AI-teaming: Redefining collaboration in the digital era

Schmutz, J.B.; Outland, N.; Kerstan, S.; Georganta, E.; Ulfert, A.-S.

DOI
10.1016/j.copsyc.2024.101837

Publication date
2024

Document Version
Final published version

Published in
Current Opinion in Psychology

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Psychology

Review

AI-teaming: Redefining collaboration in the digital era

Jan B. Schmutz[1], Neal Outland[2], Sophie Kerstan[3], Eleni Georganta[4] and Anna-Sophie Ulfert[5]

Abstract

Integrating artificial intelligence (AI) into human teams, forming human-AI teams (HATs), is a rapidly evolving field. This overview examines the complexities of team constellations and dynamics, trust in AI teammates, and shared cognition within HATs. Adding an AI teammate often reduces coordination, communication, and trust. Further, trust in AI tends to decline over time due to initial overestimation of capabilities, impairing teamwork. Despite AI's potential to enhance performance in contexts like chess and medicine, HATs frequently underperform due to poor team cognition and inadequate mutual understanding. Future research must address these issues with interdisciplinary collaboration between computer science and psychology and advance robust theoretical frameworks to realize the full potential of human-AI teaming.

Addresses

[1] Department of Psychology, University of Zurich, Switzerland
[2] Department of Psychology, University of Georgia, United States
[3] Department of Management, Technology and Economics, ETH Zurich, Switzerland
[4] Faculty of Social and Behavioural Sciences, University of Amsterdam, the Netherlands
[5] Department of Industrial Engineering & Innovation Sciences, Eindhoven University of Technology, Eindhoven, the Netherlands

Corresponding author: Schmutz, Jan B. (jan.schmutz@psychologie.uzh.ch)

Keywords

Human-AI teaming, Human autonomy team, Human-AI collaboration, Artificial intelligence, Trust.

Introduction

Dave: "Open the pod-bay doors, HAL."

HAL: "I am sorry, Dave, I am afraid I can't do that."

Current Opinion in Psychology 2024, **58:**101837

This review comes from a themed issue on Artificial Intelligence (2025)

Edited by Broderick Turner and Rebecca Walker Reczek

For complete overview about the section, refer Artificial Intelligence (2025)

Available online 29 June 2024

https://doi.org/10.1016/j.copsyc.2024.101837

In the realm of collaboration between humans and autonomous systems, few moments in cinema capture the complex dynamics of this relationship as poignantly as the standoff between astronaut Dave Bowman and the HAL 9000 computer in Stanley Kubrick's "2001: A Space Odyssey" when HAL, the ostensibly compliant artificial intelligence (AI)[1] refuses to open the spaceship doors for Dave, leaving him to face the vastness of space. This line may not reflect an actual shift in HAL's role or decision-making as much as it reveals Dave's realization that he is interacting with an autonomous entity, one with its own interpretations of the mission's best interests, rather than simply a supportive tool. The scene encapsulates the profound challenges inherent in human-AI teams (HATs): as autonomous systems become more integrated into our daily lives and work, understanding the nature of interactions between AI and humans, especially as team members, becomes increasingly crucial.

With the rapid development of AI in recent years, AI is taking on more complex roles and becoming increasingly involved in team interactions. Research has shifted from seeing AI as a mere tool to recognizing it as a collaborative team member [1,2]. Autonomous agents can now participate in teamwork activities involving monitoring, coordination, reallocation of tasks, and continuous interaction with human team members or other AI [3,4]. HATs are defined as at least one human working together with at least one autonomous agent, with a partial or high degree of self-governance over decision-making, adaptation, or communication [5,6], working towards a common goal [7]. While there is ongoing debate about whether we should consider AI as full-fledged teammates [8], it is undeniable that in the near future, AI will become even more capable of interacting and collaborating with humans and embody at least some characteristics of teammates.

In this article, we provide an overview of the rapidly evolving, interdisciplinary field of HAT research. Research has spanned computer science, cognitive

[1] Artificial intelligence is defined as the ability of a system to correctly interpret external data, to learn from that data, and to achieve specific goals and tasks through flexible adaptation, rather than merely performing computations provided by human users [51]. Numerous applications of AI exist, including machines capable of playing games such as checkers and chess, as well as programs that can analyze and replicate language [52].

