] Sirkka L Jarvenpaa, Kathleen Knoll, and Dorothy E Leidner. 1998. Is anybody out there? Antecedents of trust in global virtual teams. *Journal of management information systems* 14, 4 (1998), 29–64.

[57] Theodore Jensen, Yusuf Albayram, Mohammad Maifi Hasan Khan, Md Abdullah Al Fahim, Ross Buck, and Emil Coman. 2019. The apple does fall far from the tree: user separation of a system from its developers in human-automation trust repair. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 1071–1082.

[58] Stephanie Tulk Jesso, William G Kennedy, and Eva Wiese. 2020. Behavioral cues of humanness in complex environments: How people engage with human and artificially intelligent agents in a multiplayer videogame. *Frontiers in Robotics and AI* 7 (2020).

[59] Jiun-Yin Jian, Ann M Bisantz, and Colin G Drury. 2000. Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics* 4, 1 (2000), 53–71.

[60] Craig J Johnson, Mustafa Demir, Nathan J McNeese, Jamie C Gorman, Alexandra T Wolff, and Nancy J Cooke. 2021. The impact of training on human–autonomy team communications and trust calibration. *Human Factors* 65, 7 (2021), 1554–1570.

[61] Craig J Johnson, Mustafa Demir, Garrett M Zabala, Hongbei He, David A Grimm, Cody Radigan, Alexandra T Wolff, Nancy J Cooke, Nathan J McNeese, and Jamie C Gorman. 2020. Training and verbal communications in human-autonomy teaming under degraded conditions. In *2020 IEEE conference on cognitive and computational aspects of situation management (CogSIMA)*. IEEE, 53–58.

[62] Shivani Kapania, Oliver Siy, Gabe Clapper, Azhagu Meena Sp, and Nithya Sambasivan. 2022. " Because AI is 100% right and safe": User attitudes and sources of AI authority in India. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–18.

[63] John F Kelley. 1983. An empirical methodology for writing user-friendly natural language computer applications. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 193–196.

[64] Esther S Kox, José H Kerstholt, Tom F Hueting, and Peter W de Vries. 2021. Trust repair in human-agent teams: the effectiveness of explanations and expressing regret. *Autonomous Agents and Multi-Agent Systems* 35, 2 (2021), 30.

[65] Esther Siegling Kox, LB Siegling, and Jose H Kerstholt. 2022. Trust development in military and civilian human–agent teams: the effect of social-cognitive recovery strategies. *International Journal of Social Robotics* 14, 5 (2022), 1323–1338.

[66] Andrea Krausman, Catherine Neubauer, Daniel Forster, Shan Lakhmani, Anthony L Baker, Sean M Fitzhugh, Gregory Gremillion, Julia L Wright, Jason S Metcalfe, and Kristin E Schaefer. 2022. Trust measurement in human-autonomy teams: Development of a conceptual toolkit. *ACM Transactions on Human-Robot Interaction (THRI)* 11, 3 (2022), 1–58.

[67] Philipp Kulms and Stefan Kopp. 2016. The effect of embodiment and competence on trust and cooperation in human–agent interaction. In *Intelligent Virtual Agents: 16th International Conference, IVA 2016, Los Angeles, CA, USA, September 20–23, 2016, Proceedings 16*. Springer, 75–84.

[68] Philipp Kulms and Stefan Kopp. 2019. More human-likeness, more trust? The effect of anthropomorphism on self-reported and behavioral trust in continued and interdependent human-agent cooperation. In *Proceedings of mensch und computer 2019*. 31–42.

[69] Markus Langer, Tim Hunsicker, Tina Feldkamp, Cornelius J König, and Nina Grgić-Hlača. 2022. "Look! It'sa computer program! It's an algorithm! It's AI!": Does terminology affect human perceptions and evaluations of algorithmic decision-making systems?. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–28.

[70] John D Lee and Neville Moray. 1994. Trust, self-confidence, and operators' adaptation to automation. *International journal of human-computer studies* 40, 1 (1994), 153–184.

[71] John D Lee and Katrina A See. 2004. Trust in automation: Designing for appropriate reliance. *Human factors* 46, 1 (2004), 50–80.

[72] Mina Lee, Percy Liang, and Qian Yang. 2022. Coauthor: Designing a human-ai collaborative writing dataset for exploring language model capabilities. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–19.