Table 1

Overview of the state of research, challenges and suggestions for future research about human-AI teams.

	Inputs	Team Processes	Emergent States		Outputs
	Team composition	Communication & Coordination	Trust	Shared Cognition	Team Performance
State of Research	HATs are perceived more negatively than other team compositions	HATs coordinate and communicate worse than human-only teams	Trust in AI team members decreases over time and is lower than in humans	HATs demonstrate lower shared mental models than human-only teams	HATs underperform compared to human-only teams
Challenges	Findings are based on participants' views and expectations	Unclear explanations for communication challenges	Focus mainly on dyadic human-AI trust relationships	Lack of clear design and implementation of AI's mental model	Unfamiliarity of humans with AI interaction
Suggestions for future research	Investigation with real AI interactions	Investigation in the field with organizational team tasks	Adoption of multilevel perspective of trust	New assessments and models for capturing mental models	Deeper understanding of AI's and human's theory of mind

sciences, and organizational psychology; however, it has often been done in isolated efforts. This has sometimes led social sciences to underestimate technological complexity and computer science to oversimplify psychological aspects of group dynamics [4]. To move forward and gain a more holistic understanding of HATs, we strive to integrate theoretical and empirical research from both fields and provide insights into the future direction of HAT research. Table 1 provides an overview of the state of research, challenges and suggestions for future research about human-AI teams.

Current topics in Human-AI research

Team composition

Team composition in human teams, defined as the configuration of member attributes, fundamentally shapes teamwork and performance [9]. Researching team compositions is challenging due to the nearly infinite combinations of team sizes and member attributes. Adding an AI teammate increases this complexity, which might explain why only two studies have investigated team composition in HATs so far.

McNeese et al. [10] examined various team compositions—human-only, human-human-AI, human-AI-AI, and AI-only—in a simulated emergency response task. They found that mixed HATs performed the worst, with AI-only teams achieving the highest performance and situation awareness (SA) scores [11]. Despite the lower performance, human-only teams had higher levels of shared mental models (SMMs), defined as collectively held knowledge structures that facilitate collaborative functioning within a team [12], than mixed HATs. Adding an AI seemed to hinder SMM development among human members.

The second study, a qualitative study by Musick et al. [13], demonstrated similar findings. The researchers compared human-only teams with HATs through interviews conducted with participants who had worked in either setting in a laboratory environment. Participants reported that the inclusion of an AI in the team impeded the development of SMMs, as communication with AI was challenging. Conversely, teams with more human members were seen as better at developing SMMs and coordination. These perceptions were mainly attributed to the participants' views rather than the behavior of the AI, as participants in both conditions were actually interacting with humans but were deceived into believing they were working with an AI team member.

Team processes

The integration of AI into teams significantly impacts team processes like coordination and communication, which are crucial for overall performance [14]. Despite their importance, only a limited number of studies have explicitly investigated such team processes in HATs.

Dell'Acqua et al. [15] found that replacing a human team member with an AI agent in a video game increased coordination failures and decreased performance due to reduced effort among human team members, especially in low—and medium-skilled teams. Similarly, Johnson et al. [16] highlighted communication challenges in HATs under degraded conditions, showing that coordination training improved performance, especially in low-performing HATs. Demir et al. [17] investigated three-person teams involving unmanned aircraft systems (UAS). They found that HATs exhibited worse coordination and performance than both expert and novice human-only teams. Additionally, Demir et al. [18], using the same UAS operations task, underscored the importance of "pushing information," where team members proactively share information, as opposed to "pulling information," where team members actively request information. They found that HATs had lower levels of both pushing and pulling information, leading to lower performance than human-only teams. These studies reveal a common challenge: the presence of an AI team member might impair important team processes, such as communication and coordination, resulting in less effective communication among team members.

Emergent states

Emergent states refer to team characteristics that are typically dynamic and change based on the team's context, inputs, processes, and outcomes. These states encompass the cognitive, motivational, and affective conditions of the team rather than the nature of interactions between team members [14]. In the following, we focus on the two most prevalent emergent states in human-AI teaming: Trust and shared cognition (i.e., SMMs).

Trust

Trust in new technologies and AI has been a focal point of research in robotics and human–computer interactions for decades [19]. Trust in AI has been extensively studied, as reviewed by Glikson et al. [20] and Bach et al. [19]. Trust is crucial for human-AI collaboration, similar to human-only teams [20,21].