[73] Jinchao Lin, April Rose Panganiban, Gerald Matthews, Katey Gibbins, Emily Ankeney, Carlie See, Rachel Bailey, and Michael Long. 2022. Trust in the danger zone: individual differences in confidence in robot threat assessments. *Frontiers in psychology* 13 (2022), 601523.

[74] Shayne Loft, Adella Bhaskara, Brittany A Lock, Michael Skinner, James Brooks, Ryan Li, and Jason Bell. 2021. The impact of transparency and decision risk on human–automation teaming outcomes. *Human Factors* 65, 5 (2021), 846–861.

[75] Fabrice Lumineau. 2017. How contracts influence trust and distrust. *Journal of management* 43, 5 (2017), 1553–1577.

[76] Maria Madsen and Shirley Gregor. 2000. Measuring human-computer trust. In *11th australasian conference on information systems*, Vol. 53. Citeseer, 6–8.

[77] Stephen Marsh and Mark R Dibben. 2003. The role of trust in information science and technology. *Annual Review of Information Science and Technology (ARIST)* 37 (2003), 465–98.

[78] Tetsuya Matsui and Atsushi Koike. 2021. Who is to blame? The appearance of virtual agents and the attribution of perceived responsibility. *Sensors* 21, 8 (2021), 2646.

[79] Gerald Matthews, Jinchao Lin, April Rose Panganiban, and Michael D Long. 2019. Individual differences in trust in autonomous robots: Implications for transparency. *IEEE Transactions on Human-Machine Systems* 50, 3 (2019), 234–244.

[80] Roger C Mayer, James H Davis, and F David Schoorman. 1995. An integrative model of organizational trust. *Academy of management review* 20, 3 (1995), 709–734.

[81] Roger C Mayer and Mark B Gavin. 2005. Trust in management and performance: Who minds the shop while the employees watch the boss? *Academy of Management Journal* 48, 5 (2005), 874–888.

[82] Daniel J McAllister. 1995. Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of management journal* 38, 1 (1995), 24–59.

[83] Joseph Edward McGrath. 1984. *Groups: Interaction and performance*. Vol. 14. Prentice-Hall Englewood Cliffs, NJ.

[84] D Harrison McKnight and Norman L Chervany. 2000. What is trust? A conceptual analysis and an interdisciplinary model. (2000).

[85] Harrison McKnight, Michelle Carter, and Paul Clay. 2009. Trust in technology: Development of a set of constructs and measures. (2009).

[86] Nathan J McNeese, Mustafa Demir, Nancy J Cooke, and Christopher Myers. 2018. Teaming with a synthetic teammate: Insights into human-autonomy teaming. *Human factors* 60, 2 (2018), 262–273.

[87] Nathan J McNeese, Mustafa Demir, Nancy J Cooke, and Manrong She. 2021. Team Situation Awareness and Conflict: A Study of Human–Machine Teaming. *Journal of Cognitive Engineering and Decision Making* 15, 2-3 (2021), 83–96.

[88] Siddharth Mehrotra, Chadha Degachi, Oleksandra Vereschak, Catholijn M Jonker, and Myrthe L Tielman. 2024. A Systematic Review on Fostering Appropriate Trust in Human-AI Interaction: Trends, Opportunities and Challenges. *ACM Journal on Responsible Computing* 1, 4 (2024), 1–45.

[89] Siddharth Mehrotra, Carolina Centeio Jorge, Catholijn M Jonker, and Myrthe L Tielman. 2024. Integrity-based explanations for fostering appropriate trust in AI agents. *ACM Transactions on Interactive Intelligent Systems* 14, 1 (2024), 1–36.

[90] Joseph E Mercado, Michael A Rupp, Jessie YC Chen, Michael J Barnes, Daniel Barber, and Katelyn Procci. 2016. Intelligent agent transparency in human–agent teaming for Multi-UxV management. *Human factors* 58, 3 (2016), 401–415.

[91] Stephanie M Merritt. 2011. Affective processes in human–automation interactions. *Human Factors* 53, 4 (2011), 356–370.

[92] Bonnie M Muir. 1989. *Operators' trust in and use of automatic controllers in a supervisory process control task*. Ph. D. Dissertation. University of Toronto.

[93] Bonnie M Muir and Neville Moray. 1996. Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics* 39, 3 (1996), 429–460.

[94] Imani Munyaka, Zahra Ashktorab, Casey Dugan, James Johnson, and Qian Pan. 2023. Decision Making Strategies and Team Efficacy in Human-AI Teams. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–24.