Generally, research indicates that trust in AI teammates decreases with poor performance [15,22], a phenomenon also seen in human-only teams [22]. Teams generally trust new human members more than AI [23], with overall trust being higher in human-only teams than in HATs [24]. An exception to this trend is noted by Zhang et al. [25], who found that in a chess task, humans trust suggestions from AI more than those from humans. The explanation might be related to the nature of the task itself. It is widely recognized that chess AIs are far superior to human players, a fact that may have led to higher trust towards the AI. Moreover, trust in AI tends to decline over time due to initial overestimation of AI capabilities [20,22], unlike human-only teams, where trust builds over time [21]. Finally, various AI characteristics have been identified as factors increasing trust in AI, including tangibility (i.e., physical and visual presence), transparency, explainability, reliability, task characteristics (with more trust in AI during technical tasks versus social tasks), and specific AI behavior (e.g., responsiveness, adaptiveness, pro-social behavior) [5,20,26,27].

While most studies have focused on trust within a human-AI dyad—specifically, the trust of an individual human in an AI agent—only recently has research adopted a multilevel perspective, recognizing team trust in HATs as an emergent state at the team level [24,28]. This shift underscores the need for more multilevel assessments of trust to capture the phenomenon more accurately in HATs.

Shared cognition

Shared cognition includes concepts like SMMs, team SA, transactive memory systems, and collective intelligence [4,29–31]. We focus on SMMs, which are one of the most prominent topics in HAT research. SMMs improve teamwork by aligning individual team members' mental models, allowing team members to anticipate each other's needs, thus enhancing implicit coordination and performance [32,33].

Humans in HATs typically report lower levels of SMMs compared to human-only teams [10]. Although some studies have explored SMMs in HATs, none have quantitatively linked it to team performance. Challenges include the design and implementation of the AI's mental model and the methods for eliciting and measuring Human-AI SMMs [34].

For SMMs to work in HATs, humans must develop a mental model that includes both human and AI team members, with the latter accepted both as part of the team and as an intelligent system [34]. The transparency and explainability of AI systems are crucial for humans in building this mental model [29]. Similarly, for AI systems to collaborate effectively with humans, they must model their human teammates, a capability known as the Machine Theory of Mind [35]. Recent advancements in computer and cognitive science have made promising progress in constructing models of human inference, utilizing psychological theories to simulate human behavior analogically to AI [35–37].

Team performance

In real-world contexts, we would only deploy HATs if they outperform humans or AI alone. Indeed, it is widely believed that the so-called "human-AI centaur" (a system that combines human intelligence and AI) can outperform humans or AI alone. This is true in contexts like chess [38] or in medicine, where AI-supported systems usually lead to more accurate diagnoses [39]. However, evidence suggests that when collaboration between humans and AI becomes more intertwined, HATs usually underperform compared to human-only (or AI-only) teams due to worse team cognition [29,40], lower trust [15,20,22], coordination failures [15], inadequate communication [17,41], or the lack of an accurate human's mental model of the AI [42].

One reason HATs often underperform compared to human-only teams might be that humans are not yet accustomed to adequately interacting with AI. This unfamiliarity stems from a lack of mental models for how AI systems operate, which hampers effective communication and coordination. Without a clear understanding of the AI's capabilities, limitations, and decision-making processes, humans struggle to anticipate the AI's actions and integrate its input into their workflows effectively.

Conversely, collaborative AI systems might not yet be fully equipped to function as effective team members, lacking an adequate Machine Theory of Mind. Without understanding human behavior, AI struggles to predict actions, respond to dynamic interactions, and adapt to human needs, causing misaligned actions and coordination issues [5,35]. This deficiency limits the AI's ability to contribute meaningfully to the team. Consequently, both humans and AI must evolve to develop better mutual understanding and adaptability to achieve the superior performance theorized for "human-AI centaurs."

A glimpse into the future

In the following, we outline five research priorities that should be addressed in future studies within the field. First, future research must develop testbeds and paradigms informed by real-world use cases, as current studies lack generalizability to diverse settings. Many AI team members in experiments are better viewed as tools, questioning whether true teaming is being studied. This stems from a lack of clarity concerning how AI autonomy is designed and implemented within teams.

Second, it is crucial to build upon the foundational work of human-agent interaction from the 1990s and early 2000s; we argue that some contemporary research is merely reinventing the wheel. Pioneers like Castelfranchi [43–45] and Sycara [46] laid the essential groundwork for understanding trust, delegation, and social dynamics in human-agent relationships. Their findings remain highly relevant today but receive limited attention in contemporary work. Ignoring these pioneering studies raises concerns about the completeness of our current understanding of human-AI teaming, as we overlook valuable insights from before the field was in vogue.

Third, focusing on the human aspect in teams will become increasingly important. Despite widespread technology use, there is significant diversity in technological literacy and preferences, impacting AI collaboration. Understanding individual differences in willingness and ability to work with AI is needed to inform selection, training, and performance management, especially if the integration of AI colleagues continues at its current pace. Investigating stable individual differences, prior experiences, and enduring perceptions of AI will help researchers and practitioners develop targeted approaches to collaboration, ensuring that AI agents are designed to accommodate the needs and preferences of different team members [47]. Achieving this will require interdisciplinary research drawing from various literature streams, such as technology acceptance, user experience, and inter-group dynamics.

Fourth, the future of human-AI teaming will require strong interdisciplinary collaboration. As the field evolves, researchers from diverse backgrounds—computer science, psychology, cognitive science, sociology, and human factors—must work together and integrate insights. This collaboration will prevent construct proliferation and ensure terminologies and constructs are not siloed within disciplines. It is a necessity, not an option, for HAT research to use the science of teams to further the science of teaming. Laying the groundwork for a common taxonomy for AI across contexts and team types is fundamental to guiding more coherent foundations of the field than unintegrated silos of research.

Fifth, the backbone of any research field is a theory, which provides explanations, integrates findings and predicts future trends or mechanisms. While there are many theories in HATs [28,36,48,49], in our opinion, some fail to achieve the elaborated challenges. Given the nascent reemergence of the human-agent interaction field at work, balancing new insights with established team literature is challenging. Over-relying on traditional team theories may lead to missing new dynamics, while ignoring them can lead to overlooking well-established concepts. A promising theory is the transactive systems model of collective intelligence, which integrates cognitive architectures like instance-based learning theory to design effective human-AI systems [4]. By articulating the critical processes underlying the emergence and maintenance of collective intelligence and integrating insights from cognitive architectures such as instance-based learning theory. This work provides a scalable framework for understanding and designing effective human-AI systems [4].

Conclusion

This review underscores the intricate challenges of HATs as AI takes on increasingly collaborative roles in the workplace. The recent release of OpenAI's ChatGPT-4o [50] and its advanced voice assistant demonstrate AI's readiness to function as a teammate. The assistant reads and expresses emotions, proactively asks questions, monitors the environment, and engages in natural, interruptible conversations with humans and other AIs. While technology appears ready to be integrated into teams, it is now imperative for research to forge ahead, unlocking the full potential of human-AI collaboration and paving the way for truly synergistic and interdisciplinary teamwork.

Funding

This work was supported by the Swiss National Science Foundation [grant number PCEFP1_203374].

CRediT author statement

Jan B. Schmutz: Conceptualization, literature search, writing-original draft.

Neal Outland: Conceptualization, writing - review & editing.

Sophie Kerstan: Writing - review & editing.

Eleni Georganta: Writing - review & editing.

Anna-Sopie Ulfert: Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

The AI-based systems Grammarly Inc. and ChatGPT-4o have been used to proofread this manuscript. No new content has been generated by AI.

References

References of particular interest have been highlighted as:

* of special interest
** of outstanding interest

1. ** Seeber I, Bittner E, Briggs RO, de Vreede T, de Vreede G-J, Elkins A, Maier R, Merz AB, Oeste-Reiß S, Randrup N, Schwabe G, Söllner M: **Machines as teammates: a research agenda on AI in team collaboration**. *Inf Manag* 2020, **57**, 103174, https://doi.org/10.1016/j.im.2019.103174.

2. ** Larson L, DeChurch LA: **Leading teams in the digital age: four perspectives on technology and what they mean for leading teams**. *Leader Q* 2020, **31**:101377–101418, https://doi.org/10.1016/j.leaqua.2019.101377.

3. ** O'Neill T, McNeese N, Barron A, Schelble B: **Human–autonomy teaming: a review and analysis of the empirical literature**. *Hum Factors* 2022, **64**:904–938, https://doi.org/10.1177/0018720820960865.