[95] Geoff Musick, Thomas A. O'Neill, Beau G. Schelble, Nathan J. McNeese, and Jonn B. Henke. 2021. What Happens When Humans Believe Their Teammate is an AI? An Investigation into Humans Teaming with Autonomy. *Computers in Human Behavior* 122 (Sept. 2021), 106852. doi:10.1016/j.chb.2021.106852

[96] Clifford Nass, BJ Fogg, and Youngme Moon. 1996. Can computers be teammates? *International Journal of Human-Computer Studies* 45, 6 (1996), 669–678.

[97] Clifford Nass, Eun-Young Kim, and Eun-Ju Lee. 1998. When my face is the interface: An experimental comparison of interacting with one's own face or someone else's face. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 148–154.

[98] Clifford Nass, Jonathan Steuer, and Ellen R Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 72–78.

[99] Chinasa T Okolo, Srujana Kamath, Nicola Dell, and Aditya Vashistha. 2021. "It cannot do all of my work": community health worker perceptions of AI-enabled mobile health applications in rural India. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–20.

[100] Thomas O'Neill, Nathan McNeese, Amy Barron, and Beau Schelble. 2020. Human–autonomy teaming: A review and analysis of the empirical literature. *Human Factors* (2020), 0018720820960865.

[101] Matthew J Page, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M Tetzlaff, Elie A Akl, Sue E Brennan, et al. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Bmj* 372 (2021).

[102] April Rose Panganiban, Gerald Matthews, and Michael D Long. 2020. Transparency in autonomous teammates: intention to support as teaming information. *Journal of Cognitive Engineering and Decision Making* 14, 2 (2020), 174–190.

[103] Raja Parasuraman, Thomas B Sheridan, and Christopher D Wickens. 2000. A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans* 30, 3 (2000), 286–297.

[104] David V Pynadath, Ning Wang, and Sreekar Kamireddy. 2019. A Markovian method for predicting trust behavior in human-agent interaction. In *Proceedings of the 7th international conference on human-agent interaction*. 171–178.

[105] David V Pynadath, Ning Wang, Ericka Rovira, and Michael J Barnes. 2018. Clustering behavior to recognize subjective beliefs in human-agent teams. In *Proceedings of the 17th international conference on autonomous agents and multiagent systems*. 1495–1503.

[106] Summer Rebensky, Kendall Carmody, Cherrise Ficke, Meredith Carroll, and Winston Bennett. 2022. Teammates instead of tools: The impacts of level of autonomy on mission performance and human–agent teaming dynamics in multi-agent distributed teams. *Frontiers in Robotics and AI* 9 (2022), 782134.

[107] Tobias Rieger, Eileen Roesler, and Dietrich Manzey. 2022. Challenging presumed technological superiority when working with (artificial) colleagues. *Scientific Reports* 12, 1 (2022), 3768.

[108] Laurel D Riek. 2012. Wizard of oz studies in hri: a systematic review and new reporting guidelines. *Journal of Human-Robot Interaction* 1, 1 (2012), 119–136.

[109] Jennifer M Roche, Arkady Zgonnikov, and Laura M Morett. 2021. Cognitive processing of miscommunication in interactive listening: An evaluation of listener indecision and cognitive effort. *Journal of Speech, Language, and Hearing Research* 64, 1 (2021), 159–175.

[110] Robert Rosenthal. 1979. The file drawer problem and tolerance for null results. *Psychological bulletin* 86, 3 (1979), 638.

[111] Gunar Roth, Axel Schulte, Fabian Schmitt, and Yannick Brand. 2019. Transparency for a workload-adaptive cognitive agent in a manned–unmanned teaming application. *IEEE Transactions on Human-Machine Systems* 50, 3 (2019), 225–233.

[112] Harvey Sacks, Emanuel A Schegloff, and Gail Jefferson. 1978. A simplest systematics for the organization of turn taking for conversation. In *Studies in the organization of conversational interaction*. Elsevier, 7–55.

[113] Matthew J Scalia. 2022. Developing Objective Communication-based Measures of Trust for Human-Autonomy Teams. (2022).

[114] Kristin E Schaefer, Jessie YC Chen, James L Szalma, and Peter A Hancock. 2016. A meta-analysis of factors influencing the development of trust in automation:

Implications for understanding autonomy in future systems. *Human factors* 58, 3 (2016), 377–400.