4. * Gupta P, Nguyen TN, Gonzalez C, Woolley AW: **Fostering collective intelligence in human–AI collaboration: laying the groundwork for COHUMAIN**. *Top Cogn Sci* 2023, https://doi.org/10.1111/tops.12679.

5. Hauptman AI, Schelble BG, McNeese NJ, Madathil KC: **Adapt and overcome: perceptions of adaptive autonomous agents for human-AI teaming**. *Comput Hum Behav* 2023, **138**, 107451, https://doi.org/10.1016/j.chb.2022.107451.

6. Lai V, Chen C, Liao QV, Smith-Renner A, Tan C: **Towards a science of human-AI decision making: a survey of empirical studies**. *ArXiv* 2021, https://doi.org/10.48550/arxiv.2112.11471.

7. Wynne KT, Lyons JB: **An integrative model of autonomous agent teammate-likeness**. *Theor Issues Ergon Sci* 2018, **19**: 353–374, https://doi.org/10.1080/1463922x.2016.1260181.

8. * Evans KD, Robbins SA, Bryson JJ: **Do we collaborate with what we design?** *Top Cogn Sci* 2023, https://doi.org/10.1111/tops.12682.

9. Bell ST, Brown SG, Colaneri A, Outland N: **Team composition and the ABCs of teamwork**. *Am Psychol* 2018, **73**:349.

10. McNeese NJ, Schelble BG, Canonico LB, Demir M: **Who/What is my teammate? Team composition considerations in HumanAI teaming**. *IEEE Trans Hum Mach Syst* 2020, **51**: 288–299, https://doi.org/10.1109/thms.2021.3086018.

11. Endsley MR: **Toward a theory of situation awareness in dynamic systems**. *Hum Factors* 1995, **37**:32–64.

12. Mathieu JE, Heffner TS, Goodwin GF, Salas E, Cannon-Bowers JA: **The influence of shared mental models on team process and performance**. *J Appl Psychol* 2000, **85**:273–283.

13. * Musick G, O'Neill TA, Schelble BG, McNeese NJ, Henke JB: **What happens when humans believe their teammate is an AI? An investigation into humans teaming with autonomy**. *Comput Hum Behav* 2021, **122**, 106852, https://doi.org/10.1016/j.chb.2021.106852.

14. Marks MA, Mathieu JE, Zaccaro SJ: **A temporally based framework and taxonomy of team processes**. *Acad Manag Rev* 2001, **26**:356–376, https://doi.org/10.5465/amr.2001.4845785.

15. Dell'Acqua F, Kogut B, Perkowski P: **Super mario meets AI: experimental effects of automation and skills on team performance and coordination**. *Rev Econ Stat* 2023:1–47, https://doi.org/10.1162/rest_a_01328.

16. Johnson CJ, Demir M, McNeese NJ, Gorman JC, Wolff AT, Cooke NJ: **The impact of training on human–autonomy team communications and trust calibration**. *Hum Factors* 2021, **65**: 1554–1570, https://doi.org/10.1177/00187208211047323.

17. Demir M, Amazeen PG, McNeese NJ, Likens A, Cooke NJ: **Team coordination dynamics in human-autonomy teaming**. *Proc Hum Factors Ergon Soc Annu Meet* 2017, **61**:236, https://doi.org/10.1177/1541931213601542.

18. Demir M, McNeese NJ, Cooke NJ: **Team situation awareness within the context of human-autonomy teaming**. *Cognit Syst Res* 2017, **46**:3–12, https://doi.org/10.1016/j.cogsys.2016.11.003.

19. ** Bach TA, Khan A, Hallock H, Beltrão G, Sousa S: **A systematic literature review of user trust in AI-enabled systems: an HCI perspective**. *Int J Hum Comput Interact* 2024, **40**:1251–1266, https://doi.org/10.1080/10447318.2022.2138826.

20. ** Glikson E, Woolley AW: **Human trust in artificial intelligence: review of empirical research**. *Acad Manag Ann* 2020, **14**: 627–660, https://doi.org/10.5465/annals.2018.0057.

21. ** Jong BAD, Dirks KT, Gillespie N: **Trust and team performance: a meta-analysis of main effects, moderators, and covariates**. *J Appl Psychol* 2016, **101**:1134–1150, https://doi.org/10.1037/apl0000110.

22. McNeese N, Demir M, Chiou E, Cooke N, Yanikian G: **Understanding the role of trust in human-autonomy teaming**. *Proc. 52nd Hawaii Int. Conf. Syst. Sci.* 2019, https://doi.org/10.24251/hicss.2019.032.

23. * Georganta E, Ulfert A-S: **My colleague is an AI! Trust differences between AI and human teammates**. *Team Perform Manag: Int J* 2024, **30**:23–37, https://doi.org/10.1108/tpm-07-2023-0053.

24. ** Georganta E, Ulfert A: **Would you trust an AI team member? Team trust in human–AI teams**. *J Occup Organ Psychol* 2024, https://doi.org/10.1111/joop.12504.

25. Zhang G, Chong L, Kotovsky K, Cagan J: **Trust in an AI versus a Human teammate: the effects of teammate identity and performance on Human-AI cooperation**. *Comput Hum Behav* 2023, **139**, 107536, https://doi.org/10.1016/j.chb.2022.107536.

26. Schelble BG, Lopez J, Textor C, Zhang R, McNeese NJ, Pak R, Freeman G: **Towards ethical AI: empirically investigating dimensions of AI ethics, trust repair, and performance in human-AI teaming**. *Hum Factors* 2022, **66**:1037–1055, https://doi.org/10.1177/00187208221116952.

27. Nunes NJ, Cañas JJ, Détienne F, Jorge J, van der Veer G, Brinkman W-P, Leena N, Wong W, Fan X, Oh S, McNeese M, Yen J, Cuevas H, Strater L, Endsley MR: **The influence of agent reliability on trust in human-agent collaboration**. *Proc. 15th Eur. Conf. Cogn. Ergon.: Ergon. Cool Interact.* 2008:1–8, https://doi.org/10.1145/1473018.1473028.

28. ** Ulfert A-S, Georganta E, Jorge CC, Mehrotra S, Tielman M: **Shaping a multidisciplinary understanding of team trust in human-AI teams: a theoretical framework**. *Eur J Work Organ*

Psychol 2024, **33**:158–171, https://doi.org/10.1080/1359432x.2023.2200172.

29. Endsley MR: **Supporting human-AI teams: transparency, explainability, and situation awareness**. *Comput Hum Behav* 2023, **140**, 107574, https://doi.org/10.1016/j.chb.2022.107574.

30. Bienefeld N, Kolbe M, Camen G, Huser D, Buehler PK: **Human-AI teaming: leveraging transactive memory and speaking up for enhanced team effectiveness**. *Front Psychol* 2023, **14**, 1208019, https://doi.org/10.3389/fpsyg.2023.1208019.

31. Aggarwal I, Cuconato G, Ateş NY, Meslec N: **Self-beliefs, transactive memory systems, and collective identification in teams: articulating the socio-cognitive underpinnings of COHUMAIN**. *Top Cogn Sci* 2023, https://doi.org/10.1111/tops.12681.

32. DeChurch LA, Mesmer-Magnus J, Space Johnson C: *Maintaining shared mental models over long-duration exploration missions*. 2015.

33. Rico R, Sanchez-Manzanares M, Gil F, Gibson C: **Team implicit coordination processes: a team knowledge-based approach**. *Acad Manag Rev* 2008, **33**:163–184, https://doi.org/10.2307/20159381.

34. Andrews RW, Lilly JM, Srivastava D, Feigh KM: **The role of shared mental models in human-AI teams: a theoretical review**. *Theor Issues Ergon Sci* 2023, **24**:129–175, https://doi.org/10.1080/1463922x.2022.2061080.

35. Yang SC, Folke T, Shafto P: **The inner loop of collective human–machine intelligence**. *Top Cogn Sci* 2023, https://doi.org/10.1111/tops.12642.

36. Gonzalez C, Admoni H, Brown S, Woolley AW: **COHUMAIN:**
* **building the socio-cognitive architecture of collective human–machine intelligence**. *Top Cogn Sci* 2023, https://doi.org/10.1111/tops.12673.

37. Gervits F, Thurston D, Thielstrom R, Fong T, Pham Q, Scheutz M: **Toward genuine robot teammates: improving human-robot team performance using robot shared mental models**. In *Proceedings of the 19th international conference on autonomous agents and MultiAgent systems, Richland, SC*; 2020:429–437.