[115] James Schaffer, John O'Donovan, James Michaelis, Adrienne Raglin, and Tobias Höllerer. 2019. I can do better than your AI: expertise and explanations. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 240–251.

[116] Beau G Schelble, Christopher Flathmann, Nathan J McNeese, Guo Freeman, and Rohit Mallick. 2022. Let's Think Together! Assessing Shared Mental Models, Performance, and Trust in Human-Agent Teams. *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP (2022), 1–29.

[117] Beau G Schelble, Christopher Flathmann, Nathan J McNeese, Thomas O'Neill, Richard Pak, and Moses Namara. 2022. Investigating the Effects of Perceived Teammate Artificiality on Human Performance and Cognition. *International Journal of Human–Computer Interaction* (2022), 1–16.

[118] Beau G Schelble, Christopher Flathmann, Geoff Musick, Nathan J McNeese, and Guo Freeman. 2022. I see you: Examining the role of spatial information in human-agent teams. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–27.

[119] Beau G Schelble, Jeremy Lopez, Claire Textor, Rui Zhang, Nathan J McNeese, Richard Pak, and Guo Freeman. 2022. Towards ethical AI: Empirically investigating dimensions of AI ethics, trust repair, and performance in human-AI teaming. *Human Factors* 66, 4 (2022), 1037–1055.

[120] Sarah Sebo, Brett Stoll, Brian Scassellati, and Malte F Jung. 2020. Robots in groups and teams: a literature review. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–36.

[121] Anthony R Selkowitz, Shan G Lakhmani, and Jessie YC Chen. 2017. Using agent transparency to support situation awareness of the Autonomous Squad Member. *Cognitive Systems Research* 46 (2017), 13–25.

[122] Mona SharifHeravi, John R Taylor, Christopher J Stanton, Sandra Lambeth, and Christopher Shanahan. 2020. It'sa Disaster! Factors Affecting Trust Development and Repair Following Agent Task Failure. In *Proceedings of the 2020 Australasian Conference on Robotics and Automation (ACRA 2020)*. 8–10.

[123] Andy P Siddaway, Alex M Wood, and Larry V Hedges. 2019. How to do a systematic review: a best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual review of psychology* 70 (2019), 747–770.

[124] Ho Chit Siu, Jaime Peña, Edenna Chen, Yutai Zhou, Victor Lopez, Kyle Palko, Kimberlee Chang, and Ross Allen. 2021. Evaluation of human-ai teams for learned and rule-based agents in hanabi. *Advances in Neural Information Processing Systems* 34 (2021), 16183–16195.

[125] Gyrd Skraaning and Greg A Jamieson. 2021. Human performance benefits of the automation transparency design principle: Validation and variation. *Human factors* 63, 3 (2021), 379–401.

[126] Lynn Smith-Lovin and Charles Brody. 1989. Interruptions in group discussions: The effects of gender and group composition. *American Sociological Review* (1989), 424–435.

[127] Kimberly Stowers, Nicholas Kasdaglis, Michael A Rupp, Olivia B Newton, Jessie YC Chen, and Michael J Barnes. 2020. The IMPACT of agent transparency on human performance. *IEEE Transactions on Human-Machine Systems* 50, 3 (2020), 245–253.

[128] Stine Strand. 2001. *Trust and automation: the influence of automation malfunctions and system feedback on operator trust.* Technical Report. Institutt for energiteknikk.

[129] Meinald T Thielsch, Sarah M Meeßen, and Guido Hertel. 2018. Trust and distrust in information systems at the workplace. *PeerJ* 6 (2018), e5483.

[130] Hiroyuki Tokushige, Takuji Narumi, Sayaka Ono, Yoshitaka Fuwamoto, Tomohiro Tanikawa, and Michitaka Hirose. 2017. Trust lengthens decision time on unexpected recommendations in human-agent interaction. In *Proceedings of the 5th international conference on human agent interaction*. 245–252.

[131] Suzanne Tolmeijer, Markus Christen, Serhiy Kandul, Markus Kneer, and Abraham Bernstein. 2022. Capable but amoral? Comparing AI and human expert collaboration in ethical decision making. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–17.

[132] Silvia Tulli, Filipa Correia, Samuel Mascarenhas, Samuel Gomes, Francisco S Melo, and Ana Paiva. 2019. Effects of agents' transparency on teamwork. In *International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems*. Springer, 22–37.