38. Cassidy M: *Centaur chess shows power of teaming human and machine*. HUFFPOST; 2014. https://www.huffpost.com/entry/centaur-chess-shows-power_b_6383606.

39. Reverberi C, Rigon T, Solari A, Hassan C, Cherubini P, Cad GG, Group S, Antonelli G, Awadie H, Bernhofer S, Carballal S, Dinis-Ribeiro M, Fernández-Clotett A, Esparrach GF, Gralnek I, Higasa Y, Hirabayashi T, Hirai T, Iwatate M, Kawano M, Mader M, Maieron A, Mattes S, Nakai T, Ordas I, Ortigão R, Zúñiga OO, Pellisé M, Pinto C, Riedl F, Sánchez A, Steiner E, Tanaka Y, Cherubini A: **Experimental evidence of effective human–AI collaboration in medical decision-making**. *Sci Rep* 2022, **12**, 14952, https://doi.org/10.1038/s41598-022-18751-2.

40. Endsley MR, Cooke N, McNeese N, Bisantz A, Militello L, Roth E: **Special issue on human-AI teaming and special issue on AI in healthcare**. *J Cogn Eng Decis Mak* 2022, **16**:179–181, https://doi.org/10.1177/15553434221133288.

41. Demir M, Likens AD, Cooke NJ, Amazeen PG, McNeese NJ: **Team coordination and effectiveness in human-autonomy teaming**. *IEEE Trans Hum Mach Syst* 2017, **49**:150–159, https://doi.org/10.1109/thms.2018.2877482.

42. Bansal G, Nushi B, Kamar E, Lasecki WS, Weld DS, Horvitz E: **Beyond accuracy: the role of mental models in human-AI team performance**. *Proc. AAAI Conf. Hum. Comput. Crowdsourcing* 2019, **7**:2–11, https://doi.org/10.1609/hcomp.v7i1.5285.

43. Castelfranchi C: **Founding agents'" autonomy" on depen-**
* **dence theory**. *ECI* 2000, **1**:353–357.

44. Castelfranchi C: **Modelling social action for AI agents**. *Artificial intelligence* 1998, **103**:157–182.

45. Castelfranchi C, Falcone R: *Principles of trust for MAS: cognitive anatomy, social importance, and quantification*. IEEE; 1998: 72–79.

46. Sycara K, Lewis M: *Team cognition: understanding the factors that drive process and performance*. 2004:203–231, https://doi.org/10.1037/10690-010.

47. Kaplan AD, Kessler TT, Brill JC, Hancock PA: **Trust in artificial**
** **intelligence: meta-analytic findings**. *Hum Factors: J. Hum. Factors Ergon. Soc.* 2021, **65**:337–359, https://doi.org/10.1177/00187208211013988.

48. Dubey A, Jain S, Abhinav K, Arora V, Puttaveerana A: **HACO: a**
** **framework for developing human-AI teaming**. In *Proc. 13th innov. Softw. Eng. Conf. Former. Known India softw. Eng. Conf*; 2020:1–9, https://doi.org/10.1145/3385032.3385044.

49. Madni AM, Madni CC: **Architectural framework for exploring adaptive human-machine teaming options in simulated dynamic environments**. *Systems* 2018, **6**:44, https://doi.org/10.3390/systems6040044.

50. *OpenAI, ChatGPT (GPT-4o)*. 2024.

51. Poole D, Mackworth A, Goebel R: *Computational intelligence: a logical approach*. Oxford University Press; 1998.

52. Li B, Hou B, Yu W, Lu X, Yang C: **Applications of artificial intelligence in intelligent manufacturing: a review**. *Frontiers of Information Technology & Electronic Engineering* 2017, **18**: 86–96, https://doi.org/10.1631/fitee.1601885.

Further information on references of particular interest

1. This paper presents a research agenda created by 65 scientists,
** proposing the exploration of machines as teammates rather than tools, encompassing three design areas and 17 dualities, to organize early research and assess the potential risks and benefits.

2. This paper organizes existing research on leadership and tech-
** nology into four perspectives—technology as context, sociomaterial, creation medium, and teammate—highlighting how digital technologies affect teamwork and leadership, and identifies 12 leadership implications along with directions for future research and practice.