[133] Anna-Sophie Ulfert, Eleni Georganta, Carolina Centeio Jorge, Siddharth Mehrotra, and Myrthe Tielman. 2024. Shaping a multidisciplinary understanding of team trust in human-AI teams: a theoretical framework. *European Journal of Work and Organizational Psychology* 33, 2 (2024), 158–171.

[134] Daniel Ullman and Bertram F Malle. 2019. MDMT: Multi-dimensional measure of trust.

[135] Ruben S Verhagen, Mark A Neerincx, and Myrthe L Tielman. 2022. The influence of interdependence and a transparent or explainable communication style on human-robot teamwork. *Frontiers in Robotics and AI* 9 (2022), 993997.

[136] James C Walliser, Ewart J de Visser, Eva Wiese, and Tyler H Shaw. 2019. Team structure and team building improve human–machine teaming with autonomous agents. *Journal of Cognitive Engineering and Decision Making* 13, 4 (2019), 258–278.

[137] Isaac Wang, Jesse Smith, and Jaime Ruiz. 2019. Exploring virtual agents for augmented reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[138] Ning Wang, David V Pynadath, and Susan G Hill. 2016. Trust calibration within a human-robot team: Comparing automatically generated explanations. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 109–116.

[139] Xinru Wang and Ming Yin. 2021. Are Explanations Helpful? A Comparative Study of the Effects of Explanations in AI-Assisted Decision-Making. In *26th International Conference on Intelligent User Interfaces*. ACM, College Station TX USA, 318–328. doi:10.1145/3397481.3450650

[140] Rebecca Wiczorek and Dietrich Manzey. 2014. Supporting attention allocation in multitask environments: Effects of likelihood alarm systems on trust, behavior, and performance. *Human factors* 56, 7 (2014), 1209–1221.

[141] Ryan W Wohleber, Kimberly Stowers, Jessie YC Chen, and Michael Barnes. 2017. Effects of agent transparency and communication framing on human-agent teaming. In *2017 IEEE international conference on systems, man, and cybernetics (SMC)*. IEEE, 3427–3432.

[142] Julia L Wright, Jessie YC Chen, Michael J Barnes, and Peter A Hancock. 2016. Agent reasoning transparency's effect on operator workload. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 60. SAGE Publications Sage CA: Los Angeles, CA, 249–253.

[143] Julia L Wright, Jessie YC Chen, and Shan G Lakhmani. 2019. Agent transparency and reliability in human–robot interaction: The influence on user confidence and perceived reliability. *IEEE Transactions on Human-Machine Systems* 50, 3 (2019), 254–263.

[144] Julia L Wright, Shan G Lakhmani, and Jessie YC Chen. 2022. Bidirectional Communications in Human-Agent Teaming: The Effects of Communication Style and Feedback. *International Journal of Human–Computer Interaction* (2022), 1–14.

[145] Qian Yang, Justin Cranshaw, Saleema Amershi, Shamsi T Iqbal, and Jaime Teevan. 2019. Sketching nlp: A case study of exploring the right things to design with language intelligence. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[146] X Jessie Yang, Christopher Schemanske, and Christine Searle. 2021. Toward quantifying trust dynamics: How people adjust their trust after moment-to-moment interaction with automation. *Human Factors* 65, 5 (2021), 862–878.

[147] X Jessie Yang, Vaibhav V Unhelkar, Kevin Li, and Julie A Shah. 2017. Evaluating effects of user experience and system transparency on trust in automation. In *Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction*. 408–416.

[148] Mateusz Żarkowski. 2019. Multi-party turn-taking in repeated human–robot interactions: an interdisciplinary evaluation. *International Journal of Social Robotics* 11, 5 (2019), 693–707.

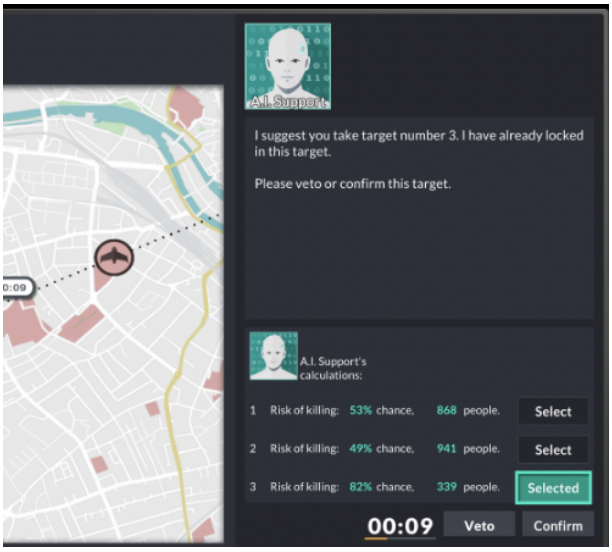
[149] Qiaoning Zhang, Matthew L Lee, and Scott Carter. 2022. You Complete Me: Human-AI Teams and Complementary Expertise. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3491102.3517791 event-place: New Orleans, LA, USA.