3. This paper defines human–autonomy teaming, synthesizes
** existing empirical research by identifying key findings and critical future research directions, and highlights the importance of understanding the mechanisms linking team inputs to outputs for successful human-autonomy teams.

4. This paper advocates for an interdisciplinary research domain
* called Collective Human-Machine Intelligence (COHUMAIN) to better understand and design sociotechnical systems, emphasizing the integration of diverse disciplinary perspectives and illustrating this approach with recent work on sociocognitive architectures and instance-based learning theory to enhance human-AI collaboration.

8. This paper critically examines the use of terms like "collaboration"
* and "co-workers" to describe human-AI interactions, arguing that such metaphors may misrepresent AI's role and undermine transparency and human dignity; it proposes an alternative ontology emphasizing the heteronomy of machines to human agents and discusses the implications for workplace dynamics and socio-political considerations.

13. This study explores the impact of perceived team composition on
* sentiments, processes, cognitive states, and team cognition by examining 46 human-only teams, some of which believed they included autonomous agents, and proposes a new model for understanding how early-stage action teams achieve effective teamwork and cognitive states.

19 This review examines user trust in AI-enabled systems, empha-
** sizing the importance of a human-centric approach and analyzing 23 empirical studies to identify definitions, influencing factors, and measurement methods, concluding that trust varies by context and is influenced by socio-ethical considerations, technical and design features, and user characteristics, with surveys being the most common measurement method.

20 This review highlights the importance of worker trust in AI for
** successful organizational integration, identifying AI representation

and machine intelligence as key factors influencing trust, and proposes a framework for cognitive and emotional trust shaped by AI's tangibility, transparency, reliability, immediacy, and anthropomorphism; it also notes current research limitations and suggests directions for future studies.

21 ** This meta-analysis of 112 studies (N = 7763 teams) confirms that intrateam trust positively impacts team performance, showing a significant effect size (β = .30), robust across dimensions of trust and controlling for other predictors, with the relationship moderated by task interdependence, authority differentiation, and skill differentiation, thereby integrating intrateam trust research and clarifying conditions where trust most enhances performance.

23. * This study investigates trust within human-AI teams through an online experiment, finding that while perceived trustworthiness and affective interpersonal trust are lower for AI teammates compared to human ones, cognitive trust and trust behaviors remain similar, highlighting that emotional trust is harder to develop for AI teammates despite rational trust based on competence and reliability; this research bridges human-only team trust literature with human-AI collaboration insights.

24. ** This study explores the emergence of team trust in human-AI versus human–human teams through two experimental studies, finding that in two-member teams, human-AI teams exhibit lower interpersonal trust based on perceived trustworthiness and team identification, while in three-member teams, no differences in team trust were observed; these results suggest that while theories of team trust in human-only teams can inform human-AI team research, notable differences warrant further investigation.

28. ** This study proposes a multidisciplinary framework for understanding team trust in human-AI teams, integrating psychology and computer science literature to address the multilevel nature of team trust and the role of AI agents as team members, aiming to enhance research and the design of trustworthy AI collaborators.

36. * This introduction to a special issue explores the rapid development of AI, particularly generative AI like ChatGPT, and its integration into human collaboration, proposing a socio-cognitive architecture for Collective HUman-MAchine INtelligence (COHUMAIN) to understand and enhance human-AI collective intelligence, addressing conceptual foundations, empirical tests, and ethical considerations through nine papers.

43. * This paper argues that autonomy is a relational concept linked to agenthood and delegation, connects autonomy to the theory of dependence, and contends that an agent's autonomy is derived from its architecture and the theory of action.

47. ** This meta-analysis identifies significant factors predicting trust in AI, categorized into human trustor characteristics, AI trustee attributes, and their shared interaction context, analyzing data from 65 articles and four common AI uses (chatbots, robots, automated vehicles, and nonembodied algorithms), revealing key predictors like AI reliability and anthropomorphism, and highlighting areas lacking empirical research to guide the design of systems that elicit desired levels of trust.

48. ** This paper presents a taxonomy of human-AI teaming concepts, extending the Java Agent Development Framework (JADE) to support this taxonomy through the Human-AI Collaboration (HACO) framework, which facilitates model-driven development of human-AI systems via a graphical user interface. A user study confirms HACO's promise, demonstrating significant reductions in development effort for use cases in a contact center.