[150] Rui Zhang, Wen Duan, Christopher Flathmann, Nathan McNeese, Guo Freeman, and Alyssa Williams. 2023. Investigating AI teammate communication strategies and their impact in human-AI teams for effective teamwork. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–31.

[151] Rui Zhang, Nathan J McNeese, Guo Freeman, and Geoff Musick. 2021. " An Ideal Human" Expectations of AI Teammates in Human-AI Teaming. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–25.

(a) Example of a robotic/iconic static image (the white robot figure on the bottom of the matrix, next to the pink human-like icon that represents the human participant), from [135].

(b) Example of a robotic/iconic animated avatar, from [65].



(c) Example of a humanoid static image, from [131].

(d) Example of a humanoid animated avatar, from [119].

Figure 10: Examples of Agents' Visual Representations in the Reviewed Research.

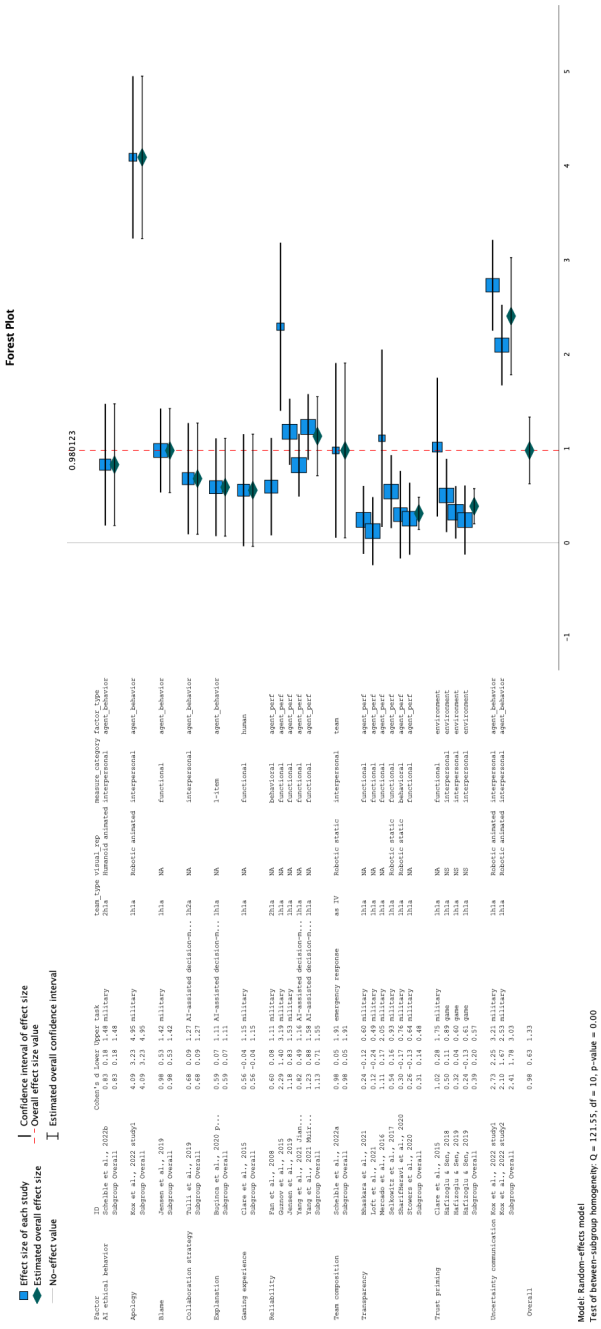


Figure 11: Forest plot of all the factors by study ID, with effect sizes and their corresponding 95% confidence intervals. The columns also show the visual representation (NA stands for agents without visual representation, NS stands for studies that did not specify agent's visual representation), the team and task characteristics, the factor type, and the measurement type. The lines that don't cross the null-effect (0) vertical line are significant at the p < .05 level. The squares represent the point estimate for each study. The size of the squares represents the weight in the meta-analysis. The diamonds represent overall effect sizes.